Paula Kotzé Gary Marsden Gitte Lindgaard
Janet Wesson Marco Winckler (Eds.)

LNCS 8119

Human-Computer Interaction – INTERACT 2013

14th IFIP TC 13 International Conference
Cape Town, South Africa, September 2013
Proceedings, Part III

3 Part III

ifip

Springer

Lecture Notes in Computer Science 8119

Paula Kotzé Gary Marsden Gitte Lindgaard
Janet Wesson Marco Winckler (Eds.)

Human-Computer Interaction – INTERACT 2013

14th IFIP TC 13 International Conference
Cape Town, South Africa, September 2-6, 2013
Proceedings, Part III

Springer

Volume Editors

Paula Kotzé
CSIR Meraka Institute, PO Box 395, Pretoria 0001, South Africa
E-mail: pkotze1@csir.co.za

Gary Marsden
University of Cape Town, Department of Computer Science
Private Bag X3, Rondebosch 7701, Cape Town, South Africa
E-mail: gaz@cs.uct.ac.za

Gitte Lindgaard
Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada
E-mail: gitte.lindgaard@carleton.ca

Janet Wesson
Nelson Mandela Metropolitan University, Department of Computing Sciences
PO Box 77000, Port Elisabeth 6031, South Africa
E-mail: janet.wesson@nmmu.ac.za

Marco Winckler
University Paul Sabatier, Institute of Research in Informatics of Toulouse (IRIT)
118 Route de Narbonne, 31062 Toulouse Cedex 9, France
E-mail: winckler@irit.fr

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-40476-4 e-ISBN 978-3-642-40477-1
DOI 10.1007/978-3-642-40477-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013945558

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

INTERACT 2013 was the 14^{th} of a series of INTERACT international conferences supported by the International Federation for Information Processing (IFIP) Technical Committee 13 on Human–Computer Interaction.

This year, INTERACT was held in Cape Town (South Africa), organized by the Nelson Mandela Metropolitan University (Port Elizabeth) and the Meraka Institute of Council for Scientific and Industrial Research (Pretoria) in collaboration with the University of Cape Town.

The Conference theme for INTERACT 2013, "Designing for Diversity," recognizes the interdisciplinary, multidisciplinary and intercultural spirit of human–computer interaction (HCI) research and practice. The conference welcomes research and reports of practice that acknowledge diverse disciplines, abilities, cultures, and societies, and that address both the technical and social aspects of HCI. Within the broad umbrella of HCI, the conference sought contributions addressing new and emerging HCI disciplines, bridging cultural differences, and tackling important social problems.

Like its predecessors, INTERACT 2013 highlighted, to both the academic and the industrial world, the importance of the HCI discipline and its most recent breakthroughs on current applications. Both experienced HCI researchers and professionals, as well as newcomers to the HCI field, interested in designing or evaluating interactive software, developing new interaction technologies, or investigating overarching theories of HCI, found in INTERACT 2013 an exciting forum for communication with people of similar interests, to encourage collaboration and to learn.

INTERACT 2013 brought the conference to South Africa and Africa for the very first time. The African tradition of HCI focuses very much on the human and social aspects of HCI, recognizing the diversity of its people and the circumstance in which they go about their everyday lives. We hope that INTERACT 2013 will be remembered as a conference that brought the diversity of HCI research to the forefront, making the computerized world a better place for all, regardless of where they come from.

INTERACT 2013 took place 29 years after the first INTERACT held in September 1984 in London, UK. The IFIP Technical Committee 13 aims to develop 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.

We thank all the authors who chose INTERACT 2013 as the venue to publish their research. This was again an outstanding year for the conference in terms of submissions in all the technical categories, especially since the conference moved away from the traditional predominantly European venues. In total, we received 639 submissions. Of these, 270 submissions were accepted:

- 128 as full research papers
- 77 as short research papers
- 31 as interactive posters
- 2 as industrial programme papers
- 4 as panels
- 1 as a special interest group
- 1 as a tutorial
- 9 as workshops
- 9 to the African Masters Consortium
- 8 to the Doctoral Consortium

The acceptance rate for the full and short research papers was 31% and 45%, respectively.

A Programme Committee meeting consisting of the Technical Programme Chairs and the Track Chairs, as well as member of IFIP Technical Committee 13, preceded the final decision on which submissions to accept. This powerful effort was only possible thanks to the diligent work of many people. Our sincere gratitude goes to the almost 700 members of our International Programme Committee who willingly assisted and ensured the high quality of the INTERACT Conference papers was properly maintained. Although some people had to be bullied into reviewing (sorry about that), everyone submitted their reviews on time without a murmur of complaint. Thank you all for the effort that you so obviously put into this task. A special thank you must go to our Track Chairs, who put in a tremendous amount of work to ensure that quality was maintained throughout.

In addition, we have to thank the members of the Organizing Committee, the staff at the Council for Industrial and Scientific Research, Nelson Mandela Metropolitan University and the University of Cape Town for their unflagging assistance with all aspects of planning and managing the many administrative and organizational issues. We also have to thank our student volunteers for making sure that everything ran smoothly at the conference itself.

Finally, we wish to express a special thank you to the Proceedings Publication Chair, Marco Winckler, who painstakingly put this volume together.

September 2013

Paula Kotzé
Janet Wesson
(INTERACT 2013 Conference Chairs)
Gary Marsden
Gitte Lindgaard
(INTERACT 2013 Technical Programme Chairs)

IFIP TC13

Established in 1989, the International Federation for Information Processing Technical Committee on Human–Computer Interaction (IFIP TC13) is an international committee of 30 member national societies and seven 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 relationship 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 worldwide.

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; modeling 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 multimedia systems to match system needs and organizational practices, etc.).

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.

IFIP TC13 stimulates working events and activities through its Working Groups (WGs) and Special Interest Groups (SIGs). WGs and SIGs 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.

- 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 will promote the use of knowledge, concepts, methods and techniques that enable 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 TC13. It aims to establish a study and research program that will combine both scientific work and practical applications in the fields of HCI and visualization. It will integrate 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.
- SIG 13.1 (HCI and International Development) aims to promote the application of interaction design research, practice and education to address the needs, desires and aspirations of people in the developing world; support and develop the research, practice and education capabilities of HCI institutions and organizations based in the developing world; develop links between the HCI community in general, and IFIP TC13 in particular, with other relevant communities involved in development, especially IFIP WG 9.4 Computers in Developing Countries.

- SIG 13.2 (Interaction Design and Children) aims to provide a forum for all things relating to interaction design and HCI where the intended users or appropriators of the technology or service are children. The definition of children is broad rather than narrow, including toddlers and teenagers, but the core work, currently at least, is with children in junior schools.

New Working Groups and Special Interest Groups are formed as areas of significance to HCI arise. Further information is available at the IFIP TC13 website: http://www.tc13.org

IFIP TC13 Members

Australia
Judy Hammond
Australian Computer Society

Australia
Andreas Holzinger
Austrian Computer Society

Belgium
Monique Noirhomme-Fraiture
Fédération des Associations
Informatiques de Belgique

Brazil
Simone Diniz Junqueira Barbosa
Brazilian Computer Society (SBC)

Bulgaria
Kamelia Stefanova
Bulgarian Academy of Sciences

Canada
Heather O'Brien
Canadian Information Processing
Society

Chile
Jaime Sánchez
Chilean Society of Computer Science

China
Chengqing Zong
Chinese Institute of Electronics

Cyprus
Panayiotis Zaphiris
Cyprus Computer Society

Czech Republic
Vaclav Matousek
Czech Society for Cybernetics
and Informatics

Denmark
Torkil Clemmensen
Danish Federation
for Information Processing

Finland
Kari-Jouko Räihä
Finnish Information Processing
Association

France
Philippe Palanque
Société des Electriciens
et des Electroniciens (SEE)

Germany
Tom Gross
Gesellschaft für Informatik

Hungary
Cecilia Sik Lanyi
John V. Neumann Computer Society

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

India
Anirudha Joshi
Computer Society of India

Ireland
Liam J. Bannon
Irish Computer Society

Italy
Fabio Paternò
Italian Computer Society

Japan
Yoshifumi Kitamura
Information Processing Society of Japan

Kenya
Dan Orwa
Computer Society of Kenya

Korea
Gerry Kim
KIISE

Malaysia
Chui Yin Wong
Malaysian National Computer Confederation

New Zealand
Mark Apperley
New Zealand Computer Society (NZCS)

Nigeria
Chris C. Nwannenna
Nigeria Computer Society

Norway
Dag Svanes
Norwegian Computer Society

Poland
J.L. Kulikowski
Poland Academy of Sciences

Portugal
Joaquim A. Jorge
Associacão Portuguesa para o Desenvolvimento da Sociedade da Informação (APDSI)

Singapore
Henry B.L. Duh
Singapore Computer Society

South Africa
Janet L. Wesson
The Computer Society of South Africa

Spain
Julio Abascal
Asociación de Técnicos de Informática (ATI)

Sweden
Jan Gulliksen
Swedish Computer Society

Switzerland
Solange Ghernaouti
Swiss Federation for Information Processing

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

UK
Andy Dearden
British Computer Society

USA-Based
John Karat
Association for Computing Machinery (ACM)

Expert Members
Paula Kotzé (South Africa)
Gitte Lindgaard (Australia)
Marco Winckler (France)
Masaaki Kurosu (Japan)
Nikos Avouris (Greece)
Zhengjie Liu (China)

Working Group Chairpersons

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

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

WG13.3 (HCI and Disability)
Gerhard Weber, Germany

WG13.4 (also 2.7) (User Interface Engineering)
Fabio Paternò, Italy

WG13.5 (Human Error, Safety, and System Development)
Chris Johnson, UK

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

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

SIG 13.1 – HCI and International Development
Andy Dearden, UK

SIG 13.2 – Interaction Design and Children
Janet Read, UK

Conference Organizing Committee

INTERACT 2013
Conference Chairs
Paula Kotzé, South Africa
Janet Wesson, South Africa

Technical Programme Chairs
Gary Marsden, South Africa
Gitte Lindgaard, Australia

Long Research Papers Chairs
Gary Marsden, South Africa
Gitte Lindgaard, Australia

Short Research Papers and Interactive Posters Chairs
Peter Forbrig, Germany
Monique Noirhomme, Belgium

Industrial Programme Chairs
Susan Dray, USA
Shikoh Gitau, Kenya
Helga Letowt-Vorbek, South Africa

Keynotes Chairs
Gary Marsden, South Africa
Paula Kotzé, South Africa
Janet Wesson, South Africa

Workshops and Tutorials Chair
Helene Gelderblom, South Africa

Doctoral Consortium Chairs
Philippe Palanque, France
Darelle van Greunen, South Africa
Paula Kotzé, South Africa

African Masters Consortium Chairs
Darelle van Greunen, South Africa
Paula Kotzé, South Africa

Panels and Special Interest Groups Chairs
Andy Dearden, UK
Hannah Thinyane, South Africa
Judy van Biljon, South Africa

Demonstrations and Interactive Experiences Chairs
Stacy Hobson, USA
Kasper Jensen, Namibia

Proceedings Publication Chair
Marco Winckler, France

Student Volunteer Chairs
Thomas Reitmaier, Germany
Funmi Adebesin, South Africa
Darelle van Greunen, South Africa

General Conference Administration and Logistics
Marlene Jivan, South Africa
Alida Veldsman, South Africa

International Programme Committee Members

Cristina, Cachero, Spain
Paul Cairns, UK
Jeff Calcaterra, USA
Karen Calteaux, South Africa
Licia Calvi, The Netherlands
Eduardo Calvillo Gamez, Mexico
José Campos, Portugal
Pedro Campos, Portugal
Maria-Dolores Cano, Spain
Cinzia Cappiello, Italy
Robert Capra, USA
Jorge Cardoso, Portugal
Stefan Carmien, Spain
Maria Beatriz Carmo, Portugal
Luis Carriço, Portugal
Tiziana Catarci, Italy
Augusto Celentano, Italy
Teresa Chambel, Portugal
Olivier Chapuis, France
Monchu Chen, Portugal
Yunan Chen, USA
Kelvin Cheng, Singapore
Keith Cheverst, UK
Sonia Chiasson, Canada
Yoram Chisik, Portugal
Luca Chittaro, Italy
Yu-kwong Chiu, USA
Georgios Christou, Cyprus
Andrea Civan Hartzler, USA
Torkil Clemmensen, Denmark
Todd Cochrane, New Zealand
Luis Coelho, Portugal
Nathalie Colineau, Australia
Karin Coninx, Belgium
Stéphane Conversy, France
Michael Cooke, Ireland
Nuno Correia, Portugal
Maria Francesca Costabile, Italy
Celine Coutrix, France
Nadine Couture, France
Anna Cox, UK
Leonardo Cunha de Miranda, Brazil
Edward Cutrell, India
Raimund Dachselt, Germany
Carla Maria Dal Sasso Freitas, Brazil
Peter Dalsgaard, Germany
Jose Danado, Portugal
Chi Tai Dang, Germany
Peter Dannenmann, Germany
John Darzentas, Greece
Berardina De Carolis, Italy
Tjerk de Greef, The Netherlands
Elena de la Guia Cantero, Spain
Alexander De Luca, Germany
Giorgio De Michelis, Italy
Luigi De Russis, Italy
Boris de Ruyter, The Netherlands
Marco de Sá, USA
Clarisse de Souza, Brazil
Ruth de Villiers, South Africa
Andy Dearden, UK
Alexandre Demeure, France
Parag Deshpande, Sweden
Heather Desurvire, USA
Nicoletta Di Blas, Italy
Ines Di Loreto, Italy
Paulo Dias, Portugal
Lily Diaz, Finland
Claire Diederich, Belgium
Anke Dittmar, Germany
Tanja Döring, Germany
Susan Dray, USA
Carlos Duarte, Portugal
Emmanuel Dubois, France
Cathy Dudek, Canada
Andreas Duenser, New Zealand
Sophie Dupuy-Chessa, France
Hamid Ekbia, USA
Linda Elliott, USA
Niklas Elmqvist, USA
Thomas, Erickson, USA
Dominik Ertl, Austria
Parisa Eslambolchilar, UK
Giorgio Faconti, Italy
João Falcão e Cunha, Portugal
Pedro Faria Lopes, Portugal
Robert Farrell, USA
Ava Fatah gen Schieck, UK
Jean-Daniel Fekete, France
Daniel Felix, Switzerland

Xavier Ferre, Spain
Mirko Fetter, Germany
Sebastian Feuerstack, Brazil
Danyel Fisher, USA
Francisco Florez-Revuelta, UK
Daniela Fogli, Italy
Eelke Folmer, USA
Joan Fons, Spain
Manuel Joao Fonseca, Portugal
Peter Forbrig, Germany
Derek Foster, UK
Marcus Foth, Australia
Teresa Franqueira, Portugal
Andre Freire, Brazil
Mathias Frisch, Germany
Dominic Furniss, UK
Luigi Gallo, Italy
Teresa Galvao, Portugal
Nestor Garay-Vitoria, Spain
Roberto García, Spain
Vaibhav Garg, USA
Anant Bhaskar Garg, USA
Jose Luis Garrido, Spain
Franca Garzotto, Italy
Jean-marc Gauthier, USA
Miguel Gea, Spain
Patrick Gebhard, Germany
Sven Gehring, Germany
Helene Gelderblom, South Africa
Cristina Gena, Italy
Nahum Gershon, USA
Werner Geyer, USA
Florian Geyer, Germany
Gautam Ghosh, Norway
Patrick Girard, France
Shikoh Gitau, Kenya
Vlado Glavinic, Croatia
Kentaro Go, Japan
Daniel Gonçalves, Portugal
Jose Luis Gonzalez Sanchez, Spain
Alice Good, UK
Phil Gosset, USA
Nitesh Goyal, USA
Cornelia Graf, Austria
Nicholas Graham, Canada
Andrina Granic, Croatia
Raphael Grasset, Austria
Olivier Grisvard, France
Tom Gross, Germany
Joshua Gross, USA
John Grundy, Australia
Tiago Guerreiro, Portugal
Yves Guiard, France
Nuno Guimaraes, Portugal
Jan Gulliksen, Sweden
Sainath Gupta, India
Adriana Mihaela Guran, Romania
Francisco Luis Gutierrez Vela, Spain
Mieke Haesen, Belgium
Hans Hagen, Germany
Lars Hallnäs, Sweden
Kim Halskov, Denmark
Martin Halvey, UK
Judy Hammond, Australia
Mark Hancock, UK
Richard Harper, UK
Joseph Hategekimana, Norway
Kirstie Hawkey, Canada
Elaine Hayashi, Brazil
Ruediger Heimgaertner, Germany
Ingi Helgason, UK
Bart Hengeveld, The Netherlands
Michael Herczeg, Germany
Marlien Herselman, South Africa
Morten Hertzum, Denmark
Wilko Heuten, Germany
Michael Hildebrandt, Norway
Martin Hitz, Austria
Stacy Hobson, USA
Christina Hochleitner, Austria
Stefan Holmlid, Sweden
Clemens Holzmann, Austria
Kasper Hornbæk, Denmark
Heiko Hornung, Brazil
John Hosking, Australia
Juan Pablo Hourcade, USA
Ko-Hsun Huang, Taiwan
William Hudson, UK
Stéphane Huot, France
Tim Hussein, Germany

Ebba Hvannberg, Iceland
Ioanna Iacovides, UK
Junko Ichino, Japan
Netta Iivari, Finland
Shamsi Iqbal, USA
Petra Isenberg, France
Minna Isomursu, Finland
Ido A. Iurgel, Germany
Ole Sejer Iversen, Denmark
Robert Jacob, USA
Mikkel Jakobsen, Denmark
Jacek Jankowski, Ireland
Anders Jansson, Sweden
Kasper Jensen, Namibia
Chris Johnson, UK
Simon Jones, UK
Joaquim Jorge, Portugal
Rui Jose, Portugal
Martino Jose Mario, Brazil
Anirudha Joshi, India
Christophe Jouffrais, France
Marko Jurmu, Japan
Don Kalar, USA
Martin Kaltenbrunner, Austria
Santosh Kalwar, Finland
Matthew Kam, USA
Bridget Kane, Ireland
Victor Kaptelinin, Sweden
Mayur Karnik, Portugal
Kostas Karpouzis, Greece
Konstantinos Kazakos, Australia
Ryan Kelly, UK
Pramod Khambete, India
Vassilis-Javed Khan, The Netherlands
Olga Khroustaleva, USA
Hyungsin Kim, USA
Jayne Klenner, UK
Gerhard Knolmayer, Switzerland
Christophe Kolski, France
Ilpo Koskinen, Australia
Paula Kotzé, South Africa
Kostas Koukouletsos, Greece
Subramanian Krishnamurthy, India
Antonio Krüger, Germany
Sari Kujala, Finland
Hannu Kukka, Finland
Juliusz Kulikowski, Poland
Ann Lantz, Sweden
Rosa Lanzilotti, Italy
Astrid Twenebowa Larssen, Ghana
Marta Larusdottir, Iceland
Yann Laurillau, France
Marco Lazzari, Italy
Eric Lecolinet, France
Jong-Seok Lee, Korea
Juha Leino, Finland
Teemu Heikki Anttoni Leinonen, Finland
Jair Leite, Brazil
Barbara, Leporini, Italy
Sophie Lepreux, France
Helga Letowt-Vorbek, South Africa
Stefano Levialdi, Italy
Henry Lieberman, USA
Gitte Lindgaard, Australia
Zhengjie Liu, China
Sara Ljungblad, Sweden
Steffen Lohmann, Germany
Nicole Lompré, France
Víctor M. López Jaquero, Spain
Fabien Lotte, France
Stephanie Ludi, USA
Bernd Ludwig, Germany
Andreas Luedtke, Germany
Christopher Lueg, Australia
Artur Lugmayr, USA
Jo Lumsden, UK
Christof Lutteroth, New Zealand
Kris Luyten, Belgium
Saturnino Luz, Ireland
Baillie Lynne, Austria
José A. Macías, Spain
Allan MacLean, UK
Rui Madeira, Portugal
Joaquim Madeira, Portugal
Christian Maertin, Germany
Charlotte Magnusson, Sweden
Ana Gabriela Maguitman, Argentina
Murni Mahmud, Malaysia
Angela Mahr, Germany

Fabio Pittarello, Italy
Andreas Pleuß, Ireland
Klaus Pohl, Germany
Roberto Polillo, Italy
Vesna Popovic, Australia
Dorin Mircea Popovici, Romania
Roman Popp, Austria
Benjamin Poppinga, Germany
Chris Power, UK
Mathura Prasad Thapliyal, India
Raquel Prates, Brazil
Costin Pribeanu, Romania
Angel Puerta, USA
Andreas Pusch, Germany
Alexandra Queirós, Portugal
Victor M.R. Penichet, Spain
Kari-Jouko Räihä, Finland
Nitendra Rajput, India
Ismo Rakkolainen, Finland
David Raneburger, Austria
Roberto Ranon, Italy
Dimitrios Raptis, Denmark
Matthias Rauterberg, The Netherlands
Janet Read, UK
Madhu Reddy, USA
Stuart Reeves, UK
Patrick Reignier, France
Rene Reiners, Germany
Thomas Reitmaier, Germany
Christian Remy, Switzerland
Karen Renaud, UK
Malte Ressin, UK
Simos Retalis, Greece
Yann Riche, USA
Thomas Rist, Germany
Guillaume Rivière, France
Toni Robertson, Australia
Simon Robinson, UK
Carsten Röcker, Germany
Armanda Rodrigues, Portugal
Yvonne Rogers, UK
Markus Rohde, Germany
Teresa Romão, Portugal
Mario Romero, Sweden
Jose, Rouillard, France
Mark Rouncefield, UK
Nicolas Roussel, France
Banani Roy, Canada
Jaime Ruiz, USA
Saqib Saeed, Pakistan
Pascal Salembier, France
Antti Salovaara, Finland
Krystian Samp, Ireland
Paulo Sampaio, Brazil
Jaime Sanchez, Chile
Bengt Sandblad, Sweden
Frode Eika Sandnes, Norway
Vagner Santana, Brazil
Carmen Santoro, Italy
Teresa Sarmento, Portugal
Corina Sas, UK
Christine Satchell, Australia
Cheryl Savery, Canada
Anthony Savidis, Greece
Dominique L. Scapin, France
Thomas Schlegel, Germany
Michael Schmidt, Germany
Kevin Schneider, Canada
Johannes Schöning, Belgium
Gerhard Schwabe, Switzerland
Marc Seissler, Germany
Dev Sen, Canada
Montserrat Sendin Veloso, Spain
Tacha Serif, Turkey
Ehud Sharlin, Canada
Aditi Sharma Grover, South Africa
Moushumi Sharmin, USA
Huihui Shi, Germany
Cecília Sik Lányi, Hungary
Bruno Silva, Brazil
Paula Alexandra Silva, Portugal
Frutuoso Silva, Portugal
Joao Carlos Silva, Portugal
Milene Silveira, Brazil
Gavin Robert Sim, UK
Carla Simone, Italy
Mikael B. Skov, Denmark
Andrew Smith, South Africa
Brian Smith, USA
David Smith, USA

Danny Soroker, USA
Kenia Sousa, Belgium
Martin Spindler, Germany
Jan Stage, Denmark
Christian Stary, Austria
John Stasko, USA
Markus Stolze, Switzerland
Simone Stumpf, UK
Noi Sukaviriya, USA
S. Shyam Sundar, USA
Alistair Sutcliffe, UK
David Swallow, UK
Anthony Tang, Canada
Charlotte Tang, USA
John Tang, USA
Jean-Claude Tarby France
Franck Tarpin-Bernard, France
Deborah Tatar, USA
Luis Teixeira, Portugal
Carlos Teixeira, Portugal
António Teixeira, Puerto Rico
Harold Thimbleby, UK
Hannah Thinyane, South Africa
Jakob Tholander, Sweden
Martin Tomitsch, Australia
Ilaria Torre, Italy
Stefan Trausan-Matu, Romania
Manfred Tscheligi, Austria
Nikolaos Tselios, Greece
Simon Tucker, UK
Susan Turner, UK
Phil Turner, UK
Leon Urbas, Germany
Kaisa Väänänen-Vainio-Mattila, Finland
Heli Väätäjä, Finland
Teija Vainio, Finland
Leonel Valbom, Portugal
Pedro Valderas, Spain
Judy van Biljon, South Africa
Thea van der Geest, The Netherlands
Darelle van Greunen, South Africa
Bradley van Tonder, South Africa
Lexvan Velsen, The Netherlands
Lynette van Zijl, South Africa
Jean Vanderdonckt, Belgium
Radu-Daniel Vatavu, Romania
Manuel Veit, France
Ángel Velázquez-Iturbide, Spain
Janneke Verhaegh, The Netherlands
Arnold P.O.S. Vermeeren, The Netherlands
Bart Vermeersch, Belgium
Jo Vermeulen, Belgium
Nadine Vigouroux, France
Giuliana Vitiello, Italy
Arnd Vitzthum, Germany
Stephen Voida, USA
Jörg Voskamp, Germany
Ljiljana Vukelja, Switzerland
Dhaval Vyas, India
QuanWei Wang, China
Ian Warren, New Zealand
Gerhard Weber, Germany
Gerald Weber, New Zealand
Carolyn Wei, USA
Alexandra Weilenmann, Sweden
Astrid Weiss, Austria
Janet Wesson, South Africa
Mikael Wiberg, Sweden
Peter Wild, UK
David Wilfinger, Austria
Marco Winckler, France
Heike Winschiers-Theophilus, Namibia
Wolfgang Woerndl, Germany
Peter Wolkerstorfer, Austria
Michael Wright, UK
Min Wu, USA
Heng Xu, USA
Yeliz Yesilada, UK
James Young, Canada
Ray Yun, USA
Massimo Zancanaro, Italy
Xiaolong (Luke) Zhang, USA
Martina Ziefle, Germany
Jürgen Ziegler, Germany
Gottfried Zimmermann, Germany
Martin Zimmermann, Germany

Additional Reviewers

Álvaro Fernández
Anke Lehmann
Asbjørn Følstad
Burkay Sucu
Carlos Pereira
Cecilie Brynskov
Diana Lemme
Florian Lettner
Haiyan Jia
Ilaria Liccardi
J.L. Hiah
Jiaqi Nie
Jo Iacovides
Luciana Cardoso de Castro Salgado
Luis Duarte
Markel Vigo
Myriam Lewkowicz
Rui Madeira
Rui Nóbrega
Samuel Silva
Sandy Gould
Sofia Reis
Thomas Birn
Timm Linder
Vinícius Segura
Xavier Le Pallec
Yasmin Al Zokari

Table of Contents – Part III

Long and Short Papers (Continued)

Mobile Usage and Techniques

Mobile UX and Privacy Concerns

Model-Based User Interface Design

Multimodal User Interface Design

Multimodality, Cross-Platform Studies

Narratives in Design

Navigation Aids

Novel User Interfaces

Passwords: e-Authentication

Physical Ergonomics

Road Safety

Seniors and Usability

Social Behaviour, Collaboration and Presence

Social Collaborative Interaction

Social Media

Software Development

Designing Mobile Phone Interfaces for Age Diversity in South Africa: "One-World" versus Diverse "Islands"

Karen Renaud[1], Rénette Blignaut[2], and Isabella Venter[3]

[1] School of Computing Science, University of Glasgow, Glasgow, G12 8RZ, UK
[2] University of the Western Cape, Department of Statistics and Population Studies, Private Bag X17, Bellville, 7535 South Africa
[3] University of the Western Cape, Department of Computer Science, Private Bag X17, Bellville, 7535 South Africa
karen.renaud@glasgow.ac.uk, {rblignaut,iventer}@uwc.ac.za

Abstract. Designing for diversity is a laudable aim. How to achieve this, in the context of mobile phone usage by South African seniors, is a moot point. We considered this question from two possible perspectives: *universal* (one-world) versus *focused* design (designing for diverse "islands" of users). Each island would be characterised by a measure of relative homogeneity in terms of user interface needs. Our particular focus in this paper is age diversity. The universal approach attempts to deliver a design that can be all things to all people – meeting the needs of all users within one user interface. The islander approach delivers specific and different designs for islands within a diverse world. To determine which the best approach would be, in the South African context, we dispatched a team of student researchers to interview participants from an older generation, on a one-to-one basis. It was beneficial to deploy aspiring designers to carry out this research because we wanted to confront aspiring researchers with the differences between their own and other generations' usage of, and attitudes towards, mobile phones. Our study found that there were indeed significant age-related differences in mobile phone usage. Our research delivered insights that led to a model of the factors impacting mobile phone usage of the senior generation as a series of *filters* between the user and their device. We conclude that the island approach is more suitable for age-specific design. This approach might well become less fitting as a more technologically experienced population ages, but at present there is a clear need for an age-sensitive mobile interface design.

Keywords: Mobile Phones, Design, Age.

1 Introduction

The world is a-changing, in different ways and at varying paces. One technology that has changed lives across the planet with dizzying speed is the mobile phone. These devices have become so interwoven with our daily lives that it is hard to imagine a world without them. Indeed, people object when they are denied their phones, even for short periods, evidenced by opposition to the ban on usage during take-offs and

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 1–17, 2013.

landings of airplanes [1]. Interestingly, there are now calls for this ban to be lifted [2]. Our focus in this paper is generational impact on the use of mobile phones, and how this ought to be accommodated in the design of phone interfaces.

Many countries world-wide are grappling with a shift from a demographically young to a demographically old society and the challenges this constitutes in terms of care and having enough young people to finance pensions. This has been a relatively slow but inexorable shift as birth rates have declined and health advances have prolonged life spans. The breakneck speed at which technology has advanced in the same period is something the older generation often finds daunting.

The diffusion of mobile phones into society provides a powerful example of the rate of change the elderly have witnessed in their lifetimes. Whereas it took 95 years for 90% of American homes to get landline phones, it took only 20 years for mobile phones to penetrate the market to the same extent. The young of today are used to this rate of change [3]. There is evidence that the older generation experience particular and challenging difficulties using this mobile technology [4] [5]. Children use mobile phones intuitively [6] but for older users it often seems an alien device [7].

A Google trends search for "senior phone" delivered the graph in **Fig. 1**, which suggests that there is an increasing and sustained interest, worldwide, in mobile phones for the older user. What is not yet clear is what approach should be taken to accommodate the needs of this growing user group.

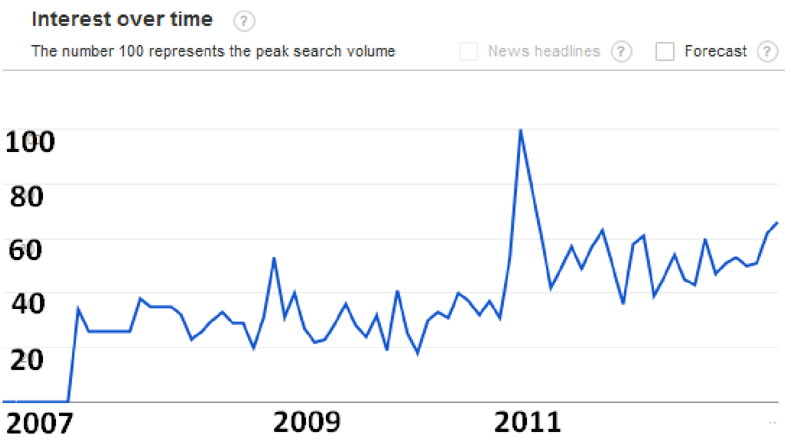


Fig. 1. Google Trends Search for the term "senior phone" carried out 15 January 2012 (http://www.google.com/trends/explore#q=%22senior%20phone%22)

The rest of this paper is organized as follows. Section 2 contemplates the different approaches to designing for diversity. Section 3 presents the design of the study we carried out to determine whether there is a need for an age-sensitive approach to designing mobile phone user interfaces. Sections 4, 5 and 6 detail the three distinct parts of the research, the first carried out by the student researchers, the second the conclusions about mobile phone usage by the young and the third modeling mobile phone usage by the older participants. Section 7 discusses the findings and Section 8 concludes.

2 Designing for Diversity

It is beneficial to ensure that we set out with a clear understanding of the core concept of diversity. Synonyms for diversity are *variety, multiplicity* and *difference*. These suggest designing for ease of use without the expectation of the user having any specific set of skills or prior experience. Users have goals and needs and wish to use the interface to satisfy them: they do not wish to have to undertake training, read manuals or to struggle to use the device. There are at least two ways of designing for diversity.

- The first is to design a *universal* interface offering a wide range of functionality and personalisation features. This option accommodates a wide range of needs and wants, being all things to all users and is followed by smart phone manufacturers. The owner of the smart phone can tailor the phone to their own personal tastes and needs, with a myriad of options and settings available to facilitate the process. Many mobile phones allow the owner to choose a ring tone, background, layout and set of applications, and is referred to as the *one-world* approach [8]. The snag might be that we run the risk of not meeting anyone's needs properly, of delivering the same sub-optimal interface to everyone.
- The second option is to perform *segmentation*, to provide a range of interfaces and devices, each focused on satisfying the needs of a specific group of users. This approach assumes that users can indeed be segmented into islands of relative homogeneity. The users within these islands are considered to share more characteristics with fellow islanders than they do with the inhabitants of other islands, but within each island there is still much diversity. This can be referred to as the *island* approach. This option can have negative connotations and be considered discriminatory. To address this, islands have to delineate with great sensitivity and care. This approach is argued for by Head & Ziolkowski who proposed segmenting mobile phone users in a meaningful way and moving beyond the "one-size-fits-all" model [9].

Fig. 2. Eliminating Complexity by Reducing Functionality (The Age UK Mobile Phone on the right)(http://www.ageuk.org.uk/Products/Communications/Mobile-phones/)

Which approach should be followed in designing mobile phone interfaces for age diversity, the *one-world* or the *island* approach? Prensky makes a strong case for designing for age-based groups [10]. **Fig. 2** (left) offers an extreme example of the most popular way of solving difficulties experienced by the elderly: withdrawing and reducing functionality to the bare minimum. As with many other seemingly quick and easy solutions to intractable problems this approach is amusing rather than effective since it severely reduces the usefulness of the device. Some mobile phone manufacturers have indeed identified the elderly as a group justifying the island approach. Most of the "senior" phones on offer provide severely reduced functionality and almost child-like interfaces (see **Fig. 2** -right). The underlying assumption is unpalatable: the elderly are incapable or don't need all the functionality other users need. While the island approach might well be justified, they appear to have failed in their execution because these phones do not accommodate the diverse needs of all "island" members, and might actually alienate some with their simplistic design.

The failure of the mobile phone industry to meet the needs of older users with their targeted device and interface could be due to the island approach, as a concept, being a fallacy, at least as far as age is concerned [11]. It could also be because the design has veered too far towards an unwarranted assumption of island homogeneity, not acknowledging islander diversity.

The viability of a generation-sensitive island approach, rather than a one-world approach, needs to be studied in order to determine how to design mobile phone interfaces to accommodate the a wide range of generational diversity.

2.1 Purpose of the Study

Our aim is to reveal diversity or homogeneity within and between generational "islands", with the end purpose of approving or denouncing generation-specific design. It could be that there are no, or merely shaky, grounds for the island approach [11]. The evidence might well convince interface designers that the one-world option is the wiser. An ethnographic study will help to deliver this insight.

2.2 Researcher Choice

The critical question is: who should undertake such a study? One of the main concerns is that whoever does the study will be influenced by the natural human tendency to stereotype older people. We all have a tendency to assign people to classes or categories and then to exercise judgement based on this categorisation [12]. This is particularly true of age-specific groups, since age is immediately apparent when we meet someone for the first time. Cuddy and Fiske also point out that "out-groups" (as the young would consider the old or *visa-versa*) are seen to be homogeneous. In terms of how the young see the elderly, this leads to expectations of competence (low) and friendliness (high). Thus two observers might interpret the same incident very differently, especially when they come from two different generations.

A researcher might have an unconscious bias towards the island population, which would influence their interpretation. It is important for the reported findings to be validated, to ensure that they are not merely the product of pre-conceived notions. Fortunately, Caspi reports that inter-age contact serves to reduce stereotype-like attitudes [13]. Moreover, ensuring that individuals from two distinct groups interact has been shown to reduce stereotypical characteristic attributions [14]. Thus, if we incorporate inter-generational contact, on a one-to-one basis, into the study, we can reduce the stereotyping effect. This will lessen the likelihood that researchers merely report pre-conceived notions rather than real effects.

There is another way of contemplating age-specific research. Neugarten, Moore and Lowe argue that older people are more aware of age-appropriate behaviours than are younger members of society [15]. Thus the younger do not have invariant expectations of age-specific behaviour, as do the older. This means that, in contrast to the stereotyping perspective, which suggests pre-conceived expectations, this perspective suggests that they would not have as strongly held behavioural expectations as the older participants. This might make them more open to contemplating the situation as it is, as opposed to merely confirming their own pre-conceived notions about what people of the older generation do and think.

Hence it made sense to dispatch trainee user interface designers to carry out this research. This delivered other benefits too. Where clearly we would like to confront seasoned designers themselves with the findings of this study, it is even better to task designers themselves to carry out the study. This kind of engagement is a very valuable way of exposing interface designers to the needs and opinions of a group of users they wouldn't normally encounter. Moreover, it would be especially valuable for aspiring user interface designers to experience this at the outset of their careers before they establish less than helpful design habits and become less malleable and more likely to hold pre-conceived notions [16].

There is another consideration too. In previous eras the older generation trained and apprenticed the young; they had something valuable to impart, which suggests that the young respected and learnt from them. There is worrying evidence of a shift in attitude. Netz and Ben-Sira asked a range of generations to rate adolescents, middle-aged and elderly people on four semantic categories [17]. All participants rated the "older" category lowest of all the semantic categories. Sharps, Price-Sharps and Hanson compared attitudes to the elderly in Thailand and America and their study confirmed these kinds of negative attitudes towards the elderly in both countries [18]. The widespread use of the word "ageism" is perhaps the most prescient reminder of the fact that the elderly are increasingly considered a burden to society, displacing them from their erstwhile strong position of respect and source of sage advice. Going back to our previous discussion about stereotyping, and the benefit of ensuring that individuals from different groups meet each other, it would seem helpful for students to carry out the research so that a sense of mutual respect can be fostered.

Hence we decided to set this as a task for students, aspiring human-interface designers, to carry out as coursework. The student researchers can undeniably deliver a different and fresh perspective, delivering potentially new and different insights since they are not constrained and influenced by their previous designs and activities.

3 Study Design

The study was structured as shown in **Fig. 3**. In stages 1 and 2 we set coursework for students undertaking a third-year Computer Science Human Computer Interface course. They were to interview an older mobile phone user about their use of their mobile phones and to teach them how to use a function on their phone. They had to assume that a mobile phone company, wanting to target the over-65 market, employed them.

The students were required to submit a joint report detailing their findings and making recommendations for improvement of the device and/or interface. We provided the framework for the interview to ensure that all students investigated the same aspects. This questionnaire addressed: background information about the participant, their mobile phone experience and usage, and how they felt about their mobile phones. No personal information was collected so anonymity was preserved. To familiarise students with ethical practice they were instructed to ask the participants to complete a consent form prior to the interview. We confirmed their findings by carrying out a search of relevant academic publications reporting the findings of empirical studies in this area.

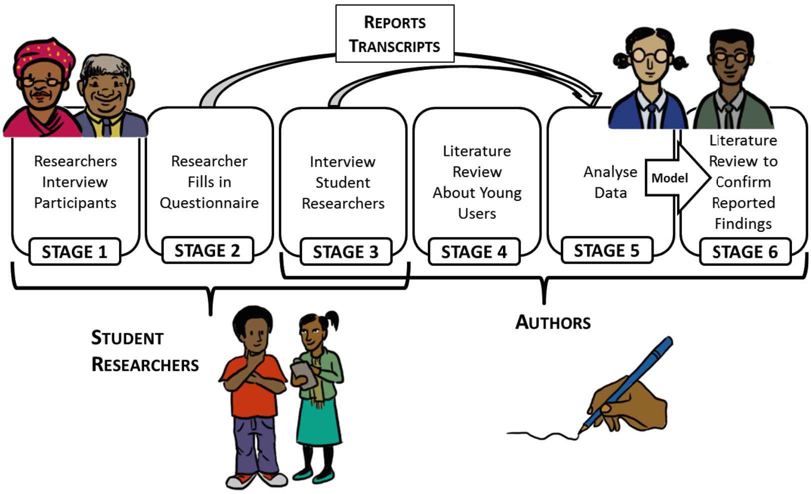


Fig. 3. Study Stages

In stage 3 we instructed the student researchers to also complete the questionnaire so that we could carry out a comparative analysis. We also interviewed the students to gather qualitative data to support stage 4. In stage 4 we consulted the academic literature to establish an understanding of their typical attitudes to, and use of, mobile phones. During stages 5 and 6 the reports, questionnaires and interview data were analysed qualitatively and quantitatively. The analysis of the qualitative data supported identification of similar or contradictory themes [19].

4 Stages 1 and 2: Student Researchers' Reports

Fifty student researchers collected data from 99 participants (aged 65 years and older). The students were asked to recommend changes to the interface design that could improve mobile phone usage. Case sampling was used to select the participants "where a profile of attributes or characteristics that are possessed by an 'average', typical person or case is identified, and the sample is selected from these typical people or cases" [19, p. 143]. To explore their participants' experience of their mobile phones, they interviewed the participants using a standard questionnaire of 65 questions developed by the authors. Students sometimes found it difficult to find willing participants for the project but a few students obtained permission to interview residents in retirement villages or homes. Those with relatives in the vicinity mostly interviewed family or family friends. Most student interviewers interviewed a friend or acquaintance (see **Fig. 4**).

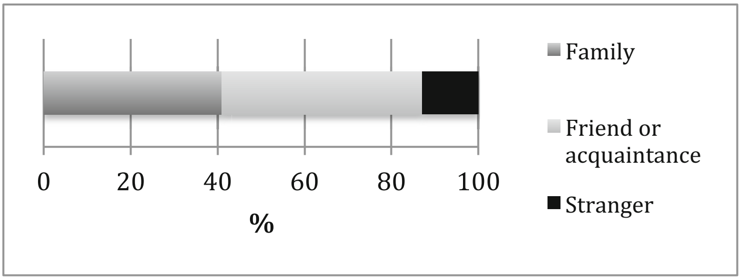


Fig. 4. Students' relationship to their participants

4.1 Questionnaire Results

The older participants were mostly unskilled (42%), although 37% were skilled, 11% were retired and 9% were unemployed. The majority (37%) of the participants had between 6 and 12 years of schooling (see **Fig. 5**).

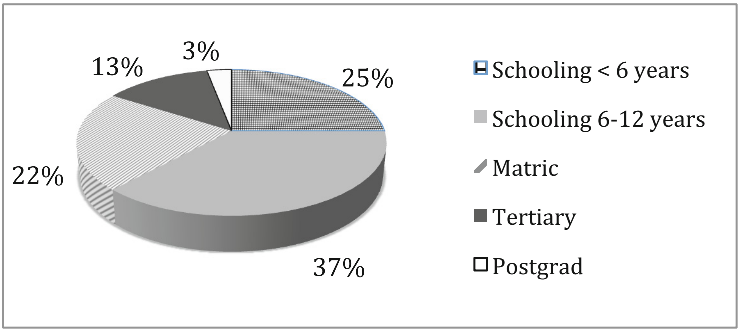


Fig. 5. Highest qualification of elderly participants

Significantly more students indicated that their English (reading and writing) skills were very good (X^2 = 30.95, p < 0.0001) (X^2 = 13.53, p= 0.0012). Students used their cell phones more regularly than the older generation (X^2 =18.58, p < 0.0001). Although 65% of the participants indicated that they enjoyed their phones, more students (97%) indicated that they enjoyed their phones (X^2 =15.93, p=0.0003). More students also indicated that they always kept their mobile phones within reach (X^2 =10.78, p=0.0010). As expected, students are more likely to use abbreviations when writing SMSs (85%) (X^2=51.70, p < 0.0001). Students were more likely to have smart phones (91% versus 31%) than their participants (X^2=45.50, p < 0.0001). Significantly more students found the mobile phone easy to operate, enjoyable, appealing and friendly compared to their participants.

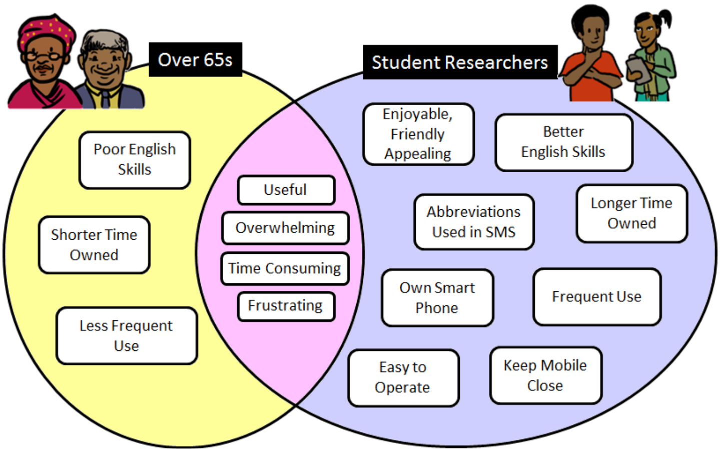


Fig. 6. Similarities and differences

A statistically significant number of participants found the mobile phone confusing at times. Phones were obtained from similar sources and charged similarly.

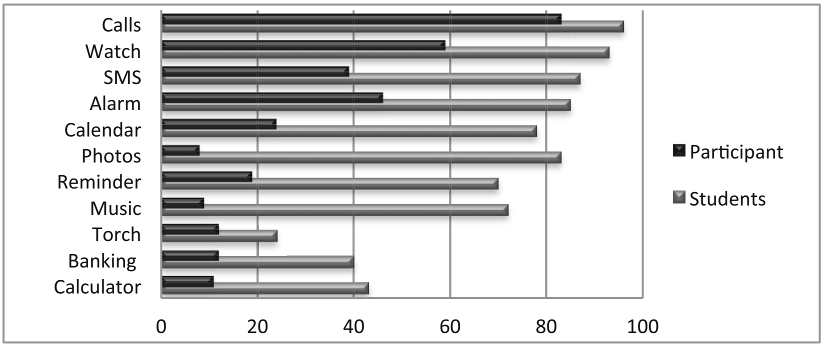


Fig. 7. Phone functions used indicated as a percentage

No differences were found between the two groups in terms of their feelings of frustration when operating the phone or it being overwhelming or time consuming. A summary of all the findings is depicted in **Fig. 6** and can be seen in Table 1 in Appendix A.

Phone usage, as expected, was very different in the two study groups – all functions, except calls, were used statistically significantly more regularly by students than the elderly participants (see **Fig. 7**). Both groups found the mobile essential, useful, liked the phone, indicated that it made life easier and that it offered value for money (see **Fig. 8**). Similarly both groups indicated that there were some aspects of their mobile phones that they would like changed/improved. These included: the keyboard, the screen, the camera and other aspects such as sound quality and increased battery life.

5 Stages 3 and 4: Literature on the Younger Mobile Phone User

The young seem totally at ease with their mobile phones, almost considering the device to be an extension of themselves. To understand how the young feel about their phones we asked students to fill in the same questionnaires the participants completed. From this, a number of themes emerged, related to how they felt about their phones. The research literature confirmed these themes (see **Fig. 9**):

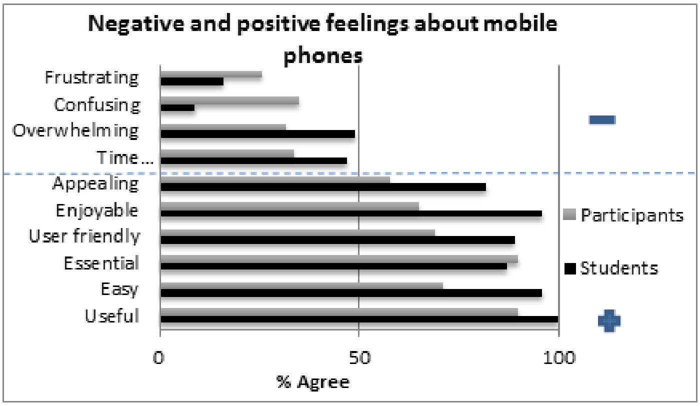


Fig. 8. How participants and students felt about their mobile phone

1. **Use Frequently** & **Keep their phones close**: Walsh, White and Young found evidence that the young are extremely attached to their phones, exhibiting symptoms of addiction [20]. They use their phones to enliven "dead time" while commuting or queuing, to maintain a sense of awareness of others not in the same location and use the phones to finalise arrangements for meetings in real time [21]. Both Park and Wikle argue that some users become addicted to their mobile phones [22] [23]. If this happens, the device will easily become characterized as **time-consuming.**

2. **Enjoyable, friendly, appealing, useful:** in terms of these aspects it is difficult to know which comes first, the mobile or the prospect of enjoyment. Dickinger, Arami and Meyer suggest that the prospect of enjoyment drives adoption [24]. Once they start using the phone, it is likely that their enjoyment will lead to more usage, with enjoyment and usage reinforcing each other, leading to whole-hearted adoption [25].
3. **Frustrating, overwhelming:** This was an unexpected result: that even the young sometimes experience negative feelings towards their phones. Thompson, Hamilton and Rust refers to *feature fatigue* being experienced by mobile phone users, which might be what these students are experiencing [26]. This aspect is confirmed by Head and Ziolkowski [9].
4. **Emotional Attachment:** It is interesting to note that young people consider their phone as a way of expressing their sense of self and use it as a fashion statement [27]. Vincent investigated emotional attachment to mobile phones, and suggests that its multiple roles and the activities it facilitates play a role in this attachment [28].

Fig. 9. Young people and their mobile phones

Many of these findings were also confirmed by the Pew Internet survey of mobile use by teens in April 2010 [29]. What is interesting about this collection of themes is the apparent contradiction between approbation and opprobrium. The students rely on and use their phones frequently, but some still express a sense of being overwhelmed by their phones and admit to a measure of frustration. Nafus and Tracy carried out a longitudinal study into the use of mobile phones in the UK. They found that teenagers in their study were very enthusiastic about their phones, but that attitudes changed after 20, with some participants expressing outright hatred of their phones. They acknowledged the usefulness of the phones but had reservations about their negative social consequences. The authors postulate that the ubiquity of phones challenges their emerging sense of individuality. They report that the over-30s in their study appeared indifferent to their phones. It could be that our students, in the 20-30 age group, are in the process of shifting away from unqualified enthusiasm to a more mediated utility-oriented usage of their phones [30].

6 Stages 5 and 6: The Younger Generation's Reports about the Older Mobile Phone User

Team reports, as well as interviews, were used to gather qualitative data. A grounded theory approach was used to analyse the reports: no pre-conceived themes were used to guide the analysis: the analysis allowed a number of dominant themes to emerge from the text. All comments referring to the elderly and the challenges they faced in using and learning mobile phone functions were extracted, and categorised.

The themes that emerged suggest a pipeline-like filtering process between the phone, its myriad capabilities and functionalities, and the elderly user. It became clear that reduced functionality was a very strong theme. Some comments suggested that this was affected by the complexity of the interfaces but for some functions the strength of the need for use of the functionality provided sufficient impetus for mastery of the function, despite difficulties. Another strong theme was that of age-related impairment. All reports mentioned at least one of visual, auditory or dexterity impairment, which made the use of the phone challenging for the elderly participants. Some also referred to age-related memory difficulties. Another anecdotal theme was related to the perception that the participants appeared to learn more slowly than the researchers did themselves. There was the sense that this was impeding their use of the mobile device. There was also mention of education, gender and location effects on usage. A final strong crosscutting theme was that of change resistance and mistrust, with many of the participants being averse to changing the way they did things. Fig. 10 presents the themes as funnels, or filters, which intrude between the phone and its owner, making the device less useful and usable than it could be.

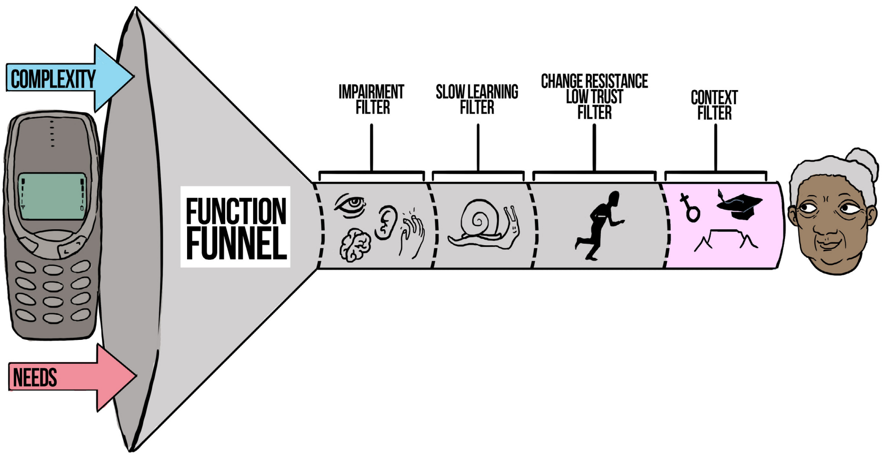


Fig. 10. Factors impacting on older mobile phone users

There is conflicting evidence in the literature about the impact of education on usage, such that we also cannot confirm that education, by itself, impacts older mobile phone users' use of their mobile phones [31] [32]. The former study detected no educational effect while the latter did. Bagchia, Kirsa and López [33] also did not find an educational effect on ICT diffusion, yet Lee did observe an educational effect [34]. Finally, Boateng carried out a study of micro-traders in Ghana and was not able to show any educational effect in terms of mobile phone usage [35]. It seems that one cannot reliably isolate the effects of education in impacting technology usage. Education, or lack thereof, like gender, is also a strong indicator of socio-economic status in Africa. Interestingly, the students also referred to a location effect. When the students referred to “location” they were probably referring to the areas participants lived in that reliably (in South Africa) serve as an indicator of their socio-economic standing. Rice & Katz were indeed able to confirm that this impacts mobile phone usage [36]. This, too, is confirmed by the Unicef report on the use of mobile phones in South Africa. Hence the gender, location and educational effects reported by the students arguably all fall under the socio-economic status umbrella, and this has indeed been shown to impact on mobile phone usage.

There is clear evidence confirming the other factors reported by the student researchers. There is confirmatory evidence of older mobile phone users making use of a small subset of core functionality [37], and that their usage is determined by their needs [38]. Turner, Love and Howell found evidence that older users did indeed use their phones less than younger users [39]. Ziefle and Bay found evidence that the complexity of the provided functions often discouraged older users from using them [40]. There is also evidence of age-related impairments getting in the way of their use of the phones [41] and there is evidence that they do indeed appear to learn more slowly than their younger counterparts (Holzinger, Searle & Nischelwitzer, 2007). They also resist change [42] and many mistrust technology [43].

Hence most of the identified factors are confirmed, suggesting that the students did indeed reliably identify age-specific difficulties in the use of mobile phones. The refined model is shown in **Fig. 11**. The ordering of the filters is not intended to be relevant or prescriptive. The relative impact and ordering will be individual-specific.

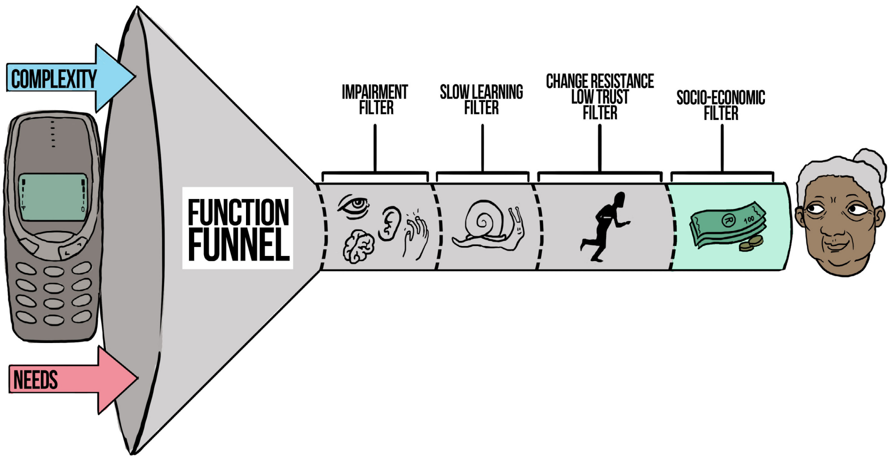


Fig. 11. A model of the factors that influence mobile phone usage by older users

7 Discussion

There are striking differences between young and old in terms of their use of, and attitudes towards, mobile phones. Whereas the older generation appear to struggle with the phone interface and multiple functionalities, what is conspicuously absent from the literature is any mention of difficulties being experienced by young people using their phones. There were no reports of perceived complexity or any sense of their being put off by the speed of technology advance – they embrace the new opportunities these advances open up. Kreutzer carried out a study of the use of mobile phone features by high school children in the same part of South Africa as this study. He reports that, during one particular day, almost all students used phones, even those who did not own phones themselves. The most used aspect was personal communication, but the second most used functionality was related to entertainment (games, music, photos, videos). This type of use is very rare amongst the older participants [44].

The older generation, on the other hand, were cautious, mistrustful, almost censorious. They owned and used their phones but there was no sense that they enjoyed them the way the young did. They used them purely in a goal-directed fashion to satisfy their needs. Fernández refers to mobile phones having a peripheral position in the older person's life, which is very different from the way young people view their mobile phones [45].

Yet there were similarities between generations. Both groups found the mobile phone useful, suggesting that it supported their goal-directed needs. Conci, Pianesi and Zancanaro surveyed 740 older people and found that their mobile phone usage was primarily utilitarian driven, which appears to confirm this. The motives that drive older users to use mobile phones, they argue, are much the same as those that drive younger users [46].

We will conclude this section with a note of caution. It would be foolhardy to assume a high level of homogeneity within island groups. For example, Aoki and Downs analysed college students' use of their mobile phones and identified five groups within their sample, displaying very different attitudes towards their mobile phones even within a group of fairly similar age, interest and educational level. In designing for "islands" we have to ensure that we identify the shared characteristics, and design to accommodate those, but the design must also be sufficiently flexible to accommodate the wide diversity *within* the generational island [47].

8 Conclusion

Daniel and co-researchers argue strongly for an understanding of context (social, economic and political dynamics) of ICT usage before any intervention is contemplated [48]. Here we have studied the use of mobile phones by the older population in South Africa, precisely in order to understand this context. We contemplated two approaches in this paper: designing for *one-world* or an *island-specific* design. We were able to depict the experiences (**Fig. 7**) and attitudes (**Fig. 8**) of the young researchers and the older participants and the differences are prominent and conspicuous. What we see is the older users grappling with a new and unfamiliar technology that often seems to overwhelm them with its complexity. We see usage of a reduced range of functionality and a measure of frustration at an inability to master this device that seems to mean so much to the younger generation. The young, on the other hand, appear to love their phones, use them constantly and seemingly couldn't do without them. Would it be possible to produce one design for these two groups? We do not believe this to be feasible or practical. In terms of age-sensitive mobile phone design, especially in a developing world context, we therefore argue that island-specific design is the wiser approach.

References

[1] Perry, T.S., Geppert, L.: Do portable electronics endanger flight? The evidence mounts. Spectrum 33(9), 26–33 (1996)

[2] Eddy, N.: FCC asks FAA to allow broader in-flight device use (December 09, 2012), `http://www.eweek.com/mobile/fcc-asks-faa-to-allow-broader-in-flight-device-use/?kc=EWWHNEMNL12122012EOADKC=` (accessed December 12 2012)

[3] Tapscott, D.: Growing up digital: the rise of the net generation. McGraw-Hill, New York (1997)

[4] Gelderblom, J.H., Van Dyk, T.J., Van Biljon, J.A.: Mobile phone adoption: do existing models adequtely capture the actual usage of older adults? In: Proceedings of the SIACSIT 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (2010)

[5] Neves, B.B., Amaro, F.: Too old for technology? How the elderly of Lisbon use and perceive technology. The Journal of Community Informatics. Special Issue: Community Informatics and Older Persons 8(1), 495 (2012)

[6] Campbell, M.: The impact of the mobile phone on young people's social life, Centre for social change research (2005)

[7] Salkowitz, R.: Generation blend: managing across the technology age gap. John Wiley & sons (2008)

[8] Chattratichart, J., Brodie, J.: Envisioning a mobile phone for all ages. In: Interact 2003, Zurich (2003)

[9] Head, M., Ziolkowski, N.: Understanding student attitudes of mobile phone features: Rethinking adoption through conjoint, cluster and SEM analysis. Computers in Human Behavior 28(6), 2331–2339 (2012)

[10] Prensky, M.: Listen to the natives. Educational Leadership 63(4), 8–13 (2006)

[11] Bayne, S., Ross, J.: The "digital native" and "digital immigrant": a dangerous opposition. In: Annual Conference of the Society for Research into Higher Education (SRHE), Brighton (2007)

[12] Cuddy, A.J., Fiske, S.T.: Doddering, but dear: Process, content and function in stereotyping of elderly people. In: Ageism, pp. 3–26. MIT Press, Cambridge (2002)

[13] Caspi, A.: Contact hypothesis and inter-age attitudes: A field study of cross-age contact. Social Psychology Quarterly 47(1), 74–80 (1984)

[14] Erber, J.T., Etheart, M.E., Szuchman, L.T.: Age and forgetfulness: Perceivers' impressions of targets' capabilites. Psychology and Aging 7(3), 479–483 (1992)

[15] Neugarten, B.L., Moore, J.W., Lowe, J.C.: Age norms, age constraints, and adult socialization. American Journal of Sociology 70(6), 710–717 (1965)
[16] Ariely, D., Norton, M.: How actions create - not just reveal - preferences. Trends in Cognitive Sciences 12(1), 13–16 (2008)
[17] Netz, Y., Ben-Sira, D.: Attitudes of young people, adults, and older adults from three-generation families toward the concepts "ideal person", "youth", "adult" and "old person". Educational Gerontology 19(7), 607–621 (1993)
[18] Sharps, M.J., Price, J.L.: Attitudes of young adults toward older adults: evidence from the United States and Thailand. Educational Gerontology 24(7), 655–660 (1998)
[19] Cohen, L., Manion, L., Morrison, K.: Research methods in Education, 5th edn. RoutledgeFalmer, London (2001)
[20] Walsh, S.P., White, K.M., Young, R.M.: Over-connected? A qualitative exploration f the relationship between Aistralian youth and their mobilephones. Journal of Adolescence 31(1), 77–92 (2008)
[21] Ito, M., Okabe, D.: Technosocial situations: Emergent structurings of mobile e-mail use. In: Ito, M., Okabe, D., Matsuda, M. (eds.) Peronal, Protable, Pedestrian: Mobile Phones in Japanese Life, pp. 257–273. MIT Press, Cambridge (2005)
[22] Park, W.K.: Mobile phone addiction. In: Ling, R., Pedersen, P.E. (eds.) Mobile Communicaitons: Re-negotiation of the Social Sphere, pp. 253–272. Springer, London (2005)
[23] Wikle, T.A.: America's cellular telephone obsession: New geographies of personal communication. Journal of American and Comparative Cultures 24(1/2), 123–128 (2001)
[24] Dickenger, A., Arami, M., Meyer, D.: The role of perceived enjoyment and social norm in the adoption of technology with network externalities. European Journal of Information Systems 17, 4–11 (2008)
[25] Turel, O., Serenko, A., Bontis, N.: User acceptance of hedonic digital artifacts: A throey of consumption values perspective. Information and Management 47(1), 53–59 (2010)
[26] Thompson, D.V., Hamilton, R.W., Rust, R.T.: Feature fatigue: when product capabilities become too much of a good thing. Journal of Marketing Research XLII, 431–442 (2005)
[27] Katz, J.E., Sugiyama, S.: Mobile phones as fashion statements: The co-creation of mobile communication's public meaning. Mobile Communicaitons: Computer Supported Cooperative Work 31, 63–81 (2005)
[28] Vincent, J.: Emotional attachment to mobile phones: An extraordinary relationship. Mobile World: Computer Supported Cooperative Work 2, 93–104 (2005)
[29] Lenhart, A., Ling, R., Campbell, S., Purcell, K.: Teens and mobile phones (April 20, 2010), `http://pewinternet.org/Reports/2010/Teens-and-Mobile-Phones.aspx` (accessed December 12, 2012)
[30] Nafus, D., Tracey, K.: Mobile phone consumption and concept of personhood. In: Katz, J.E., Aakhus, M. (eds.) Perpetual Contact: Mobile Communicaitons, Provate Talk, Public Performance, pp. 206–220. Cambridge University Press, Cambridge (2002)
[31] Rasanen, P., Kouvo, A.: Linked or divided by the Web? Internet use and sociability in four European countries. Information, Communication & Society 10(2), 219–241 (2007)
[32] Gutierrez, L.H., Gamboa, L.F.: An approximation to the digital divide among low income people in Colombia, Mexico and Peru: Two composit indexes. Universidad del Rossario, Bogota (2008)
[33] Bagchia, K., Kirsa, P., Lopez, F.: The impact of price decreases on telephone and cell phone diffusion. Infomrtion & Managment 45(3), 183–193 (2008)
[34] Lee, J.-W.: Education for technology readiness: prospects for developing countries. Journal of Human Development 2(1), 115–151 (2001)

[35] Boateng, R.: Mobile phones and micro-trading activities - conceptualizing the link. Info 13(5), 48–62 (2011)
[36] Rice, R.E., Katz, J.E.: Comparing Intrnet and mobile phone usage: digital divides of usage, adoption, and dropouts. Telecommunications Policy 27(8-9), 597–623 (2003)
[37] Renaud, K.V., Van Biljon, J.: Worth-centered mobile phone design for older users. Universal Access Information Society 9(4), 387–403 (2010)
[38] Kurniawan, S., Mahmud, M., Nugroho, Y.: A study of the use of mobile phones by older persons. In: Proceedings of CHI EA 2006 Extended Abstracts on Human Factors in Computing Systems, New York (2006)
[39] Turner, M., Love, S., Howell, M.: Understanding emotions experienced when using a mobile phone in public: the social usability of mobile (cellular) telephones. Telematics and Informatics 25(3), 201–215 (2008)
[40] Ziefle, M., Bay, S.: How older adults meet complexity: Aging effects on the usability of different mobile phones. Behaviour & Infomration Technology 24(5), 375–389 (2005)
[41] Renaud, K.V., van Biljon, J.: Predicting technology acceptnce and adoption by the elderly: A qualitative study. In: Proceedings of SAICSIT 2008 (2008)
[42] Gilly, M.C., Zeithaml, V.A.: The elderly consumer and adoption of technologie. Journal of Consumer Research 12, 353–357 (1985)
[43] Benamati, J., Serva, M.A.: Trust and distrust in online banking: The role in developing countries. Information Technology for Development 13(2), 161–175 (2007)
[44] Kreutzer, T.: Assessing cell phone usage in a South African township school. International Journal of Education and Delopment using ICT 5(5) (2009)
[45] Fernandez, M.: Interactions with and through mobile phones: what about the elderly population? In: Third European Communcations Conference, Hamburg (2010)
[46] Conci, M., Pianesi, F., Zancanaro, M.: Useful, social and enjoyable: mobile phone adoption by older people. In: Intract 2009, Uppsala (2009)
[47] Aoki, K., Downes, E.J.: An analysis of young people's use of and attitudes toward cell phones. Telematics and Informatics 20(4), 349–364 (2003)
[48] Daniel, M., Skuse, A., Slater, D., Tacchi, J., Chandola, T., Cousins, T., Horst, H., Kwamu, J.: Final Report. Information Society: Emergent technologies and development communities in the South (2005), http://94.126.106.9/r4d/PDF/Outputs/Mis_SPC/R8232FTR.pdf (accessed January 08, 2013)

Appendix A

Table 1. Elderly participant and student comparisons

Variable	Options	Participant	Student	X^2	p-value
Resident	City	65	46	4.94	0.0848
	Small town	21	36		
	Rural	14	18		
English reading skills	Not well	18	0	30.95	<0.0001**
	Reasonably well	42	11		
	Very well	40	89		
English writing skills	Not well	21	0	13.53	0.0012**
	Reasonably well	36	32		
	Very well	43	68		
Length of mobile ownership	< 1 year	6	2	8.94	0.0115*
	1-3 years	22	4		
	3+ years	72	94		
How was mobile obtained	Bought	49	56	3.20	0.3612
	Gift	43	33		
	Contract	6	11		
	Other	2	0		
How often is mobile used	< Once per week	11	0	18.58	<0.0001**
	Once a day	24	2		
	> Than once a day	63	98		
Keep mobile close	Yes	80	100	10.78	0.0010**
	No	20	0		
How mobile is charged	Home battery	17	15	3.17	0.3667
	Electricity	83	85		
Abbreviations used for SMS	Yes	21	85	51.70	<0.0001**
	No	79	15		
Phone enjoyable	Agree	65	96	15.93	0.0003**
	Not sure	23	2		
	Disagree	12	2		
Phone frustrating	Agree	26	16	2.34	0.3102
	Not sure	18	15		
	Disagree	56	69		
Makes life easier	Ambivalent	3	3	0.78	0.6761
	Yes	95	97		
	No	2	0		
Suggested changes to phone	Nothing	38	19	6.16	0.1879
	Keyboard	15	23		
	Screen	17	16		
	Camera	6	12		
	Other	24	30		
Own a smart phone	Yes	31	91	45.50	<0.0001**
	No	96	9		

Significant result $p < 0.05$; **Very significant $p<0.01$.

PointerPhone: Using Mobile Phones for Direct Pointing Interactions with Remote Displays

Julian Seifert[1], Andreas Bayer[2], and Enrico Rukzio[1]

[1] Ulm University, Institute of Media Informatics, James-Franck-Ring, 89081 Ulm, Germany
[2] University of Duisburg-Essen, Schützenbahn 70, 45127 Essen, Germany
{julian.seifert,enrico.rukzio}@uni-ulm.de, andreas.bayer@uni-due.de

Abstract. Large screens or projections in public and private settings have become part of our daily lives, as they enable the collaboration and presentation of information in many diverse ways. When discussing the shown information with other persons, we often point to a displayed object with our index finger or a laser pointer in order to talk about it. Although mobile phone-based interactions with remote screens have been investigated intensively in the last decade, none of them considered such direct pointing interactions for application in everyday tasks. In this paper, we present the concept and design space of PointerPhone which enables users to directly point at objects on a remote screen with their mobile phone and interact with them in a natural and seamless way. We detail the design space and distinguish three categories of interactions including low-level interactions using the mobile phone as a precise and fast pointing device, as well as an input and output device. We detail the category of widget-level interactions. Further, we demonstrate versatile high-level interaction techniques and show their application in a collaborative presentation scenario. Based on the results of a qualitative study, we provide design implications for application designs.

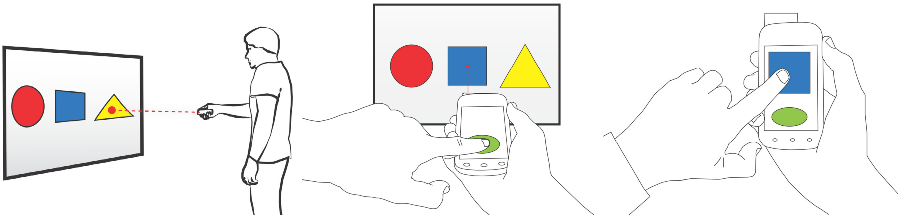
Keywords: Mobile phone, pointing, interaction, collaboration.

1 Introduction

Large screens and projections have become part of our daily lives. They support collaboration as they allow multiple users to simultaneously access information. For instance, in a meeting presentation, information is shared with multiple users. However, the control of what is being displayed is usually limited to a single user (e.g., in presentation). Others cannot share, access, or manipulate virtual objects or data on the remote display.

Direct pointing has been investigated for interaction with remote displays [19]. Direct pointing interaction is a natural way for users to select and interact with objects on a remote screen (see Figure 1(a)) [16]. However, such settings enable only few options for interactions and are limited to basic operations.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 18–35, 2013.



(a) (b) (c)

Fig. 1. Using the mobile phone as a pointing device enables versatile interactions with remote screens: (a) Pointing to targets. (b) Performing actions on the personal phone such as selecting and downloading an item. (c) Further interaction with data on the phone.

As mobile phones are ubiquitously available, they enable users to access remote displays in diverse ways [2]. For instance, downloading information from a remote display to the mobile phone for further inspection [4] or sharing information on a remote screen with others [1]. Pointing-based interactions offer an easy-to-use way of allowing users to interact with an object by pointing to it [24].

Hence, mobile phones with integrated pointing abilities enable diverse novel options for interaction in a natural and seamless way. Using the mobile phone as a pointing device and general interaction device has not been investigated, thus raising questions regarding how the phone's specific characteristics (e.g., options for input and output) and attributes (e.g., user data context) can be integrated into the interaction process.

In this paper, we contribute the detailed investigation of the novel design space of PointerPhone. PointerPhone uses mobile phones as pointing devices for new direct and natural pointing interactions with remote screens (see Figure 1). We present a classification consisting of low-level, widget-level, and high-level interaction techniques. Further, we show application examples and demonstrate the integration of diverse techniques into a collaborative meeting support application which we implemented based on a prototype system that uses laser-pointer-equipped mobile phones[1]. Furthermore, we present observations and results of a qualitative user study and we provide a catalogue of design guidelines as well as lessons learned that need to be considered when designing applications based on PointerPhone interactions.

2 Related Work

A large body of research exists in the area of remote interaction with remote information displays such as large screens or projections. These related works can be classified into the following categories: a) general pointing interaction and technologies, b) interaction with remote screens using mobile phones as a pointing device, and c) collaboration support. The review of the related work shows that no previous work investigated the combination of mobile phones with direct pointing for interaction with remote screens.

---

[1] A complementary video figure is available at `http://youtu.be/qp3pIklYLxo`

2.1 Pointing Interaction

Early work by Rekimoto investigated the potential of interacting with remote screens in a continuous work space [23]. Using the hyperdragging technique, users could move data items across multiple devices (e.g., laptops or shared screens). Here the user controlles a cursor with a mouse connected to one device. The cursor would *jump* to other screens when reaching a display border, allowing to reach out for data across devices. By using a mouse, the pointing is only indirect as users cannot literally point out to the desired target but rather move the mouse on a planar surface. Olsen and Nielsen investigated direct pointing for interaction with virtual objects based on laser pointers [19]. They analyzed interaction techniques based on camera tracking of a simple laser pointer on a remote screen (e.g., interaction with widgets, drawing, and navigation). Vogt et al. demonstrated how collaborative multi-user settings can be implemented using several laser pointers [31]. Myers et al. compared the effect of different form factors of pointing devices (a pen-like pointer, a pointer attached to a toy gun, and a pointer attached to a personal digital assistant device) for interaction with large projected screens [16]. They observed that it is essential to support interactions which do not cause much jittering of the pointing device and allow users to easily control the making of selections. Shizuki et al. investigated the potential of using pointers to interact with remote screens beyond making mouse-like selections [27]. Their findings show that simple commands for applications can be directly performed using a standard laser pointer device. Nintendo released the Wii U controller device which enables pointing to a remote screen and interaction through a small touchscreen [18]. Similarly to the PointerPhone approach, this system enables interaction on a remote and a personal screen. Yet, the Wii U game controller does not provide and store users' data such as photos, bookmarks or contacts. Also, a game controller is not always carried by the users and thus is not suited for interacting in diverse settings. Schmidt et al. introduced a pico projector-based approach where users project an interface onto smart objects that are equipped with a specific sensor [25]. Through moving the projector and thus the projected interface, users can select and activate different interface widgets. This way of pointing interaction is, however, limited to one user and does not support collaboration.

2.2 Pointing Technologies

A large variety of technologies and approaches exist to allow for the sensing of the pointing interaction. They can be classified into direct pointing (e.g., with laser pointers as in [19, 31, 29]), and inertial sensing (e.g., using gyroscope and accelerometer sensor data as in [18]). An additional approach was presented by Zizka et al.; they presented SpeckleSense which enables precise interactions with remote screens based on laser speckle motion sensing [34]. In contrast to a pointer-based approach, each interaction device handles the sensing internally. This approach also allows for the use of a fixed device which senses the motion of a user based on laser speckle; similar as interaction sensing enabled through a fix installed depth camera (e.g., [15]). With the XWand, Wilson and Shafer introduce a pointing device with the shape of a short stick, which allows users to point to and interact with objects in a smart environment [33]. This system determines with which object a user intends to interact based on data provided by different integrated sensors such as accelerometers. Hence, no information display is required and user feedback is provided implicitly.

2.3 Mobile Interaction with Remote Screens

For interacting with remote screens, mobile phones have been used in many different ways. For instance, different works examined the potential of gestures where the user pretends to throw an object towards the remote screen while holding the mobile phone in their hand (e.g., [8, 10]). Yet this only allows for imprecise interactions, and selecting items is not possible. Also, different television set manufacturers provide applications for mobile phones which replace the standard remote control: the mobile phone provides an interface for selecting the channel and controlling the volume (e.g., [12]). However, interaction is limited to these simple actions and to steering a cursor through the phone's touch screen. Boring et al. presented a unidirectional mobile phone camera-based approach for selecting and transferring data from a remote display to a personal mobile phone [4]. Different works investigated the potential to control a virtual cursor on the remote screen, using the phone camera by tracking the optical flow (e.g., [13]), or the phone's accelerometers (c.f., [6]). However, this approach is limited in terms of interaction speed. Boring et al. presented TouchProjector, which allows users to perform touch input on their mobile devices; the touch events are then transmitted to a remote screen [5]. Baur et al. investigated interaction based on the metaphor of virtual projection whereby the phone camera captures an external screen to sense the spatial relation of the phone to the screen. Users can exchange data and interact with virtual objects on the remote screen within the camera frustum by performing touch interactions on the mobile phone [3].

The PointerPhone approach presented in this paper, in contrast to the aforementioned works, allows to interact directly with specific targets on the remote screen by pointing in a natural way. Users do not have to hold the mobile phone like a camera as the case when using TouchProjector for remote touch interaction. Also, in contrast to camera-based approaches for steering a cursor, which only allow indirect pointing, PointerPhone allows pointing that is natural for the user, fast, and direct.

2.4 Mobile Phone as Pointing Device

Rukzio et al. investigated users' preferences of mobile interaction with smart objects in their environment [24]. They compared the interaction styles of touching (using near field communication), scanning (based on Bluetooth), and pointing. For pointing, they attached a laser pointer to a mobile phone for interacting with smart objects such as a CD player. Pan et al. investigated how inertial sensing of a phone's position and rotation in relation to a projected remote screen can be used to support pointing using a mobile phone [20]. In their work, Rashid et al. investigated the differences of proximal and distal selection of targets using a mobile phone connected to a Wii controller

[22]. They found that distal interaction is preferred for the selection of targets and results in lower error rates compared to proximal selection. While this work provides interesting insights regarding the selection of targets, they did not focus on other aspects such as interaction with widgets or data sharing interactions.

These works investigated the use of a mobile phone combined with different pointing capabilities. However, they all do not consider the phones' specific attributes and capabilities that allow for novel interactions. PointerPhone considers the characteristics of mobile phones in terms of available sensing, input and output options, and application possibilities which have not been investigated so far.

2.5 Collaboration Support

Work on single-display groupware [30] investigated how multiple users can simultaneously interact with a single display, for instance, by using multiple mouse cursors [21]. However, in such a setting user collaboration is limited to simultaneously pointing to targets on a shared screen. Sharing personal data is not possible, although feasible through the PointerPhone approach. Connecting personal and mobile devices to a remote shared screen to support collaboration has been investigated in different works. The combination of such different devices enables multiple users to access a shared device while each also controlling their personal device [23]. Myers et al. connected multiple PDAs to a central workstation [17], allowing for data exchange. Additionally, the combination of personal and mobile devices with horizontal tabletop-like displays has been studied [26, 28]. Vogt et al. compared mouse-based and laser pointer-based interaction regarding their ability to support collaboration in small groups [32].

3 Phone Pointer Interactions

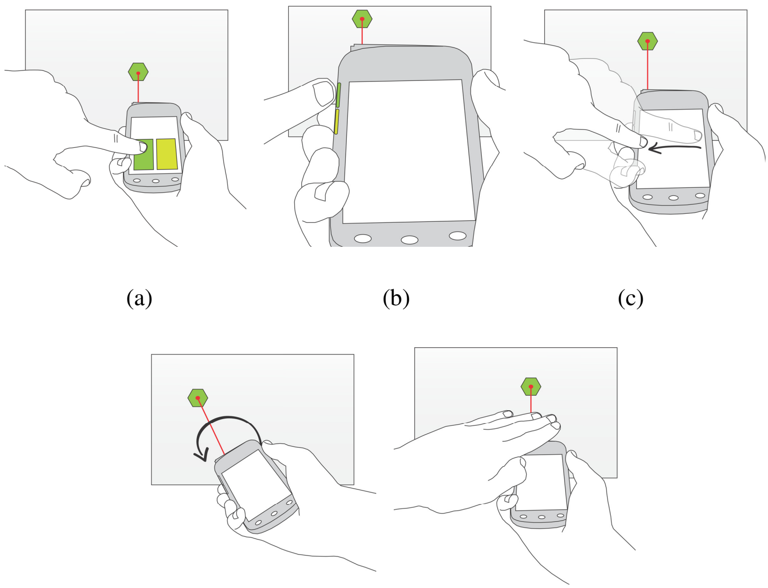
The underlying concept of using the mobile phone as a pointing device for direct interaction with remote displays is simple but at the same time versatile: users point towards targets on a remote screen in order to perform an action that is applied as the user triggers the action (e.g., selecting or editing an item, controlling widgets). The available hardware of the mobile phone and the remote display yield a number of basic attributes and possibilities for interaction. Basic attributes include whether the user is pointing and the location on the remote display where the user is pointing to. Further, each mobile device that is used as a pointing device can be distinguished through its ID. Accordingly, different users can be distinguished.

Each of these basic attributes can be used for interactions. Considering attributes isolated and combinations of different attributes result in interactions of diverse complexity. Hence, simple interactions can be used as building blocks for more complex interactions. For instance, selecting a target by *pointing* and *clicking* can be part of any activity such as selecting a file on the remote screen. Again, this activity can be part of many applications such as photo sharing or web browsing.

Accordingly, the diverse interaction possibilities can be classified into three layers of abstraction: low-level input and output, widget-level interactions, and high-level interaction and applications. This classification was chosen as it shows which options for interaction are available and how these can be integrated into application designs.

3.1 Low-Level Input and Output

Low-level Input Options. The most basic options for performing input on the mobile phone while pointing to a target on the remote display are using software buttons displayed on the phone's screen (see Figure 2(a)), using hardware buttons available on the phone's case (e.g., buttons commonly provided to control audio volume) (see Figure 2(b)), and performing gestures on the phone's touch screen (see Figure 2(c)). These options can be applied in flexible ways as they can be used either with one hand or with two hands.



(a) (b) (c)

(d) (e)

Fig. 2. Basic selection and input options supported through PointerPhone: (a) software buttons; (b) hardware buttons; (c) touch-based gestures; (d) rotation-based interaction; and (e) proximity activation.

One alternative option that avoids pressing hardware or software buttons is to trigger an action by rotating the phone along the pointing axis (see Figure 2(d)). Rotation in different directions (i.e., clockwise and counterclockwise) allows to encode different actions. For instance, a left and right click can be performed. However, rotating the phone could also result in moving the cursor away from the target.

As emphasized by Myers et al. (see [16]), using physical buttons causes unintended jitter effects which could lead to input actions on unintended targets. This is potentially also the case when interacting with software buttons or performing gestures on the mobile phone. One approach that enables users to trigger an action without touching and moving the mobile phone uses the proximity sensor in the mobile phone. While users point to a target, they move their hand close to the phone and trigger the action as their hand gets close enough (see Figure 2(e)). Similarly, users could trigger an action without moving or potentially even without touching the mobile phone through snapping with their available hand which could be sensed using the phone's microphone.

Low-level Output Options. Output options for feedback and information presentation are distributed on the users' mobile pointing device and the remote screen. The latter provides visual feedback and optionally, audio output can be provided (e.g., a television set in the user's living room remotely operated through pointing). The output options of the remote display can be targeted to one specific user only to a limited degree. If several users are using the system simultaneously, for instance, audio feedback provided by the remote screen is audible to all present persons. Visual feedback, however, can be displayed on the remote screen close to a user's pointing cursor in order to make clear at whom the feedback is targeted.

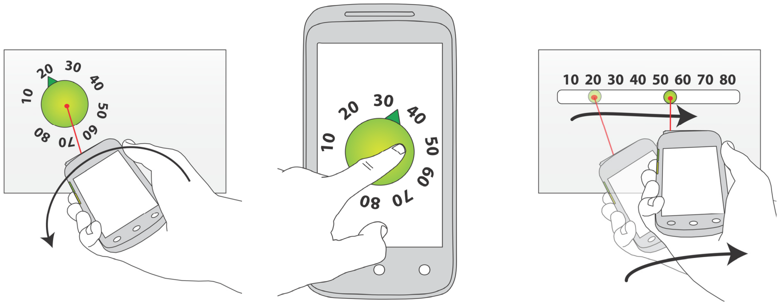
In addition, the personal mobile pointing device enables personal feedback which is not accessible to others. That is, the personal mobile devices provide visual feedback and output on their display, which is visible only to the specific user. Also, audio feedback can be provided either via speakers or headphones, which allow feedback that is not audible to others. Third, mobile devices allow for haptic feedback through vibration.

3.2 Widget-Level Interaction

Often, interaction with diverse applications requires users to specify specific pieces of information or data (e.g., numeric values, strings) in order to control the state of an application. For instance, users control the zoom level of text or specify the volume of audio data using a slider, or select an option from a list using radio buttons. To facilitate this task, many different widgets are available, including as sliders, radio buttons, and text fields, each of which supports the input of a specific data type.

When using mobile phones as pointing device, to interact with applications on a remote or shared display, users require support to interact with all kinds of standard user interface widgets. The given configuration yields up to three different options for interacting with widgets: 1) *rotation* of the mobile phone (see Figure 3(a)), 2) *proximal* interaction on the touch screen (see Figure 3(b)), and 3) *distal* interaction through direct pointing to the widget on the remote screen, selecting it and moving the pointing cursor to change the value of the widget (see Figure 3(c)).

As analyzed by Rashid et al. (see [22]), the performance of proximal and distal selection of targets (e.g., clicking a button) depends on the complexity of the tasks. For complex tasks which involve many and small targets, proximal interaction is superior to distal interaction which was superior in simple tasks (i.e., interaction with few large targets).



(a) (b) (c)

Fig. 3. Controlling widgets for entering data can be implemented in three ways: (a) Rotating the phone to change the value; (b) manipulating the proximal widget representation on the phone screen; (c) distal interaction with widgets on the remote screen.

However, not all three options for interacting with widgets apply to each widget, depending on the type of data supported. Table 1 offers an overview of standard widgets and how they can be controlled through pointing with a mobile phone.

Table 1. Overview of the widget control options

Widget	Orientation	Proximal	Distal
Turning Knob	Yes	Yes	Click & Drag
Sliders	Yes	Yes	Click & Drag
Button	Yes	Yes	Click
Radio Buttons	Yes	Yes	Click
Check Boxes	No	Yes	Click
Text Field	No	Yes	No

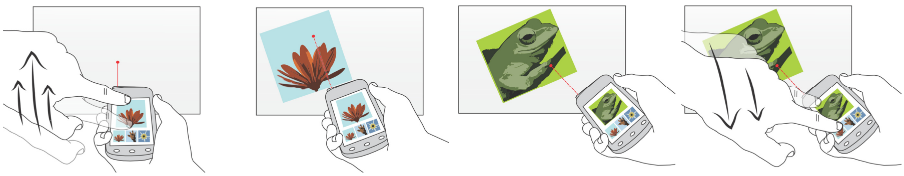
Accordingly, only widgets that are designed for the input of continuous values, that is, sliders or turning knobs, could be directly controlled via rotating the phone while pointing to them in order to change their value. Yet it is also possible to control these widgets proximally on the phone touch screen or distally on the remote screen. Standard buttons can be controlled using all three options, given that the phone rotation is mapped to a selection. Radio buttons could be selected and rotating the phone could change the selection. Check boxes are less suited for this alternative due to their size. Text fields require the user to interact with a keyboard which is most convenient for the user on the mobile phone.

3.3 High-Level Interactions and Applications

In this section, we discuss phone-based pointing interaction techniques that build on the previously discussed interactions.

Data Exchange. Users can share data that is stored on the mobile phone by transferring it to the remote screen. To do so, users select one or several items to share, point to a position on the screen, and trigger the transfer. For instance, users could select an image, point to the desired location, and perform a swipe gesture on the phone towards the remote screen (see Figure 4(a)). On the remote screen the image appears at the location of the pointing cursor (see Figure 4(b)).

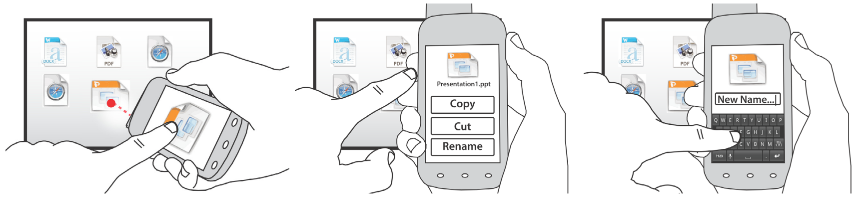
In order to receive data from the remote display, users point at the intended item (see Figure 4(c)) and trigger the transfer. As illustrated in Figure 4(d), a swipe gesture on the user's phone could be used to *pull* the item. However, any other low-level input can be applied here.



(a) (b) (c) (d)

Fig. 4. Transmitting an item from the mobile phone to the remote screen ((a) and (b)) and vice versa picking up an item from the screen ((c) and (d))

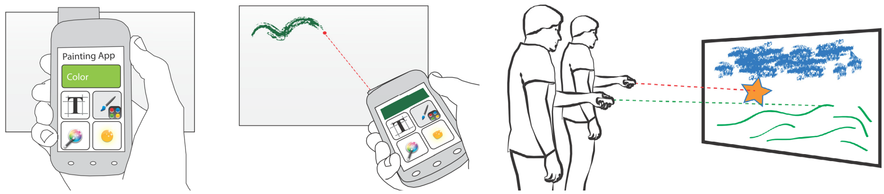
Proximal Context Menus. Pointing-based interaction with a remote display through a mobile phone supports the handling of meta information of items such as files which are displayed on the remote display. For instance, context menus are often used in order to change the name of a file. These provide a list of possible options that can be applied to the selected file. Using the mobile phone as pointing device, users first select a file (see Figure 5(a)). The corresponding context menu is then displayed on the mobile phone (see Figure 5(b)), thus for instance facilitating the input of a new file name (see Figure 5(c)). Users are not required to keep pointing at the selected file while using the context menu.



(a) (b) (c)

Fig. 5. Proximal context menus: (a) A user selects a file. (b) The context menu is displayed on the phone. (c) A new file name can be typed in.

Drawing and Sketching. Through pointing to the remote screen, users can edit and create graphical content such as sketches. Depending on a selected tool and the corresponding parameter settings (e.g., a brush and a selected color) users can create sketches simultaneously (see Figure 6). As different phones are distinguished, each user can select different tools and settings at the same time.



(a) (b) (c)

Fig. 6. A sketching application. (a) The phone provides a palette of different tools. (b) Tools are applied through pointing. (c) Multiple users can work simultaneously.

Personal Input and Output. Personal mobile phones as pointing devices allow users to receive personal output (e.g., visual or auditory) as well as to perform input on the personal device in collaborative settings. For instance, when multiple users share a view on a web page, a single user who is interested in following a specific link can point to it and open the corresponding web page on their personal device (see Figure 7(a) and (b)). This allows users to look up additional information without interrupting or disturbing the group activity.

(a) (b) (c)

Fig. 7. Input and output on the personal device: (a) To avoid disturbing a group activity, the user may point to a link and (b) open it on their device. (c) Entering information on the personal device.

Additionally, input can be performed on the personal device which, on the one hand, avoids cluttering the remote screen with a large virtual keyboard. On the other hand, input on the personal device allows users to enter sensitive information such as a password. For instance, when a user needs to login to a user account to access some information, the password can be entered on the mobile phone (see Figure 7(c)).

In addition, different types of data such as files, geographical coordinates, contact cards, or appointments can be distinguished and, once selected through pointing to their representation on the remote screen, they can be handled with different applications on the mobile phone. For instance, a user could point to an address and select it, which opens the map application on the mobile phone and displays the given location.

Remote Control. As more television sets support additional diverse applications such as web browsers, one emerging idea is to use secondary display devices to achieve remote control [7, 9, 22]. Using the personal mobile phone for this kind of interaction allows multiple users to interact simultaneously, for instance with web pages displayed on a smart TV (see Figure 8).

(a) (b)

Fig. 8. Using the mobile phone as a remote control for browsing web pages on a TV set

3.4 Collaborative Meeting Scenario

To show how different interaction options can be integrated and used as building blocks for a realistic application context, we designed and implemented a collaborative presentation system that supports users in a meeting scenario.

Scenario: Bob is giving a presentation for some colleagues on a large projected remote screen (see Figure 9(a)). Each meeting participant is equipped with a mobile phone which can be used to point to the remote display and control a curser through phone pointing interaction techniques. Each participant's pointing cursor on the remote screen is distinguished by a different color. Hence, each meeting participant has a visual representation of who and how many users are currently pointing to specific pieces of information on the remote screen. Each projected presentation slide contains diverse pieces of information. For instance, an overview plan may allow specific views on details of the plan through the selection of a corresponding icon (see Figure 9(b)). This allows users to individually explore and access additional information without disturbing others, as pointing to an icon on the remote screen and selecting it results in a detailed preview on their personal mobile phone. Icons next to a person's name indicate that the contact card can be downloaded to the mobile phone by pointing to it and selecting it. Additional background information can easily be accessed, for example, by pointing to an image on the remote display (see Figure 9(c)). In their meeting, the participants also discuss with each other about the presented topic. This discussion and brainstorming is supported through collaborative sketching on a drawing canvas on the remote screen where to each user can contribute using their phones (see Figure 9(d)).

(a) (b) (c) (d)

Fig. 9. Collaborative presentation and meeting support (a). Additional information can be accessed through specific icons ((b) and (c)). Users can share data with others on the remote screen (d).

A video illustrates this scenario in detail: `http://youtu.be/qp3pIklYLxo`

4 Evaluation

In order to gain an understanding on how users would use the system and how they appreciate the different interaction techniques, we performed a qualitative evaluation. The aim was to gain qualitative insights regarding direct pointing-based interaction with a mobile phone and remote display.

4.1 Design

The study session consisted of two parts: After participants were introduced to the study, they performed a series of practical tasks. Tasks that involved collaboration were performed by the participant together with the investigator. After the practical tasks, they filled out a questionnaire. During the session, participants were encouraged to think aloud and continuously talk about their actions. Further, the investigator took notes and the task performance was recorded on video.

Practical tasks. Practical tasks were selected to expose participants to a broad variety of different application contexts. Participants used the PointerPhone prototype for the tasks. The following list of tasks was performed by participants in randomized order.

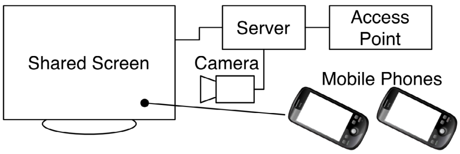
1. *Browsing*. Participants had to browse through a website that was displayed on the remote screen. The PointerPhone prototype was used for controlling widgets and link selection. They followed text-based instructions, which involved selecting links, downloading images to the phone, and interacting with widgets such as radio buttons through pointing.
2. *Photo sharing*. Selecting and transferring two photos from the phone's library to the remote display and retrieving photos from the remote screen.
3. *Sketching*. Collaborative sketching of a simple building on a shared sketching canvas on the remote screen which required different brushes (selected and configured on the mobile phone). Users also performed text entry on the phone and placing the created text on the sketching canvas for labeling the sketched items.
4. *Completing a form*. Filling in a form on the remote display which included interacting with different kinds of widgets for data input via the mobile phone.
5. *Context Menu*. Renaming, copying, and deleting files displayed on the remote screen by using a proximal context menu on the mobile phone.
6. *Playing*. Playing a simple Pong-like game involving two users who would steer the position and angle of a racket by pointing to the screen to control the translation and rotating the phone to control the angle.

Apparatus. The main components of the apparatus system are mobile phones as pointing devices and a remote display that is connected to a server computer (Figure 10(a)). The mobile devices are connected to the server through Wi-Fi for the exchange of data and commands.

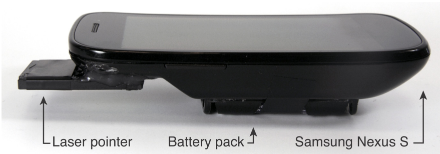
In order to achieve high-level pointing accuracy and low latency, we followed the approach of using laser pointers and camera-based tracking for the pointing task (as previously demonstrated by [16, 19]). That is, a camera is used to capture the remote display. If a user points to the remote screen, the laser pointer creates a bright point on the image which can be extracted through simple image processing. The location of the laser pointer is used to control the user's pointing cursor that is displayed on the remote screen. For distinguishing different laser pointers (and thus different users), a color filter is applied during the image processing step. This tracking approach requires only one calibration sequence before using the system until the camera or display setup is changed (i.e., if they are moved).

We extended a standard mobile phone (a Samsung Nexus S, running Android 2.3) with a laser-pointing module which can be controlled via the mobile phone's software stack (see Figure 10(b)). The laser pointer is turned on and off via a simple circuit with a photodiode that is placed right in front of the flash light LED of the mobile phone. On the Android platform [11], this component can be controlled via a standard application programming interface.

Based on the system prototype, we developed a number of mobile applications for the mobile phone and a corresponding application for the remote screen which implemented all functionalities that were required for the tasks. To allow participants to experience several possibilities of the PointerPhone interaction, the applications provide different options to perform any single action. For instance, users can make selections distally on the remote display, as well as proximally on the phone display. For the sake of consistency, the activation of the pointing (turning on the laser pointer) is the same in all applications. Short tapping on the hardware button on the bottom right activates permanent pointing. Holding the button activates the pointing until its release.



(a)



(b)

Fig. 10. (a): System components schema. (b): Mobile phone prototype.

Participants. We recruited 14 participants (7 female) aged between 20 and 31 (Mean=26). Of these, 10 were students (diverse backgrounds) and 4 were employees. After the study, they were rewarded for their effort with 10 Euro.

4.2 Observations and Design Implications

In the following, we discuss the results of the feedback sessions and, where possible, summarize findings as a set of design implications that support application designers when considering pointing interactions for their work.

1. *One- and two-handed interaction.* We observed that, during their interaction with the different applications, participants switched between using one or two hands to hold the mobile phone depending on the task. For instance, during the sketching task, 7 of the 14 participants held the phone with two hands. An additional 4 participants held the phone in the right hand and supported it with the left. Only 3 participants used one hand to hold the phone and point during the sketching task. In contrast, during the data-sharing task, 13 participants used a single hand to hold the phone, of whom 5 used the thumb of the same hand and 8 used the index finger of their other hand to interact with the phone interface. Accordingly, the manner in which users choose to hold the phone depends on the given task. For tasks that require precise pointing input (e.g., sketching), users tend to use two hands. Tasks that require less precise pointing, however, lead users to prefer one-handed operation. Hence, interfaces for the hand-held pointing device should encourage two-handed interaction when designed for tasks that require precise pointing. Inversely, interfaces for simple tasks should be adapted to one-handed usage.

2. *Selecting targets.* During the browsing task, participants could select targets such as links either distally on the remote screen or proximally on the phone's display. We observed that participants preferred to select large targets distally while small targets (e.g., text links) led to a preference to select proximally. This concurs with the findings of Rashid et al. who investigated distal and proximal target selection [22]. We observed that distal selection forced users to switch their focus from the remote display to the phone to ensure hitting the correct button. Accordingly, user interfaces for proximal interaction should be designed to allow users to keep their focus on the remote display – for instance, through the use of hardware buttons on the phone if available or a single large software button, so that the user does not have to look at the pointing device.
3. *Navigating.* When users selected an area of the remote screen that should be displayed proximal on the phone, several participants tried to interact with the proximal representation like they were used to interact with smartphone web browsers that allow navigation through dragging and zooming in and out. However, this applies only to small adjustments. Participants expressed that they can select easily an area through pointing to it which is more comfortable than navigating on the phone screen.
4. *Providing output and feedback.* We observed several times that users would focus on the remote display while they select a target there, yet they were not aware of the resulting change on the mobile phone. Inversely, this phenomenon was also observed when a user focused on the phone and performed an action which resulted in an event on the remote display. Hence, it is essential to provide cues (e.g., audio feedback or vibration) which notify users regarding resulting actions. Several participants raised the general point that the remote display should not be used to display user-specific information that is not intended for all users. For instance, when interacting with a web page, users could display tool tip information through pointing at an item for a few seconds. These should be displayed on the personal device.
5. *Controlling pointing actions.* Participants had to manually enable or disable the pointing mode through toggling a button (i.e., turning on and off the attached laser pointer). However, when a participant was engaged with performing a task, they forgot to turn the pointer on which resulted in confusion. Hence, if application workflows allow to anticipate when pointing is required, the system should automatically do so. For instance, when pointing to a web page on the remote display in order to transfer a clipping to the phone for further inspection, the pointing should be disabled automatically to prevent unintended updates.

All users indicated that they liked that they could see a cursor where they were pointing at. This indicated that alternative implementations (e.g., inertial sensing [20]) should provide a visual cursor throughout the interaction.

5 Discussion and Conclusion

In this work, we analyzed options for interaction when using a mobile phone as a pointing device for interaction with a remote display. Related work on pointing interaction, the use of mobile phones for controlling content on public and shared remote displays, as well as collaboration provides a large body of research on interaction techniques for specific contexts. Using the mobile phone as a personal pointing device provides not only powerful computing and sensing technologies but also the user's personal data context such as photos, calendars, and messages. These rich options for interaction are likely to attract the design of applications in the future. Hence, application designers considering PointerPhone applications should be supported through a design space and corresponding guidelines for using aspects from the design space.

Different options exist regarding possible implementations for the application of phone pointing-based interaction outside the laboratory setting. Using inertial sensing to determine the phone's pointing direction is a promising approach, as most available smart phones are equipped with the required sensors (i.e., accelerometer and gyroscope). However, each time a user intends to use such a system, the user needs to calibrate their phone to determine the pointing direction. Moreover, during the interaction the calibration may have to be repeated to maintain the pointing accuracy. Alternatively, direct pointing using a laser pointer can be easily added to standard smart phones. For instance, laser pointers can be plugged into the audio jack of the phone [29] and operated through an application. This approach would require the remote screen to be equipped with camera tracking. As an alternative, the laser pointer could be based on infrared light which can be sensed by specific screens (e.g., Microsoft PixelSense [14])

The investigation of the design space provides the list of basic attributes and characteristics. Further, we provide a classification of low-level interaction options regarding input and output options, widget-level interaction techniques, and high-level interaction techniques and applications. This classification into three levels of abstraction is not fixed and can be extended as it includes only selected examples for applications, and technological features for input are likely to be extended. These techniques can be used as building blocks for complex applications.

Finally, we showed how several interaction techniques can be integrated into a presentation application for a collaborative meeting context. We used a prototype implementation of the system based on mobile phones combined with laser pointers to realize a number of applications that we used for a qualitative study. Results from the study support a collection of five design recommendations that should be considered for the design of pointing-based applications.

Acknowledgements. This work has been conducted in the Emmy Noether research group Mobile Interaction with Pervasive User Interfaces funded by the German Research Foundation.

References

1. Alt, F., Kubitza, T., Bial, D., Zaidan, F., Ortel, M., Zurmaar, B., Lewen, T., Shirazi, A.S., Schmidt, A.: Digifieds: insights into deploying digital public notice areas in the wild. In: MUM 2011, pp. 165–174. ACM, New York (2011)
2. Ballagas, R., Borchers, J., Rohs, M., Sheridan, J.: The smart phone: a ubiquitous input device. IEEE Pervasive Computing 5(1), 70–77 (2006)
3. Baur, D., Boring, S., Feiner, S.: Virtual projection: Exploring optical projection as a metaphor for multi-device interaction. In: CHI 2012. ACM, Austin (2012)
4. Boring, S., Altendorfer, M., Broll, G., Hilliges, O., Butz, A.: Shoot & Copy: Phonecam-based information transfer from public displays onto mobile phones. In: Mobility 2007, pp. 24–31. ACM, New York (2007)
5. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: CHI 2010, pp. 2287–2296. ACM, New York (2010)
6. Boring, S., Jurmu, M., Butz, A.: Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In: OZCHI 2009, pp. 161–168. ACM, New York (2009)
7. Cesar, P., Bulterman, D.C., Jansen, A.J.: Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In: Tscheligi, M., Obrist, M., Lugmayr, A. (eds.) EuroITV 2008. LNCS, vol. 5066, pp. 168–177. Springer, Heidelberg (2008)
8. Chehimi, F., Rukzio, E.: Throw your photos: an intuitive approach for sharing between mobile phones and interactive tables. In: Ubicomp 2010 Adjunct, pp. 443–444. ACM, New York (2010)
9. Cruickshank, L., Tsekleves, E., Whitham, R., Hill, A., Kondo, K.: Making inter- active tv easier to use: Interface design for a second screen approach. The Design Journal 10(3), 41–53 (2007)
10. Dachselt, R., Buchholz, R.: Natural throw and tilt interaction between mobile phones and distant displays. In: CHI EA 2009, pp. 3253–3258. ACM, New York (2009)
11. Google Inc. Android developers (2013), http://developer.android.com/index.html (last access January 15, 2013)
12. Google Inc. LG TV remote 2011 - Android Apps on Google Play (2013), https://play.google.com/store/apps/details?id=com.clipcomm.WiFiRemocon&hl=en (last access January 15, 2013)
13. Jeon, S., Hwang, J., Kim, G.J., Billinghurst, M.: Interaction with large ubiquitous displays using camera-equipped mobile phones. Personal Ubiquitous Comput. 14(2), 83–94 (2010)
14. Microsoft. Welcome to Microsoft Pixelsense (2013), http://www.microsoft.com/en-us/pixelsense/default.aspx (last access January 15, 2013)
15. Morris, M.R.: Web on the wall: insights from a multimodal interaction elicitation study. In: ITS 2012, pp. 95–104. ACM, New York (2012)
16. Myers, B.A., Bhatnagar, R., Nichols, J., Peck, C.H., Kong, D., Miller, R., Long, A.C.: Interacting at a distance: measuring the performance of laser pointers and other devices. In: CHI 2002, pp. 33–40. ACM, New York (2002)
17. Myers, B.A., Stiel, H., Gargiulo, R.: Collaboration using multiple PDAs connected to a pc. In: CSCW 1998, pp. 285–294. ACM, New York (1998)
18. Nintendo. Wii u official site at Nintendo (2012), http://www.nintendo.com/wiiu (last access December 18, 2012)
19. Olsen Jr., D.R., Nielsen, T.: Laser pointer interaction. In: CHI 2001, pp. 17–22. ACM, New York (2001)

20. Pan, G., Ren, H., Hua, W., Zheng, Q., Li, S.: Easypointer: what you pointing at is what you get. In: CHI EA 2011, pp. 499–502. ACM, New York (2011)
21. Pawar, U.S., Pal, J., Gupta, R., Toyama, K.: Multiple mice for retention tasks in disadvantaged schools. In: CHI 2007, pp. 1581–1590. ACM, New York (2007)
22. Rashid, U., Kauko, J., Häkkilä, J., Quigley, A.: Proximal and distal selection of widgets: designing distributed ui for mobile interaction with large display. In: MobileHCI 2011, pp. 495–498. ACM, New York (2011)
23. Rekimoto, J., Saitoh, M.: Augmented surfaces: a spatially continuous work space for hybrid computing environments. In: CHI 1999, pp. 378–385. ACM, New York (1999)
24. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., Chin, J.: An experimental comparison of physical mobile interaction techniques: touching, pointing and scanning. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 87–104. Springer, Heidelberg (2006)
25. Schmidt, D., Molyneaux, D., Cao, X.: Picontrol: using a handheld projector for direct control of physical devices through visible light. In: UIST 2012, pp. 379–388. ACM, New York (2012)
26. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: DIS 2012, pp. 318–327. ACM, New York (2012)
27. Shizuki, B., Hisamatsu, T., Takahashi, S., Tanaka, J.: Laser pointer interaction techniques using peripheral areas of screens. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2006, pp. 95–98. ACM, New York (2006)
28. Sugimoto, M., Hosoi, K., Hashizume, H.: Caretta: a system for supporting face- to-face collaboration by integrating personal and shared spaces. In: CHI 2004, pp. 41–48. ACM, New York (2004)
29. Tangram. Tangram - Smart Dot Pro (2013), `http://www.tangramdesignlab.com/smartdot/en/` (last access January 15, 2013)
30. Tse, E., Histon, J., Scott, S.D., Greenberg, S.: Avoiding interference: how people use spatial separation and partitioning in SDG workspaces. In: CSCW 2004, pp. 252–261. ACM, New York (2004)
31. Vogt, F., Wong, J., Fels, S.S., Cavens, D.: Tracking multiple laser pointers for large screen interaction. In: Extended Abstracts of ACM UIST 2003, pp. 95–96 (2003)
32. Vogt, F., Wong, J., Po, B.A., Argue, R., Fels, S.S., Booth, K.S.: Exploring collaboration with group pointer interaction. In: CGI 2004, pp. 636–639 (June 2004)
33. Wilson, A.D., Shafer, S.: XWand: UI for intelligent spaces. In: CHI 2003, pp. 545–552 (2003)
34. Zizka, J., Olwal, A., Raskar, R.: Specklesense: fast, precise, low-costandcompact motion sensing using laser speckle. In: UIST 2011, pp. 489–498. ACM, New York (2011)

Situating Asynchronous Voice in Rural Africa

Nicola J. Bidwell[1,2] and Masbulele Jay Siya[3]

[1] Department of Informatics, University of Pretoria
[2] CSIR: Meraka Institute
[3] Mankosi AA, Eastern Cape, 5140, South Africa
nic.bidwell@gmail.com, jaysiya26@gmail.com

Abstract. Designing for oral users in economically poor places has intensified efforts to develop platforms for asynchronous voice. Often these aim to assist users in rural areas where literacy is lowest, but there are few empirical studies and design tends to be oriented by theory that contrasts the mental functions of oral and literate users, rather than by local practices in social situations. We describe designing an Audio Repository (AR) based on practices, priorities and phone-use in rural Africa. The AR enables users to record, store and share voice files on a shared tablet and via their own cell-phones. We deployed the AR for 10 months in rural Africa and illiterate elders, who have few ways to use free or low-cost phone services, used it to record meetings. Use of, and interactions with, the AR informed the design of a new prototype. They also sensitized us to qualities of collective sense-making that can inspire new interactions but that guidelines for oral users overlook; such as the fusion of meaning and sound and the tuning of speech and bodily movement. Thus, we claim that situating design in local ways of saying enriches the potential for asynchronous voice.

Keywords: Oral users, Rural Africa, Asynchronous voice, Social media.

1 Introduction

Most of the world's 800 million illiterate adults live in economically poor regions; in fact, interpolating Worldbank indicators, in Africa fewer people can read and write than there are cell-phone subscriptions [23]. Cell-phones offer many new opportunities to people in technology-sparse regions but they may also amplify gaps between oral and literate users and this has motivated recent efforts to develop *asynchronous voice* platforms for rural settings where literacy is lowest. Such efforts include interactive voice applications and services for cell-phones to record and access content produced locally, or by authorities or NGOs [19, 27, 31, 39] or participate in citizen journalism or online social networks [25, 29]. However, there are few empirical studies [29] and design tends to be oriented by theory that contrasts the mental functions of oral and literate people, not by practices in social situations. In this paper we describe a system we deployed in rural South Africa to enable users to record, store and share voice files for free. We aim to show that situating design in local ways of saying can inspire new interactions for asynchronous voice; thus, we start by claiming that theory currently orienting design for oral users omits factors that can prompt ideation.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 36–53, 2013.

1.1 Design Principles for Oral Users

Most design guidelines for oral users in economically poor regions [e.g. 13, 32, 37] draw on Ong's theory about cognition and the internalisation of writing [28]. Ong (1982) claimed that literate people perceive words as reified things with an existence and meaning that is separate from speech sounds and this affects memory, categorization and verbal-explanation [28]. Thus, guidelines often contrast the formulaic, prosaic and experiential thinking of oral users with the objective, logical and analytic skills of literate users [e.g. 37]. However, many studies show that literacy alone is not responsible for great shifts in mental functioning. Extensive experimental studies and ethnographic and survey-based descriptions of literacy find that certain writing systems and certain reading and writing activities foster certain forms of thinking [e.g. 36]. For instance, Qur'anic literacy improves performance on some memory tasks and literacy in a certain Liberian script enhances phonological discrimination [36]. These, and other studies [15], also show the equally specific effects of schooling on verbal explanation. Further, while guidelines emphasise oral users' deficits in abstract thinking [37], studies in Africa show that oratures involve abstract and other complex structures [1, 17], harmonic song and polyrhythm execute intricate patterns and that polylingualism, common in Africa, fosters a sensitivity to how a language works [10].

Ong's theory appeals to perspectives that deem thinking to be cognitive, tie learning to information transfer and/or model design problems as tasks. These approaches best suit studying individual capabilities and single-user interactions in controlled settings [14], such as to test oral users' learning from instructional video [20]. They cannot predict interactions between technology, orality, learning and the diverse resources and skills that people use to express and communicate ideas and work with information. Many authors admit that task-focused and cognitive studies do not address the complexity of mundane communication, social factors that effect learning or the politics of evaluating certain skills in certain ways [20]. Yet the epistemic status of these methods promotes design decisions that *are* insulated from real-world messiness. Much design for oral users draws on insights that are framed by specific goals but not by everyday practice, even when informants tell researchers that their main information source is "word of mouth" [e.g. 22]. Indeed, authors can advise grounding design in "practices of community knowledge building and transmission" [37] and "how" as well as "what" people say [32] but then extract requirements, using general-ized models of information transfer, rather than produce ideas within local dialogue.

1.2 Body and Voice

Ingold (2007) argues that Ong's theory inadequately accounts for the ways the printing press and certain literacy pedagogy disconnect graphic inscription from the gesture of writing [16]. For instance, children in the West first encounter letters as tangible objects (e.g. play-bricks) and are drilled to copy their form rather than the gesture that created it [16]. Certainly, cognitive approaches to oral users tend to ignore the actions that produce representational forms. For instance, Medhi, *et al* (2007) suggest that illiterate people recognize meanings in drawings better than photo-realistic images because they contain less information [21] but not because drawings are traces of hand movements. Interpretations that apply Ong's theory usually conflate print and hand-writing; for example to attribute disruptions in experiential information to writing itself not the printed forms that people write into [e.g. 32]. Overall such analyses separate thought/mind from action/body and neglect the knowing and reasoning constituted by physical manipulation, movement or sensory/motor memory [18].

Relegating bodily abilities to ergonomics deprives design of many constituents of orality. For instance, in Africa, gestures or mnemotechnical devices that attach verbal utterances to places or objects [1, 10] can inspire new ways to support oral practices such as tangible designs that link stories and beadwork [38]. Responding to extra-linguistic and meta-communicative phenomena, such as the mutual tuning of people's voices and bodies even when they talk using phones and cannot see one another [26], may enhance interactions. For instance, it might improve the efficacy of Interactive Voice Response (IVR), which is currently inferior to interacting with a live operator [31]. Recognising embodied social risks and emotions [18] can also guide decisions on interaction modes; for instance, while at ease using IVR to access football updates, users preferred to input to phone keypads in accessing information about HIV [27].

1.3 Designing for Oral Practice

Designing interfaces for asynchronous voice based on principles that *compensate* for what users did not attain via the pedagogies of certain literacies neglects how sense-making is always situated in social contexts. Oral societies use many skills and tools to produce meanings. Mutual fluency in specific oral forms enables members to convey and stabilize information [1] and sustain social bonds that are integral to cohesion and resilience in poor, rural communities. However, design efforts for oral users mainly separate single-user and co-present interactions with technologies such as cell-phones, despite the prevalence of proximate use and intermediary access by users in economically poor areas [e.g. 42]. Work in asynchronous voice is often inspired by the way local radio has contributed to poverty relief by enabling people to share locally relevant information regularly. Most efforts in rural areas aim to support specific interests or domains for social upliftment, such as agriculture, education or healthcare. However, understanding the local communication ecology in which information is shared is vital to avoid amplifying existing gaps in access to technology between genders, generations and areas [e.g. 2, 9, 33]. For instance, women were less able than men to participate in an Interactive Voice Forum (IVF) in Tanzanian farming communities [39]; relatively few farmers given access to an IVF in India conveyed what they learnt to those who were excluded [29]; and, social relations in and around rural clinics contributed to the failure of a healthcare communicator in South Africa [41].

Our design of the Audio Repository (AR), described here, sought to respond to practices and priorities, economics and phone-use in South Africa's rural Eastern Cape. Thus, next, we introduce the setting and our approach to designing to assist local communication. We outline factors that inspired the AR and aspects of its design. Then we summarise use of, and interactions with, the AR and pose some specific and more general implications for designing interactions with asynchronous voice.

2 Setting

Our research was set in Mankosi, along Eastern Cape's Wild Coast. It was prompted by our (Nic's) ethnography in a neighbouring area [3], which shaped the design of a mobile digital storytelling application (Fig. 1) [5] and also exposed difficulties local Xhosa people have in communicating between villages. There are 580 households in Mankosi and inhabitants have extensive kin in twelve villages spread over $30km^2$ of very hilly land. Between 500 and 900 people live in each village, as many temporarily migrate for work or school. Households survive on around $150 per month, mostly in payments from remote family members and government grants. Up to six adults and eight children live in each homestead, comprised of clusters of thatched, mud-brick rondavals, an occasional tin-roofed 2-room dwelling, animal corrals and a garden for subsistence crops. Hedges or fences around homesteads do not enclose domestic life. Children often move between homesteads and across the day people walk along many paths to communal pasture, forest, taps and dams where they collect water, fire-wood and grasses, graze animals, tend plots and make mud-bricks. There is only one bus a day which takes up to 2-hrs to traverse Mankosi and does not go to all villages.

Like 36% of South Africa's population, inhabitants are governed by a Tribal Authority, which in Mankosi consists of the Headman, a Subheadman in each village and three messengers. The Headman's and Subheadmen's homesteads are sites for local administration, from hosting meetings to notarising proof of residence papers. Post-apartheid Tribal Authorities are separate from other political bodies. Headmen relate by clan to Chiefs and inherit their role patrilineally, but they can replace Subheadmen and permit women to assume these roles. The Headman is paid a government stipend but, like almost all inhabitants, he cannot afford a car or to connect to mains electricity; still, he is, amongst the 20% who can afford to send children to school in a city.

3 Ethnographic Action Research

We learnt about Mankosi's communication ecology and engaged inhabitants in articulating requirements for information sharing and deploying prototypes in Ethnographic Action Research [40]. Critical to our methods,and accountability, are that Jay and eight other Local Researchers (LRs) generated data, advised in design, trialed systems and translated linguistically and culturally; and, we all lived in Mankosi - LRs since birth and Nic for 20 months from 2010 to 2013. We produced data using mixed methods [e.g. 4, 7] including observations, contextual inquiry, interviews, diary studies, auto-ethnography and deploying, and training people to use, different prototypes. Few local people speak English and, while Nic's rudimentary isiXhosa has improved, in most activities Nic spoke in English, others spoke in isiXhosa and LRs translated. We recorded activities in handwritten notes and, sometimes, photos and video.

LRs insisted we consult the Tribal Authority so we began by gaining approval to experiment with communication systems that might assist inhabitants, and perhaps in turn others like them elsewhere. Subsequently, we observed or participated in many community meetings and met with the Tribal Authority almost 40 times to discuss ideas, plans and problems in designing, trialing and maintaining systems (Fig. 1). Shortly after our first meeting LRs and other people began video-recording conversations, storytelling and presentations. They accumulated over 50 items featuring some 60 people, aged 14 to 80 years, which yielded data on values, priorities and oral style.

We also undertook over 300 interviews, individually or in groups, specifically about communication practices and use of phones, phone services and prototypes. We first interviewed 141 people about phone ownership and use and, as this showed the prevalence of using free Callback services, we interviewed another 16 people about Callback [4]. Then we generated data on the ways people manage mixtures of communication in their daily routines in diary studies across periods of 4 to 10 days. This involved individual or group interviews at the start and end, and 72 short individual interviews that reviewed phone-use in between. We included six male and six female older owners of low-end phones, half of whom are illiterate; and, ten younger, literate owners, some of whom we interviewed using a text-based, chat service that is popular elsewhere in South Africa. A UK researcher, visiting for two weeks, also conducted 17 group or individual interviews with 28 younger feature-phone owners [11].

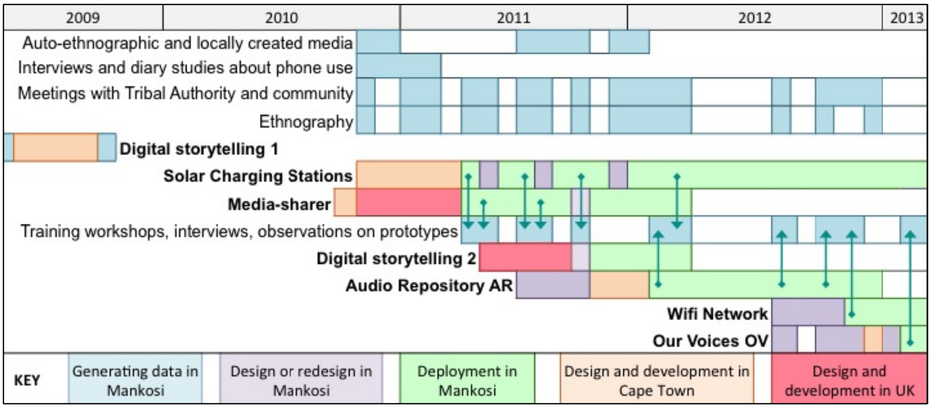


Fig. 1. Timeline of activities to gather data, design, develop and deploy and evaluate prototypes

We shared our data with UK researchers who developed two prototypes, for creating, storing and sharing media. We deployed their first prototype, a media-sharer, in April 2011, which yielded insights for our design of the Audio Repository that we introduced in early 2012 (Fig. 1). The media-sharer ran on two tablets each deployed with solar-powered, cell-phone Charging Stations [7] as we hoped that people would use the media-sharer when they came to charge their own phone for free. We will soon report the design of the media-sharer, which uses real-time, text-based chat between users' own phones but sends/receives photos later between users' phones and communally accessible tablets. Inhabitants chose two sites for the Stations that were 2.5Km, or 25-mins walk, apart: in the homesteads of the Headman and a Subheadmen in Mankosi's poorest area, Ridge [7]. We trained LRs to use the tablet, media-sharer and Stations and then ran seven workshops on the media-sharer with 24 inhabitants, mostly members of the Tribal Authority or Community Association. The media-sharer was unsuited to most local phones, so we also gave low-end Bluetooth-enabled phones to Tribal Authority and Community Association members. We observed each Station site for over 80-hrs on different days of the week, at different times of day and different times in the year and interviewed 40 people who left or collected phones at Charging Stations [7]. We logged use of the media-sharer automatically; who charged phones each day in notebooks; and, the distance between Stations and the homesteads of 40 Station users. In the ensuing year few of the 700 people who charged phones [7] used the media-sharer. We tried to promote use of the tablet by uploading locally created audio commentaries, photos and videos of local soccer matches and exploring with players how tablets might support the league. In October 2011, UK researchers introduced a digital storytelling prototype for Android smartphones that interfaces with the media-sharer (Fig. 1). They left two phones and a tablet with LRs to loan to people [11], but nobody has used the application yet.

LRs verbally translated audiovisual records and we transcribed audio and/or annotated video. We thematically coded data after each activity and cross-linked themes between activities. We used descriptive statistics to analyse responses to closed questions in interviews and gave written descriptions of varying thickness to UK researchers. We validated and revised interpretations of themes hermeneutically, as dialogue and events produced new meanings, and recently began detailed analyses of video to explore themes that often occurred when we cumulatively integrated codes.

4 Inspiration for Designing the Audio Repository

Our data on co-present communication and phone use in rural Africa confirms, disputes and significantly adds to that already reported in HCI4D [e.g. 4, 6, 7]. Here we present issues that often arose in thematic analyses and oriented designing the Audio Repository. We illustrate some of these issues with ethnographic details.

4.1 Orality and Deliberate Qualities in Speaking

Inhabitants spend much time talking to greet, support, learn, gossip or debate and some 40% of adults, including elders in the Tribal Authority, do not read and write. There are few reading and writing materials locally, even in schools, and people's most common experience of print is a phone's interface. Many recognize the role of orality in local identity and belonging, for instance one young man linked his talent for poetry, inherited from an uncle with whom he performed at ceremonies, to "fighting the westlife" on behalf of the Xhosa. The vast range of video inhabitants recorded focused on the speaker not the visual setting. People said audio recording was a useful feature of phones [4] and some used feature-phones to record local performances, events, conversations, jokes and stories as well as music from the radio [11]. In fact participants in workshops on the media-sharer often recorded themselves speaking or singing rather take photos. The preferences of men to record their speech and women to record song align with practices in ceremonies and church.

Inhabitants prefer to use phones for voice calls over other services, especially if they are older. For instance, SMS comprised only 1.4% of older people's phone-use in diary studies. People rarely use free services that send an SMS to recipients who do not pick-up, nor leave voicemail as it consumes airtime they would prefer to use to talk. Despite favoring voice, inhabitants make brief calls as average weekly spending on airtime limits speaking to just 4 minutes at basic rates and few phones can access cheaper data rates for VOIP [4]. This constrains expression, especially as isiXhosa can take three times longer than its English translation, and means phone use is more often planned than impromptu. Inhabitants mostly interact with phones in their homesteads and rarely while walking along paths/roads, in transport or in meetings. They said that prior to owning a cell-phone others phoned or sent a message to someone else in their village who purposively walked to bring them a message or the phone.

Deliberate qualities in speaking also arise in local governance, which is based on customary structures and community meetings. Up to 70 inhabitants (55% male, 45% female, and 60% older than 30 years) attend weekly meetings, in the Headman's homestead, that address Mankosi's internal issues and/or host counselors or municipal officers. Meetings often take 3 hours and can entail walking for over 2-hrs to attend. This limits turnout. People sit according to gender and age to speak a person raises a hand, stands and removes his hat, if he is a man. People attend to the speaker and nobody interrupts. The Tribal Authority aims to reach consensus in meetings and, after opening a meeting, the Headman usually listens silently until the end when he states how an issue would be resolved or discussed further. The Tribal Authority's voluntary secretary writes minutes and attendee's details on paper but does not disseminate these due to cost, infrastructure and because Subheadmen and Headman's messengers are conduits for information across villages. This fuels allegations about the Tribal Authority's inertia, opacity and "forgetfulness". The Tribal Authority said that audio recordings of meetings would "stop people lying", which approximates to being held accountable, and could inform those unable to attend and other inhabitants also hoped communication systems might improve information dissemination.

4.2 Differences in Access to Phones and Literacies in Phone Use

Some 30% of inhabitants over 14 years old do not own a phone, only 25% own feature-phones and people under the age of 35 are twice as likely to own a phone than older people [4]. Older and/or illiterate people often said that they were "not educated to use the phone" or did not use free or low cost services and call rates as they could not "read properly". Indeed, the amount of print in the media-sharer contributed to its lack of use. Younger and/or literate phone owners have many tactics to use text-based services to coordinate communication, seek/offer support or sustain a presence in other's lives. They choose to use these based on their recipients' technological or print literacy [4]. Younger people often use free services and low-cost IRC chat services to communicate with friends and people in cities, including those they do not know. Older people perceive this as playful and irrelevant to local communication. Generational differences in phone-use are acerbated by authority protocols and a dramatic population decline between the ages of 20 and 44 years ('skip generation'). People of parental age are those most likely to have died from illness (often related to a local HIV prevalence of 29%) or to work in cities and mines. Grandparents often head households but respect for elders can make younger people feel uneasy about teaching them to use cheap services. Indeed, LRs often interacted with phones for elders in workshops. Illiteracy is not stigmatized but Tribal Authority members were shy to reveal that they could not read. The Headman refused to try the media-sharer, though with practice other elders used the tablets' direct manipulation interface easily.

Print and technological literacies have different implications for privacy in phone use. Phone owners said phones helped to make their lives more private and their patterns of calling or walking between homesteads had changed, so now they chose to "shout-out" or make a deliberate missed call ("buzz") depending on their reason to communicate. Those with low literacy gain help from trusted intermediaries and close family members know each other's PIN codes and have various tactics to prevent another's use of their airtime. However, an illiterate mother was concerned that people took photos of others without consent and, meanwhile, we had to discourage LRs, aiming to prevent theft, from saying that shared equipment can track users.

4.3 Temporality and Phone Use

Perceptions of opacity and delay in local governance communication are situated in access to phones and phone services, the pace of life and frequent co-present contact within villages. People spend much time outdoors in Mankosi and an ambient orality pervades daily life as they call between homesteads, sing in gardens, whistle to animals and greet and chat as they walk to communal resources [7]. Most phone owners acquired phones between 6 months and 4 years ago and prioritise using phones to sustain supportive connections [4]. Their phone use tends to split between contact with people living nearby and much farther away. For instance, older participants in diary studies had seen, or predicted to see, nearly a third of the people with whom they had interacted by phone within a day of that interaction; and, another half of their phone interactions were with close family members or friends in places that entail at least a day of travel. Only 13% of older people's phone use was contact with people in other villages in Mankosi and only 10% of their use specifically arranged meeting.

Cost and access to infrastructure affects the temporality of phone use. For instance, some 50% of phone-owners' diaries recorded at least one day when they did not use their phone at all. Firstly, cost limits immediacy in communicating. Most phone-users spend R7 (1.1 US$) per week on airtime as pay-as-you-go subscribers to MTN's network [4], which entails keeping a small airtime credit. Also inhabitants usually pay on average $0.70 to charge their phones, often at spazas (small shops) and shebeens (informal bars) [4] and can go for a week without charging so conserve charge by switching off phones [7]. Secondly, cellular network coverage and electricity access limit connectivity. People undertake daily tasks in places with unstable connections and can leave phones to charge for two days if charging is free [7]. Thirdly, low-end phones do not always display that they received calls when they were off. Finally, access to phones or assistance via an intermediary effects temporality; for instance, an illiterate mother always waits for her daughter to walk home from school to read aloud and reply to her messages.

5 The Audio Repository

We designed the AR to run on the 7inch Samsung Galaxy tablets that inhabitants had kept at the Stations, that had charged up to 20 phones a day in the past year. The AR, developed in Java and SQLlite, consists of a database and API to make, access and share recordings and create and join groups. We sought to respond to local practices and rapidly deploy a working prototype to allay frustration with the media-sharer. Jay visited the developer, Thomas, in Cape Town. Their design discussions were shaped by Jay's experiences in trialling the media-sharer, Thomas' design of our original digital storytelling application [5] and the contrast of the city and Mankosi [34]. Here we outline only the main conceptual, interface and interaction aspects of our design.

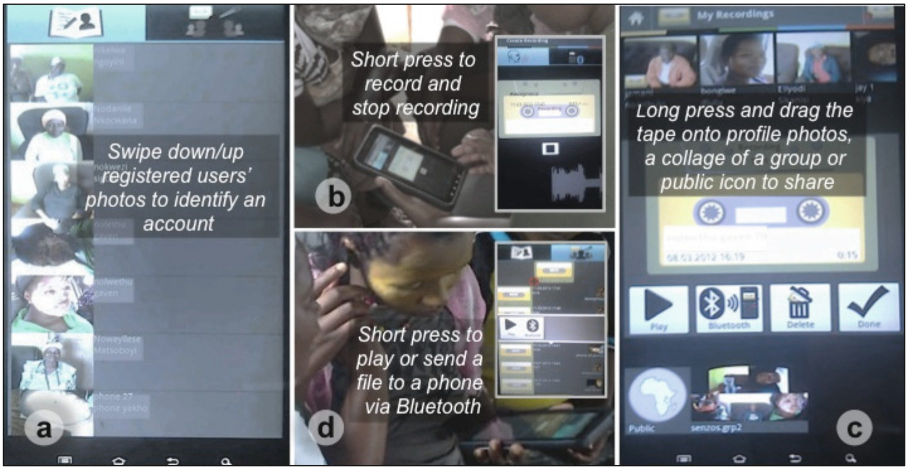


Fig. 2. Interfaces to find account (a), record (b), share (c), play or send (d) recordings

All users can record and listen to voice files using the AR on the tablet, whether or not they have an account. However, unregistered users can only interact with files that are made ´public´. Sharing files does not require owning a phone, unlike for the media-sharer, but users can send/receive files to/from cell-phones over Bluetooth (Fig. 2d). We tested the AR with Nokia 2730 and 2330 Classic phones, since some 60% of local phones are Nokia [4], but it also works for other low-end Bluetooth phones with enough unused storage space. Recording functions are merged in one button, beneath a cassette tape icon, so when users press the red circle, to record, it turns into a black square to stop. The AR also provides other feedback that it is recording (Fig. 2b).

By registering for an account users can store audio files; send/receive files to/from other registered users; and, create and enable access to "Groups" of registered users. Users register with their first and second name; a photo, taken using the tablet; and an alphabetic/numeric password, since people use PINs on their phones and many illiterate inhabitants can recognize 10 to 20 names or numbers. Users scroll vertically through profile photos to find their account (Fig. 2a) as inhabitants recognized people's photos in the media-sharer. To assist identifying files we used a basic folksomic strategy by displaying those with whom a user shares a file [34]. We symbolized sharing with a palm-up, open hand because it is more relaxed than the deferential local gesture of giving and receiving with both arms.

Users scroll horizontally through profile photos to find other users' accounts to share with (Fig. 2c). Deliberate qualities in speaking and passing messages inspired designing interactions for sharing and 'unsharing' files based on slow gestures not standard ways that aim to increase the speed of interactions [34]. To share a file a user 'long-presses' on a tape icon and drags a small copy of the tape over target users' photos, a collage of Group members' photos, or the public icon (Fig. 2c). The tape drops onto the corner of the photo, when the user lifts his/her finger, where it stays to show the recording is shared.

6 Deploying and Using the Audio Repository

Inhabitants of Mankosi used the AR for 10 months. We introduced and sought opinions about it in 7 workshops lasting 2-hrs to 4-hrs. LRs taught 50 women and men to use the AR, in groups of 2 to 8, and suggested they use Callback [4] to alert others that they had shared a recording. We interviewed 23 people, in four groups, about the AR and owning and sharing digital media and observed useage after deployment. After 9 months we analysed recordings, interviewed users of the AR, and discussed with other inhabitants their perceptions of use and ways to improve access to the AR.

6.1 Use and Potential Use

Inhabitants were very enthusiastic about the AR, and twice 30 women arrived for workshops when we expected six. They said they would use the AR to leave updates about events (e.g. church, choir) or messages for family members that some also said might become mementoes. Community Association members said the AR would help to record meetings as written minutes did not always accurately report what was said. However, despite wide interest, access to the AR was restricted. The Charging Station at Ridge was damaged a week after residents assumed post-trial responsibility for it and conflict arose [7]. The Tribal Authority decided to remove and relocate that Station and tablet and agreed that women could select a new site in Ridge but women did this only recently. Meanwhile, the Headman moved the Station in his homestead away from the office where he stored the tablet and later began to charge the tablet using a system installed for a subsequent WiFi network [6, 35]. He is wary of untrained use and to conserve battery turned on the tablet only to record or listen to files. Thus, people who came to charge their phones did not have easy access to the AR.

The AR was used almost exclusively to record meetings the Headman attended on 17 different days. The Headman stored all recordings in his account: 35 of meetings and 12 that he used to practice or that contained no audible sounds. Mostly there were less 5 recordings a month but in one month there were 17 and there were two months without any recordings. Once the tablet's battery totally drained which meant the date was incorrect for 50% of files and tablet had been re-set to a Cyrillic script. A third of the files recorded information about issues affecting inhabitants, a third recorded "cases" of tribal law, 20% recorded events in Mankosi (e.g. meetings to end the Stations' trial) or elsewhere hosted by the Chief or municipality. One file recorded a song from the radio and another a discussion about a useability problem with the tablet. The Headman said the AR was vital as key Tribal Authority members had been less available than usual and said that one day he was alone when another Headman arrived to contest Mankosi's boundary so, while a child went to alert his secretary, he used the AR to record the exchange. The Tribal Authority said that others were impressed by their advancement in using the AR at Chief's meetings and five Headmen (governing around 50,000 people in total) had asked how they too could have an AR.

The Headman asked us to delete temporary accounts made in training workshops at Ridge, as other Headmen had reproached him for having too many female friends. Most files (78%) contained only male voices and across the set of files 88% of voices were male. We were able to differentiate between voices in 33 files and found an average of 5.4 different male voices per recording. We discerned an average of 2.9 different females in files containing female voices. There were two files that recorded only female voices amongst a set of five in which a municipality speaker, at the opening of a pre-school, spoke about the use of children's social grant.

The Headman calls the tablet his "witness" and said he had not deleted any files but he rarely shared recordings using the AR. This might relate to other inhabitants' comments, in workshops, that there would be "a fight" or "trouble" that entails negotiating with the recorder, if a person contested sharing a recording of them. He recorded topics that he said need to be "referenced" in the future. Some 16% of the files recorded discussions of protocol and inherited authorities or responsibilities and included information about dreams, ancestors and the Royal family. Some of this dialogue occurred in the 14% of recordings that contained disciplining speech acts by, and apologies to, the Tribal Authority. These and other files also related to various disputes within Mankosi or with another community. Most disputes arose in Tribal Authority cases (e.g. paying "damages" to the family of a pregnant woman; fights at soccer matches or within families; physical ostracism and verbal insults). A third of files recorded discussions about environmental resources - 17% explicitly about land, including permission to build; and, 14% about specific assets (e.g. sand). Some 14% of files contained dialogue about the Community Trust, which is expected to manage income from resources and a few files related to local facilities and jobs (e.g. informing workers on local government projects about wage delays).

6.2 Observed Interactions

In workshops illiterate and literate, older and younger, women and men readily learnt and taught each other to use the AR. The Headman eagerly learnt to use the AR (unlike the media-sharer) and taught other Tribal Authority members who learnt less swiftly. Video showed people huddled around tablets, bent their bodies together to listen and held the tablet for each other (Figs. 2b, 2d, 3a). They pointed to the screen to assist others and some made rhythmic hand gestures to depict interactions (Fig. 3b).

The AR orders profiles alphabetically by first name (Fig. 2a) but people found their accounts, or other users, more easily using photos. This is partly because nearly 50% of female first names start with 'N' (50% of male names start with 'S', 'M' or 'A'. The vertical display of alphabetically ordered photos (Fig. 2a) also affects other meanings. For instance, the Headman said that a photo directly above his own connoted disrespect as it depicted a young man using a rap-style gesture "over" his head.

Fig. 3. Subheadmen (a) and older and younger women (b) gestured (white circle) as they interacted and assisted each other in workshops on the AR. Men carry the tablet (yellow circles) to record community and Tribal Authority members speaking in meetings (c, d).

The Headman decided an issue warranted recording based on who spoke and "what they start to say". Only two files recorded a whole meeting and files usually recorded one main topic. Some 70% of separate files that were made on the same day recorded discussions of different topics. When separate files were recorded on the same day, at most two files recorded the same or a similar topic. Some 14% of files contained long pauses without speech, this was mostly because in meetings a man carried the tablet between each person as s/he stood to speak (Fig. 3c, d). His movement and the tablet's size afforded visibility and twice in meetings Community Association members resisted recording as their permission had not been sought.

The tablet's orientation between people during recording made 9% of files inaudible. Further, 51% of files were recorded outdoors and the wind was audible in a third of these, which made understanding speech in two recordings impossible. Some 21% of files also recorded music, children or other ambient noise.

The Headman used the AR to remind himself of topics after a meeting and "to deliver information" to the Tribal Authority. The Headman listened along with other members, rather than sharing asynchronously, partly because he is the most proficient user. His secretary, however, listened alone to enhance minutes of meetings.

People found a part of a recording easily. Of the Headman's recordings, 22% were shorter than 10-mins; 26% of 10 to 20-mins; 44% of ½ to 1-hr; and, 7% exceeding 1-hr, including one over 4-hrs. He used the time bar that displays when the audio plays to move through the file, but this functionality does not work for recordings of over 1-hr. Inhabitants found a specific file visually less easily, even with few to search. Identifying a recording based on those with whom it was shared did not help the Headman who did not share via the AR. Instead, he looked at the date on the tape, but this was incorrect for 50% of files because the battery had drained.

7 Implications for Design

Use of, and interactions with, the AR shows that portable, asynchronous voice platforms can support oral practices in rural Africa, such as in local governance. To do so, however, it is vital to improve audio quality so it performs better outdoors. Tablets do not yet suit low-electricity conditions but their size and direct-manipulation interfaces offer special affordances for recording and interacting with voice files. Practices in using the AR suggest that accounting for temporal and spatial qualities in co-present interactions might be as important to design as, say, the speed of interacting with information and disseminating it widely. These insights are novel to reports of designing for oral users in technologically sparse settings. Thus, grounding design in local oral practices heightened our sensitivity to conceptual issues that neither cognitive models nor guidelines based on Ong's theory [28] anticipate. We discuss these sensitivities after we outline our current design response to use of the AR in Mankosi.

7.1 Our Voices

Encouraged by enthusiasm for the AR we recently introduced a new prototype for sharing recordings to Mankosi: Our Voices (OV). The Tribal Authority's control of the AR, selectivity in recording and co-present sharing did not widen dissemination but, on the other hand, it improved aspects of information reliability amongst those who have few ways to use phone services. To improve access, and the AR's fragile strategy of manual back-up to external SD cards, OV runs on three tablets at different sites and enables sharing files between tablets when they connect to a wireless mesh network that Jay recently set-up with new collaborators [35]. Community Association members who oversee a tablet in south Mankosi, 1-hr walk from the Headman, say they hope to demonstrate good practice in recording and sharing voice-files using OV.

Although inhabitants appreciated the concept of creating Groups of users the Tribal Authority did not share recordings to Groups in the AR. Nor did they delete files. Further, inhabitants said that a person who records another owns the audio recorded and, if they share it, then all recipients own it. Thus, OV does not include Groups or permit deletion but enables re-sharing files and displays the provenance of sharing.

Many co-present interactions and users' remarks (e.g. the Headman's comparison of the tablet when reset to a Cyrillic script to "riding an unfamiliar horse") shifted our focus to bodily interactions. The AR's deliberately slow interaction for sharing files

[34] resonated with the tempos of physically sharing the tablet in workshops and meetings and suggested that interactions with direct manipulation interfaces can tangibly trace social relations. Thus, OV enables re-ordering users' profiles displayed on each tablet by long pressing and dragging profile photos; for instance, to cluster profiles according to the frequency of using a specific tablet or social roles or relations.

Enabling users to re-order accounts might reduce the time it takes them to find their accounts and other users amongst the wider user base of the distributed OV. Inhabitants do not judge the efficacy of information sharing in relation to Western time scales but electricity is a critical constraint to access, so we included other visual features to speed up search and retrieval. The OV displays the profile of those with whom a user shares most often in a drop-down box. Further, it highlights recordings that users have not yet played and sorts files in the order they were recorded, to avoid the problem of incorrect date-stamps when a tablet's battery drains.

As users sweep through their recordings OV displays a photo reminder of the event recorded by their audio file. Users can take the photo, using the tablet's camera, and attach it to a file at the time or after the event recorded. This aims to assist visually searching for files, but it also offers a way to collect locally created voice and photo instructions to populate a voice-based help and feedback system. Voice-to-text processing for speech recognition demands high quality recordings and computational and electricity resources that are infeasible for tablets in Mankosi. Thus, instead, we are attempting to integrate links between bodies and sounds into search and retrieval. The Headman recognized the topic or setting almost the instant he played a file; thus, as users sweeps through recordings OV starts to play either that file or a user's short vocal abstract about it. Users can record and attach their abstract at the time of recording the main audio file or later, in a similar way as they can tag a file with a photo.

Peer-to-peer synchronization requires temporal overlap when two tablets connect to the WiFi network. This cannot be based on clock-time, since users do not wear watches and the time on electrical devices may be wrong. Thus we are exploring environmental patterns that can prompt switching on tablets to synchronize them, as our ethnography shows inhabitants tune their practices to local rhythms (e.g. sunrise, people or animal movements) [6].

7.2 New Sensitivities in Designing for Oral Users

Developing the AR revealed that advances in audio acutely lag visual technology and our experiences in deploying the media-sharer illustrate that design privileges visual media. This visual bias restricts opportunities to design for oral practices and access by oral users. It contributes to embedding a "language" into interfaces [24] that is further shaped by designers' internalization of print. This design language introduces meanings that can have far reaching ethical consequences [24], even if designers disagree with Ong's evolutionary thesis on the role of writing in 'civilization' [28]. Some who apply Ong's theory recognize that it is the "collective that embodies the shared experience that constitutes the pool of knowledge available to the community" [32]. However, there are no reports of grounding the design of social media in local practices of speaking or situating the design of asynchronous voice systems in co-present interactions. This contributes to exclusion. For instance, researchers hoped the users they chose to participate in an IVF would transfer information to those excluded from the trial [29] but did not consider how designing the IVF for single-user interactions with phones amplifies existing exclusionary infrastructures.

Our ethnography elsewhere in rural Eastern Cape alerts us to qualities of collective sense-making with recorded information. We (Nic) observed science teaching for 5-hrs in high-school classrooms and every few minutes all the teachers said in English "are we moving together?" or "sihamba" or "sihamba sombani" in isiXhosa, which literally means walking together but also translates to "do you understand?". They used these expressions most when reading textbooks: they read aloud a piece of text and then, often mid-sentence, paused or asked "and what?" whereupon learners read aloud the end of the sentence or the next. Patterns in linking voices, pauses, meanings and print remind us of Ingold's (2007) account of early reading. People read manuscripts aloud and treated them as sets of signposts to guide them, Ingold writes, "within the landscape of memory" because meanings *occur* as words are spoken [16].

Nobody experiences reality as a static array of autonomously meaningful things [16] yet we usually start design by reifying events, agents and objects from the relations of practice that give them meaning. Consider, for instance, how modeling names as attributes of a person entity de-contextualises names (e.g. as an attribute "alias" [32]). Mankosi's inhabitants use various names according to circumstances. For example, they personalize Callbacks [4] with different names depending on why they sent a Callback; and, LRs used different identifiers for the same person in Station logs sometimes combined with other words, such as a prefix like "mama" or "sisi" or the name of one of their children [7]. Names emerge along spatial and temporal paths and recall relations with ancestors through clan, in families by birth order, marriage or parenthood and between peers in shared rites of passage. Importantly, people usually experience names as sounds. Conceptualising words as things that are separate from their vocal, bodily, affective and historical paths omits many relational aspects of sense-making. A heightened sensitivity to this relationality inclines us to embrace the fusion of meanings, sounds and the bodies that produce and experience them.

Contrary to Ong [28], Ingold proposes that it was print, not writing, that reified words because it severed the inscription from the body movement [16]. Interactions with the AR suggest that shared tablets can support the fusion of meanings, the sonority of words and the bodies that speak and hear them to share information. Their direct-manipulation interfaces offer affordances for extra-lingustic qualities of orality; such as those that arise when women in Mankosi sat closely together and gestured as they spoke and guided others to use the AR. The tablet's portability meant audio files trace spatial patterns such as pauses and sounds recorded along topokinetic paths when a person carried the tablet from one to the next in meetings and indications of the relative orientations of recorders and speakers. We have reported associations between bodies, utterances, spatiality and social relations in sense-making elsewhere in rural Africa [8], and we found that certain meanings emerge in Mankosi when gestures, social roles and the spatial arrangement of photos combined.

Speech sounds and bodily movements are not easily split into discrete events, so we advocate embracing qualities of this continuity in design. For instance, the AR's interaction style for sharing reconciles with the temporal textures of movement when people handed or carried the tablet to others or interacted together with it. This is a more nuanced approach to temporality than embedding Western experiences, conceptions and economies of time in design [6], such as to speed up searching because voice recordings listening to them is "tedious" [28]. Guidelines for oral users already suggest that rhythms in verbal utterances aid memory [37]. We propose that responding to other rhythms in information sharing can offer many more affordances in interacting with asynchronous voice, so we end by speculating about this.

In our design workshops for the original digital storytelling application a group of women fluidly took turns speaking when they recorded a story about unfolding events in a child's morning routine [4]. Their familiarity with each other and their daily routines contributed to this fluidity. The tuning of daily practices to rhythms in the environment may offer a way to prompt inhabitants to switch on and connect tablets to the network to synchronise them [6]. More profoundly, the women's fluidity in speaking one-by-one and barely without a pause between each affected, and was affected by their interactions with the recording devices [4]. Drawing on diverse research from music to neuroscience, Gill (2012) argues that the rhythmic synchrony of bodies and voices is integral to sociality and enables "being together" [12]. We all have rhythms, in breath, in movement, in gait, that we tune with our speech. Our self-synchrony supports inter-personal coordination by enabling us to adapt our rhythms to others to time our syllables and pauses as we converse. Gill cites a study in which pairs of participants, in isolated rooms who talked using microphones, finely tuned their body movements to their own and the other's voice. Their rhythms aligned with the language they spoke but concurrent movement also arose without speech [12]. Such synchrony emerges in diverse oral practices in rural Africa - women sing together in harmony, younger people adapt their gestures to elders and protocol, and men respond to each others' body movements in meetings. This may inspire designing interactions, say to link gestures to vocal input/output, that support sharing information.

8 Conclusion

We designed an Audio Repository (AR) to enable users in rural African villages to record, listen and share voice recordings, for free. Inhabitants learnt to interact with the system easily and illiterate elders, who have few ways to use free or low-cost phone services, used it to record and review community meetings for ten months. We situated the concept and interaction design of the AR in local communication practices and phone use and did not analyze users' mental functions or apply principles derived from theory about the internalization of writing. Use of, and interactions with, the AR informed the design of a new prototype. Our observations also suggest that the fusion of meaning and sound and tuning of speech and bodily movement can inform strategies for displaying, searching and sharing information. Current design guidelines do not promote such ideas as they focus on single-user interactions and separate mind and body. Thus, we hope to motivate others to enrich the potential of voice-based applications for oral users by situating design in local ways of saying.

Acknowledgements. We are very grateful to all inhabitants of Mankosi, especially LRs. We also thank Thomas Reitmaier for developing the AR and OV, and Simon Robinson and Elina Vartiainen for the media-sharer; Gary Marsden for designing the Charging system; Carlos Rey-Moreno and Zukile Roro for the WiFi mesh; Paula Kotze, Mounia Lalmas and Bill Tucker for support; and, Irina Shklovski, Beki Grinter, Sam Merritt and Heike Winschiers-Theophilus for comments on earlier drafts. Our work was funded mostly by CSIR-Meraka and partially by an UK EPSRC grant (EP/H042857/1).

References

1. Barber, K.: Text and Performance in Africa. Oral Tradition 20(2), 264–277 (2005)
2. Best, M., Smyth, T.N., Etherton, J., Wornyo, E.: Uses of Mobile Phones in Post-Conflict Liberia. Inf. Tech. & Int'l Development 6(2), 91–108 (2010)
3. Bidwell, N.J.: Anchoring Design to Rural Ways of Doing and Saying. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 686–699. Springer, Heidelberg (2009)
4. Bidwell, N.J., Lalmas, M., Marsden, G., Dlutu, B., et al.: Please call ME.N.U.4EVER: Designing for 'Callback' in Rural Africa. In: IWIPS 2011, pp. 117–138 (2011)
5. Bidwell, N.J., Reitmaier, T., Marsden, G., Hansen, S.: Designing with mobile digital storytelling in rural Africa. In: CHI 2010, pp. 1593–1602. ACM (2010)
6. Bidwell, N.J., Reitmaier, T., Rey-Moreno, C., et al.: Timely Relations in Rural Africa. In: 12th Int'l Conf. on Social Implications of Computers in Developing Countries IFIP (2013)
7. Bidwell, N.J., Siya, M., Marsden, G., et al.: Walking and the Social Life of Solar Charging in Rural Africa. Transactions on Computer-Human Interaction (2013)
8. Bidwell, N.J., Winschiers-Theophilus, H., Koch Kapuire, G., Rehm, M.: Pushing Personhood into Place: Situating Media in the Transfer of Rural Knowledge in Africa. Int'l J. Human-Computer Studies 69, 618–631 (2011)
9. Blumenstock, E.J., Eagle, N.: Divided we call: Disparities in Access and use of Mobile Phones in Rwanda. Inf. Tech. & Int'l Development 8(2), 1–16 (2012)
10. Finnegan, R.: The Oral and Beyond: Doing Things with Words in Africa. James Curry/Uni. of Chicago (2007)
11. Frohlich, D., Robinson, S., Eglinton, K., et al.: Creative Cameraphone Use in Rural Developing Regions. In: MobileHCI. ACM (2012)
12. Gill, S.P.: Rhythmic Synchrony and Mediated Interaction: towards a framework of rhythm in embodied interaction. AI & Society 27, 111–128 (2012)
13. Gorman, T., Rose, E., Yaaqoubi, J., et al.: Adapting usability Testing for Oral, Rural Users. In: CHI 2011, pp. 1437–1440. ACM (2011)
14. Harrison, S., Tatar, D., Sengers, P.: Making epistemological trouble: Third-paradigm HCI as successor science. Interacting with Computers 23(5), 385–392 (2011)
15. Hull, G., Schultz, K.: Literacy and learning out of school: a review of theory and research. Review of Educational Research 71(4), 575–611 (2001)
16. Ingold, T.: Lines: A brief History. Routledge (2007)
17. Kaschula, R.: The Bones of the Ancestors are Shaking: Xhosa Oral Poetry in Context. Juta, Cape Town (2002)
18. Klemmer, S., Hartmann, B., Takayama, L.: How bodies matter: five themes for interaction design. In: CHI 2006, pp. 140–149 (2006)

19. Kumar, A., Sheetal, K., Agarwal, S.K.: Spoken web: using voice as an accessibility tool for disadvantaged people in developing regions. Accessibility & Computing 104, 3–11 (2012)
20. Medhi, I., Menon, R., Cuttrell, E., Toyama, K.: Correlation between limited Education and Transfer of Learning. Information Technologies and Development 8(2), 51–65
21. Medhi, I., Prasad, A., Toyama, K.: Optimal audio-visual representations for illiterate users of computers. In: 16th Int'l conference on World Wide Web, pp. 873–882. ACM (2007)
22. Medhi, I., Sagar, A., Toyama, K.: Text-Free User Interfaces For Illiterate And Semiliterate Users. Technologies for Int'l Development 4(1), 37–50 (2008)
23. Melamed, C.: `http://www.globaldashboard.org/2011/02/07/what-if-you-have-a-mobile-phone-but-you-cant-read/`
24. Merritt, S., Bardzell, S.: Postcolonial language and culture theory for HCI4D. In: Ext. Abstracts CHI 2011, pp. 1675–1680. ACM (2011)
25. Mudliar, P., Donner, J., Thies, W.: Emergent practices around CGNet Swara, voice forum for citizen journalism in rural India. In: ICTD 2012, pp. 159–168. ACM (2012)
26. Murray-Smith, R., Ramsay, A., Garrod, S., et al.: Gait alignment in mobile phone conversations. In: MobileHCI 2007. ACM (2007)
27. Ndwe, T., Barnard, E., Mqhele, E., Dlodlo, E.: Effects of application type on the choice of interaction modality in IVR systems. In: South African Institute for Computer Scientists and Information Technologists Conf., SAICSIT 2012, pp. 236–242. ACM (2012)
28. Ong, J.: Orality and Literacy: Technologizing of the Word. Routledge, New York (1982)
29. Patel, N., Chittamuru, D., Jain, A., et al.: Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In: CHI 2010, pp. 733–742. ACM (2010)
30. Patel, N., Klemmer, S.R., Parikh, T.S.: An asymmetric communications platform for knowledge sharing with low-end mobile phones. In: UIST 2011, pp. 87–88. ACM (2011)
31. Patnaik, S., Brunskill, E., Thies, W.: Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice. In: Int'l Conf. on Information & Communication Technologies & Development, ICTD 2009 (2009)
32. Pitula, K., Dysart-Gale, D., Radhakrishnan, T.: A Case Study in Requirements Engineering for ICT4D. Inf. Tech. & Int'l Development 6(1), 78–93 (2010)
33. Porter, G., Hampshire, K., Abane, A., et al.: Youth, Mobility and mobile phones in Africa: a three country study. Information Tech. Development 18(2), 144–162 (2012)
34. Reitmaier, T., Bidwell, N.J., Siya, M., et al.: Designing an Asynchronous Oral Repository for Rural African Communities. In: IST-Africa. Cunningham & Cunningham (2012)
35. Rey-Moreno, C., Tucker, W.D., Bidwell, N.J., et al.: Experiences, Challenges and Lessons from Rolling Out a Rural WiFi Mesh Network. In: DEV 2013. ACM (2013)
36. Scribner, S., Cole, M.: The psychology of literacy, Harvard Uni. Cambridge, MA (1981)
37. Sherwani, J., Ali, N., Rosé, C.: Orality-Grounded HCID: Understanding the Oral User. Inf. Tech. & Int'l Development 5(4), 37–50 (2009)
38. Smith, A., Reitsma, L., van de Hoven, E., Kotze, P.: Towards Preserving Indigenous Oral Stories Using Tangible Objects. In: Culture-Computing, pp. 86–91 (2011)
39. Sullivan, B.: Freedom to Learn: Blending Interactive Voice Response and Radio. Learning with Community Media, pp. 131–135. Commonwealth of Learning (2012)
40. Tacchi, J., Slater, D., Lewis, P.: Evaluating Community Based Media Initiatives: An Ethnographic Action Research Approach. In: IT4D 2003 (2003)
41. Tucker, W.D., Blake, E.H., Marsden, G., et al.: Reflection on three years of Rural wireless Internet. Protocol communication. In: SATNAC 2007, pp. 452–457 (2007)
42. Walton, M., Hassreiter, S., Marsden, G., Allen, S.: Degrees of Sharing: Proximate Media Sharing and Messaging by Young People in Khayelitsha. In: MobileHCI 2012. ACM (2012)

A Field Trial on Mobile Crowdsourcing of News Content

Factors Influencing Participation

Heli Väätäjä[1], Esa Sirkkunen[2], and Mari Ahvenainen[1]

[1] Tampere University of Technology, Korkeakoulunkatu 1, 33720 Tampere
heli.vaataja@tut.fi, mari.ahvenainen@tut.fi
[2] University of Tampere, Kalevantie 4, 33100 Tampere
esa.sirkkunen@uta.fi

Abstract. We conducted a five-week field trial on mobile crowdsourcing of hyperlocal news content to 1) understand the readers' experiences and 2) explore factors affecting their participation. In the end of the study the participants were surveyed with an online questionnaire (17/104 respondents) and five participants were interviewed. Although respondents and interviewees were enthusiastic about the trial, the activity in the trial was low. Results indicate that participant characteristics (age, gender, participation motivations and hobbyist background in photography) and task characteristics in terms of the subjectively perceived task significance (possible impact on important issues in the environment or on community), task relevance (related to the background and participation motivation), and task engagingness have an effect on the participation. In addition, participation was influenced by the estimated needed effort vs. the expected benefit (monetary benefit or having a possibility to influence), vicinity to the assignment location, enjoyment of the activity, and the monetary reward. To plan and manage the crowdsourcing activity the news publishers need information about the characteristics of the participants, participation patterns and motivations that could be provided by the crowdsourcing platform.

Keywords: Crowdsourcing, user-generated content, hyperlocal, news, motivation, location, mobile, ubiquitous, reader, photo, Scoopshot.

1 Introduction

We have been witnessing a transformation of journalism as an industry and profession during the last ten years. It is evident that the old production models and professional cultures do not work in the age of digital, networked and socially shared media. A twofold process within the few last years has been taking place: User generated content (UGC) has been adapted more and more into news output and UGC creators have taken into account the editorial requirements of the newsrooms [1]. User-generated content (UGC) in this context refers to content that is created and contributed by non-professional reader reporters and reflects a certain amount of creative effort (adapted from [2]). So far UGC has been most often used in "soft and good news" when professional journalism is used in hard and bad news [3]. In case of eyewitness news on

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 54–73, 2013.

breaking topics, such as riots, disasters, and accidents, newsrooms have started to use social media as an important source of material [4-5].

There are especially rich possibilities to utilize UGC and crowdsourcing in the field of *hyperlocal* news, which are most often distributed free of charge to the readers in print or online. By hyperlocal news content we refer in this paper to content that pertains to a town, village, suburb, or other small, geographically defined community [6]. There are already numerous hyperlocal outlets for example in the UK which rely heavily on the readers' content, such as stories, photos and videos [6]. Especially hyperlocal newspapers collaborate with readers in content creation, because their editorial staff is few in number and the readers are well-informed of the hyperlocal, interesting and timely issues. Furthermore, other readers' material is considered trustworthy and authentic [1, 7-8].

One of the possibilities to get UGC for the hyperlocal news is mobile crowdsourcing. By mobile crowdsourcing we refer to a form of crowdsourcing, in which the initiator sends a task or makes available a task for voluntary undertaking by using mobile phones as an enabler for receiving or accessing the assignments and for submitting the news content (adapted from [9-10]). So far there is little prior research done concerning mobile crowdsourcing processes facilitated by news organizations publishing hyperlocal news, although the use of mobile assignments in news briefings have been reported [11]. This study addresses this gap. We report findings from a five week long field trial on crowdsourcing hyperlocal news content with a weekly mobile assignment. Findings are based on an online questionnaire (17/104 respondents) and interviews of five participants after the trial.

The goal of the research was to explore the characteristics of the participants and participation, as well as participants' experiences and identify the factors affecting participation. The goal for the news publisher was to use the findings in the development of their collaboration processes. In addition, the news publisher wanted to find out, whether young people could be activated to the collaboration by using a novel collaboration process and system. The main research questions were the following: How do participants experience the trial and the crowdsourcing solution? What factors affect the participation in the crowdsourcing activity? What implications are found for the development of the mobile crowdsourcing systems and processes?

To our knowledge, this is one of the first studies that reports on using mobile assignments for crowdsourcing in a real-world field trial carried out by a hyperlocal news organization. We present and discuss the findings on experiences and factors affecting participation and participation preferences. Based on the findings we provide implications for designing crowdsourcing processes and systems.

2 Related Work

Attitudes towards user-generated content (UGC) have changed during the last few years both on newsroom and reader reporters' side. Newsrooms and content creators have co-adapted their practices as newsrooms have created new processes for UGC and UGC creators have simultaneously learned the expectations of the newsroom [1].

Varying combinations of intrinsic and extrinsic motivations [13-14] are reported for reader participation in the context of news [15-17, 10]. In case of news related UGC there are differences in motivations to contribute depending on the addressees of shared content, the goal of sharing as well as the intended outlet of the submitted UGC [12]. The strongest reported motivations for contributors of stories at a German hyperlocal site *myheimat* are individually perceived creativeness, the fascination of publishing and the enjoyment of presenting one's own ideas to a larger public [12]. Respondents also report typical journalistic motivations, including balancing different perspectives and informing other people [12]. Similar motivations have been found for readers' photo content in case of hyperlocal news. In case of hyperlocal news, in which the submitter of a news photo gets a monetary incentive for a photo published in printed news, the top reported motivations are fun, an opportunity to get a monetary reward, sharing one's photos with others, and informing the wider public about local issues [10, 17].

Prior research reports on the following dimensions for users' participation and preferences in mobile crowdsourcing:

- context of use [10, 18, 20] - time, location, vicinity to task location, parallel other tasks or activity, social situation, technology/infrastructure
- assignment properties [10, 18-19] – validity (including time criticality), incentives, estimated time or other effort (e.g. travel distance) for carrying out a task
- assignment type [10, 18, 20] – photo, video, information acquisition, action (such as conducting an interview)
- patterns of planning and carrying out assignments [18, 20] – a priori planning
- mobile service(s) with features and functionalities [10, 18-19, 20] – reception or search of assignments (push, pull, location-based), submission of UGC
- use of location information [10, 18, 20] – geotagging, locating for assignments
- privacy issues [10]
- perceived risk, benefit and value of disclosing location information [10]
- organization type [10] – local or national news publisher
- incentives [10, 18]

Based on earlier literature on location-based crowdsourcing, the most preferred assignment types are photo [10, 18, 20], information acquisition [18], and video tasks [10]. Time-critical tasks and tasks with no incentive are not interesting [18] and solution should be achievable in 10 minutes [18].

Patterns of solving tasks are related to location and temporal dimensions. Tasks are preferred to be solved close to home [18], downtown [10] and in close vicinity of the current location [10, 18]. Tasks are searched for during breaks during the day at work or studies or while going home [18] and solved after work [18].

In location-based crowdsourcing – where tasks are related to a certain location and either searchable by location or at the location – searching for tasks that can be carried out in close vicinity of the home was preferred [18]. Map or augmented reality based solutions are accepted for planning of the activity [20]. If user's current location is used for sending assignments to the user, obfuscation of query for the user's location is preferred or using anonymous but precise locating [10].

We used these previously mentioned dimensions of participation preferences and findings in creating the questionnaire items and interview themes for this study.

3 Methods

3.1 Setup of the Study

The hyperlocal news publisher, Sanoma News, operating in Helsinki metropolitan area in Finland, trialed a crowdsourcing platform called Scoopshot. Scoopshot is offered by a third party that acts as an intermediary (marketplace) [21]. It enables registered users, solvers (scoopshooters), to submit photos and videos from smartphones to be sold for media companies (seekers) at the price set by the solver. Scoopshooters may also carry out tasks created by media companies and get rewarded for their accepted submissions. The incentive for simple tasks is usually at maximum two euros. Prior to the trial, the news publisher received reader's content sent as text messages (short comments), multimedia messages (photos), as well as via email and an online form (photos, stories and tip-offs).

3.2 Participants

104 participants were recruited with a website banner at the online news site (Omakaupunki.fi). The banner was visible to every third visitor of the website for two and a half days. In the end of the trial there were seventeen responses to the online questionnaire and five respondents were interviewed on a voluntary basis.

3.3 Apparatus

The participants used their own smartphones, either iPhone (9/17) or Android (8/17) during the trial. Those participants (3/17), who had not used Scoopshot prior to the trial, installed it to their smartphone when joining the trial. The rest joined the news publisher's community. Participants used Scoopshot mobile client for accessing the assignments, shooting photos and submitting user-generated photo content. The user interface of the Scoopshot mobile service client is illustrated in Figure 1.

At the time of the study the platform did not provide access to the information on the news publisher's community members, such as demographic information, activity logs other than actual UGC submitted by members in past three days, prior accomplishments of members etc. Contacting a member was only possible for current submitted UGC if the creator had allowed the phone number to be attached to the content. For professional photojournalists, the online platform enabled uploading of portfolios, creating automatically a profile based on accomplishments and their quality as well as being able to contact a specific person directly for undertaking a job.

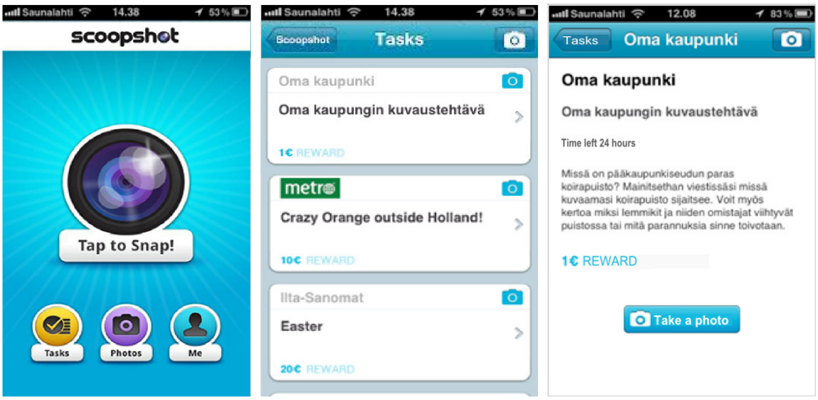


Fig. 1. Scoopshot mobile service client on iPhone. On the left, the opening page. In the middle the list of assignments. On the right an opened assignment.

3.4 Assignments

The participants were sent mobile photo assignments weekly, in total five (5) assignments (see Table 1) during March and April 2012. The editorial staff of the news publisher was responsible for planning of the assignments, sending them from the Scoopshot online portal, selecting the content to be purchased and publishing the content online and in the printed weekly news tabloid.

Table 1. Characteristics of the assignments and the activity during the trial (N = 104)

#	*Assignment theme*	*Validity*	*Scoopshot reward*	*Views*	*Submitters*	*Photos*	*Purchased*	*Printed*	*Reward for print*
1	Cleaning the environment	1wk	1 €	87	9	15	5	4	50 €
2	Tallinn shipping	12h	1 €	68	2	27	2	1	50 €
3	Noise barriers	2 d	1 €	63	2	2	0	0	-
4	Best dog park	2 d	1 €	64	2	2	2	1	2 movie tickets
5	Street conditions	1 d	50 €	56	3	7	1	1	-

The most important properties for each photo assignment were a textual description of the photo assignment to be carried out, the validity of the assignment and the reward information. Scoopshot reward was paid for all bought photos. In addition, the participants were rewarded for photos published in print media. The recipient group in all the assignments was news publisher's own Scoopshot community. As an example, the description of the assignment on dog parks (# 4) was the following:

"Where is the best dog park of the metropolitan area? Describe in your message where the dog park is located. You can also describe why pets and their owners enjoy the park or what improvements are wished for."

The degree of activity decreased during the trial (Table 1). The first news assignment "Cleaning the environment", was viewed by 87 participants. The last news assignment of the trial was viewed by 56 of the participants.

Fourteen (14/104) participants sent photos to the news assignments. Most of them (10/104) submitted photos to only one assignment. The most popular assignment based on the number of submitters was the first assignment on "Cleaning the environment" with nine submitters. "Tallinn shipping" received the highest number of photos, but there were only two submitters and one of them submitted 26 photos. Purchased photos, were published online and most of the purchased photos were used in localized printed versions of the "Vartti" tabloid.

3.5 Procedure

Participants received a weekly mobile assignment via the Scoopshot mobile client. Answering to the assignments was voluntary. In the end of the trial the participants received as a mobile assignment an invitation to participate to the online questionnaire. The questionnaire was open online for two and a half weeks. There were 17 responses to the questionnaire and respondents were compensated with two movie tickets (value 18 euros). The themes covered perceptions of the trial and assignments, participation preferences and background information. Five respondents volunteered to participate in a one hour interview. The five interviewees were compensated with two movie tickets. The semi-structured interviews were recorded and transcribed. In addition, log data of activity and generated content during the trial was collected.

3.6 Analysis of the Data

Questionnaire data was analyzed by descriptive statistics due to the small number of respondents. Interview data was analyzed by data-driven content analysis by identifying emerging themes in interviewees' perceptions and experiences.

4 Results from the Online Questionnaire

4.1 Respondents' Background

Demographics (N=17). Majority of the questionnaire respondents were male (16 male, 1 female). The youngest participant was 15 and eldest 53 years old (M = 28, SD = 10.4, Md = 26). Eight respondents were aged between 15 and 25 indicating that this type of activity attracts younger age groups to participation as was hoped for by the publisher. We use a grouping to two age groups in the analysis, that is, 15-25 years and over 25 years. Results based on age group are presented when clear difference between the groups was found and it is relevant for the presentation of the results.

Education. The most common highest level of education was vocational degree (7/17). For the rest of the respondents the highest level of education varied from comprehensive school to a higher university degree. Six of the respondents were students, but otherwise the occupations varied.

Activity in Photo and Video Shooting. Respondents were active photographers. Over half of them reported taking photos daily (10/17) and the rest of them (7/17) weekly. The frequency of video recording was lower. Only one of the respondents recorded videos daily, and the rest of them weekly (8/17) or monthly and less frequently (8/17).

Prior Usage of Scoopshot. Scoopshot was the most often reported outlet for UGC that can be rewarded for. Most of the respondents (14/17) had used Scoopshot prior to the trial, primarily for carrying out other assignments than news assignments.

Prior Sending of Photos to Media for Publication. Twelve (12/17) respondents reported sending reader's photos prior to the trial. They were asked to report in more detail how often during the last half a year they had sent photos to 1) the hyperlocal news publisher, 2) other news media and 3) Scoopshot prior to the trial. For most participants the frequency to send reader's photos to hyperlocal news or other media was low prior to the trial - either monthly, less than monthly or never. One of the respondents was an active weekly contributor to all three output channels. Another respondent was an active daily user of Scoopshot.

4.2 Assessment of the Trial and Scoopshot

Overall Assessment of the Trial. The respondents were asked to describe the trial by completing a sentence with 1-3 different endings. The question was as follows: "*How did you find the trial? Please, complete the sentence with one to three endings. In my opinion the trial was…*". There were altogether 38 different sentence completions of which 32 were categorized as positive. Trial was assessed as interesting (9/38), novel (8/38) and pleasant (6/38). The negative comments were related to wishing for more tasks or a lower number of tasks than expected. Overall, twelve respondents (12/17) described the trial only positively, and the rest both positively and negatively.

Suitability of Scoopshot. Scoopshot was assessed as suitable for receiving assignments as well as for submitting photos. Sixteen respondents (16/17) rated suitability for receiving the assignments to be between 8 and 10 (11-point scale: 0 = not suitable, 10 = extremely suitable). For sending photos the rating for suitability was somewhat lower as twelve respondents (12/17) rated the suitability to be 8 or higher.

4.3 Answering to the Assignments

Age as a Determinant. The older respondents were more active contributors than age group 15-25 years. Ten (10/17) respondents reported that they had submitted photos to the assignments. Eight of them were aged over 25 years. Six respondents of the seven who did not submit photos were aged between 15 and 25 years.

Reasons for Answering to the Assignments of the Trial. Those respondents who reported answering to the news assignments of the trial (10/17), were next asked a multiple choice question for their reasons to answer (Figure 2). Answering was motivated by an interesting assignment, enjoyment of activity when searching for a suitable subject to photograph and wanting to earn some money. Photography as a hobby, being close to a suitable location and easily finding a suitable topic to shoot the photo of were also mentioned. Other reasons mentioned in the open field were opportunity for influencing and gaining new experiences with technology.

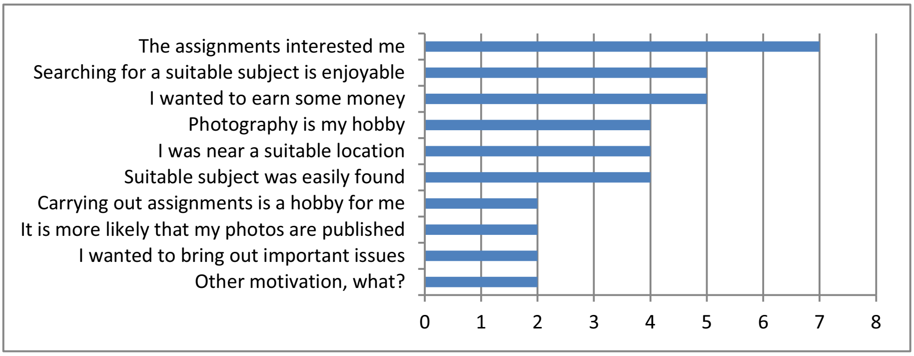


Fig. 2. Reasons for carrying out assignments (N = 10)

Reasons for Not Answering to the Assignments of the Trial. Those who did not answer to any of the news assignments in the trial (7/17) were asked for the reasons not to answer (Figure 3). Most often reported reasons for not carrying out assignments were that a suitable subject to shoot a photo was not found or the respondent was far from the location of the assignment. In addition, reason mentioned as "other" was related to the location, specifically, the respondent did not live in the area where the assignments were supposed to be carried out. Interestingly, none of the respondents reported as a reason for not answering the following options that were offered to them: "The assignments did not interest me", "I did not follow the offered tasks", and "I believe, that I will not get a reward".

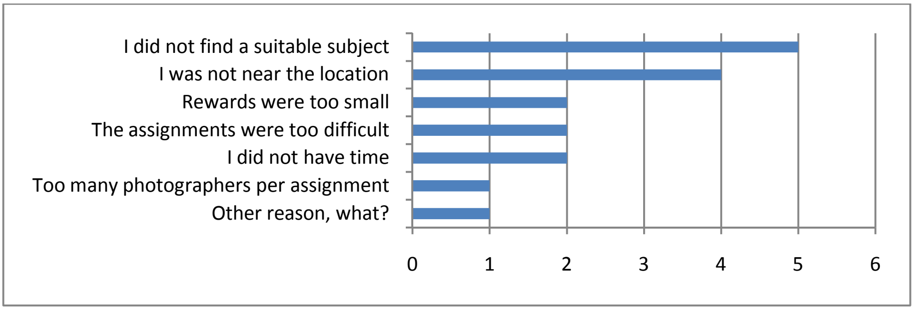


Fig. 3. Reasons for not carrying out assignments (N = 7)

4.4 Assessment of the Assignment Themes

Respondents were asked with a multiple choice question to assess the assignment themes in the trial. The question was asked in a form of sentence completion as follows: "*The themes of the assignments were voluntary cleaning, Tallinn shipping, noise barriers, the best dog park and street conditions. What do you think about the themes? In my opinion the themes were…*". We provided seven predefined attributes to be selected as the descriptive attributes and a possibility to choose "other" and to provide an open description for it (Figure 4).

To identify the predefined attributes (qualities) for this question, we asked nine externals (4 females, 5 males; age: 28 – 39 years) to describe with adjectives each of the five assignments of the trial. As an outcome 91 adjectives were collected. Adjectives with similar meaning were grouped and given a descriptive code. Seven most often reported attributes, with code counts varying from four to eight, were selected to be used in the questionnaire. This covered in total 44 of 91 original adjectives.

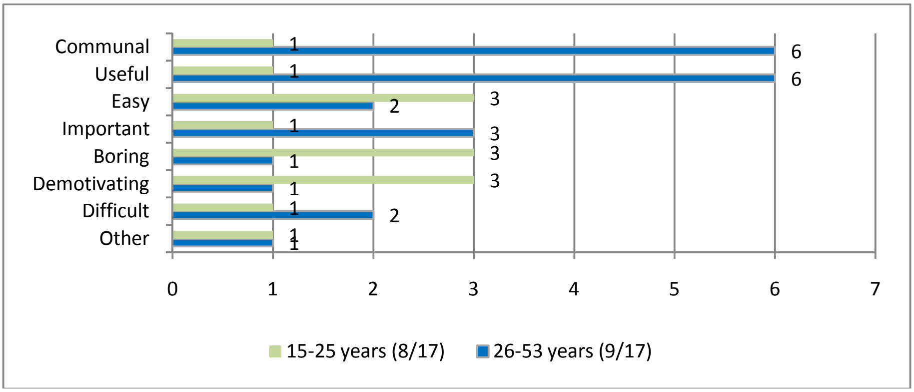


Fig. 4. Assessment of the assignment themes based on the age groups (N = 17)

Respondents assessed the themes as communal and useful, easy and important, but also boring and demotivating (see Figure 4). The young respondents (15-25 years), assessed the themes less positively than the older age group. Over half of the young respondents (5/8) assessed the themes as either boring or demotivating, whereas the older age group (over 25 years) assessed them communal and useful.

4.5 Assessment of the Other Assignment Properties in the Trial

Frequency of Assignments. Assignments in the trial were sent once a week. Respondents were asked "*In your opinion, were the assignments sent often enough*?" (Yes/No/I cannot say). Majority of the respondents (13/17) found that assignments were not sent often enough.

Suitable Length for Validity of the Assignments. The respondents were asked to rate the suitable length for validity of the sent Scoopshot assignments from four options: 1/2 day, 1 day, 2 days, and 1 week. A 5-point rating scale was used (too long, quite long, suitable, quite short, too short). The suitable length for the validity of an assignment was two days (14/17). Over half of the respondents (10/17) considered one day to be short and one week to be long. Half a day was considered to be short by all.

4.6 Dimensions on Participation Preferences

The respondents were asked with multiple choice questions preferences for assignment types, and on context related issues: in what situations, when and where they would prefer to carry out assignments [10, 22]. They were also asked to estimate from given choices the maximum distance willing to travel for and the time willing to spend on an assignment.

Preferred Assignment Types. The respondents were asked what kind of assignments they would be interested in carrying out (see Figure 5). All respondents preferred photo assignments and over half of them also video assignments (11/17) and information acquisition (10/17), such as finding out how much a kilo of new potatoes costs on a market place. There was no difference between the age groups in the preferences.

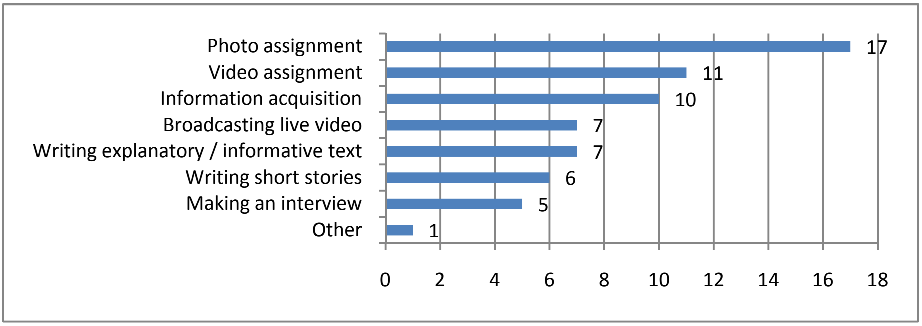


Fig. 5. Preference for carrying out different types of assignments (N = 17)

Preferred Situation (Task Context). Majority of respondents (15/17) were willing to carry out assignments on free time and over half of the respondents (11/17) when there is nothing else to do. Situations proposed by respondents in the open field were during hobbies and while waiting, i.e., when passing time.

Preferred Time (Temporal Context). Over half of the respondents (11/17) were willing to carry out assignments at any time. The most often reported specific times when willing to carry out assignments were weekends (8/17) and evenings (6/17).

Preferred Location and Vicinity (Spatial Context). All questionnaire respondents except one were willing to carry out assignments anywhere if they are in the vicinity of the assignment location (16/17) (Figure 6). Over half of the respondents (12/17) were willing to carry out tasks close to workplace, studies or near home. In addition, daily patterns of activities, such as commuting, generally during transitions between places and while shopping seem to be provide relevant times to carry out tasks.

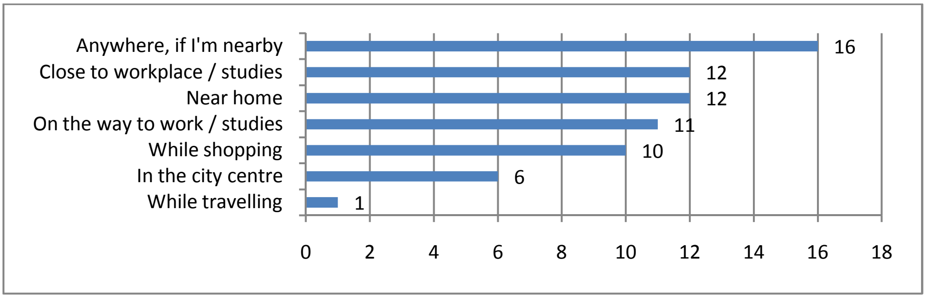


Fig. 6. Activities and locations to carry out assignments (N = 17)

Amount of Time Willing to Spend on an Assignment. Over half of the respondents (12/17) estimated willingness to spend even over a half an hour for carrying out an assignment, including the time for travel and submission (choices: 1-5 min, 10-30 min, less than an hour, more than an hour, other). However, this result may reflect more the maximum amount of time and be dependent on other factors, such as the interestingness of the assignment or the incentive.

Maximum Distance Willing to Travel for an Assignment. Choices given for estimating the maximum distance to travel were 1/2/5/10/15/over 15 kilometers. Over half of the respondents (9/17) were willing to travel at most five kilometers, whereas the rest reported willingness to travel even more than five kilometers.

4.7 Positive Attitudes Contradict the Disappointment at the Trial

In the end of the questionnaire the respondents were given a possibility to openly give feedback on the trial. Eleven respondents answered the question (11/17) giving twenty separate inputs. Four of the inputs were positive, ten negative and six neutral ideas and wishes. One of the respondents wished for continuation of the crowdsourcing activity, echoing the primarily positive assessments of the trial in the sentence completions reported earlier in subsection 4.3:

> *”This felt like sensible activity. More of this!”*

Eight of the ten negative inputs were given by respondents who did not carry out assignments or who were aged between 15 and 25 years. Two of the negative inputs were related to the rewards, being either too small compared to the amount of work or to the required distance to travel. Some respondents also expressed to be disappointed with the trial, as they had high expectations in the beginning, but due to the assignment themes their enthusiasm faded. There were wishes for broader topics and that some topics would be more challenging than the others. Topics concerning the youth were wished for, reflecting the low participation activity by the young participants:

> *“More tasks concerning the young!! We are the future, hey? :P”*

5 Results from the Interviews

5.1 Interviewees' Background

Demographics. The five volunteer interviewees were male and their age was between 26 and 53 (Md = 34) (Table 2).

Photographing. All interviewees had photography as a hobby. The primary motivations were intrinsic: they enjoy it because it is fun, creative and they constantly want to improve their skills. Three interviewees shot photos daily and two weekly (see Table 2). One interviewee found video shooting even more appealing than photographing and shared his videos in Youtube. Another one shared photos under Creative Commons license and his photos had been used for web publications, and magazines.

Prior Participation to News Making. One interviewee was an experienced contributor to news. He had started participation to news making almost 40 years ago by photographing and writing stories to the hyperlocal newspaper of his home town. The other interviewees had less participation experience to news making (see Table 2).

Table 2. Interviewees' background and characteristics

Age	Photographing as a hobby	Length of participation to news making	Created UGC types to news: Photo (P), Video (V), Story (S)			Preferred mode: Tasks (T), Spontaneous (S)		Intentional (I) vs. Randomly acting (R)		Channels used: Web form (W), MMS (M), Scoopshot (S)		
			P	**V**	**S**	**T**	**S**	**I**	**R**	**W**	**M**	**S**
26	>10 years	5 years	x	x		x	x		x	x		x
28	2 years	A few years	x			x			x		x	x
34	>10 years	Since Scoopshot	x			x		x				x
38	>10 years	Since the trial	x			x			x		x	x
53	40 years	Almost 40 years	x	x	x		x	x		x	x	x

Photographing was a pleasing way to participate in news making for all interviewees (see Table 2). Two of them also sent video material. One interviewee submitted a lot of spontaneous photos and chose the outlet media based on the photo's topicality. Two interviewees can be characterized primarily as intentional and actively seeking photojournalists, whereas three are primarily randomly acting photojournalists [15-16]. All interviewees were interested in carrying out assignments, although one of them preferred spontaneous photographing to tasks. He felt that by choosing the topics himself, he has more possibilities to influence.

5.2 Perceptions and Development Ideas on Scoopshot

Experiences with Scoopshot Prior to the Trial. Four interviewees liked the photo and assessment tasks of local companies that were initiated by a company providing directory services. One interviewee had carried out about one thousand tasks in one month. Another interviewee was not interested in simple tasks and hoped for themes with more challenge. Three interviewees had sold news photos via Scoopshot.

Suitability of Scoopshot. Scoopshot was found convenient, easy and effortless especially for assignment-based participation supporting the questionnaire findings. Moreover, it was found cost-effective as submitting is free of additional charge with a cellular data subscription or via a wireless network.

Ideas for Future Development. Some ideas for developing a mobile crowdsourcing platform arose in the interviews:

- *Community like features*: A possibility to see photos that other solvers have submitted to the assignments.
- *Incentives*: A possibility to submit news photos with no expected reward.
- *Support for planning and carrying out tasks*: Tasks available on a map view.
- *Limiting the number of solvers*: Assignments could be available to first five solvers close to the assignment location and if they are not able to submit sufficient content, the assignment would be opened to other solvers in the area.
- *Support for using solvers with good reputation*: An editor should be able to allocate an assignment to a reader reporter that is already known of one's good work. This calls for an implemented mechanism for reputation building.

Attitudes towards the Use of Location Information. Scoopshot uses the user's location information when accessing the tasks and it geotags the photos. Four of the five interviewees were positive towards the use of location information in news reporting reflecting the findings of earlier studies in this context [10, 20]. Four interviewees (4/5) found locating useful in general and they also used automatic geotagging of their photos. One interviewee described how geotagging provides added value for the viewers of the UGC by enabling location-based search for UGC and he therefore added the location information if it was missing. On the other hand, one interviewee explained how with some photos giving the exact location could be harmful, when photographing protected animals, for example. Instead of automatic geotagging, he added in these cases an approximate location to protect the source.

5.3 Participation in the Trial

Motivations to Participate in the Trial. The most common motivation to take part in the trial was interest in Scoopshot and assignments (4/5). One of the interviewees described, how he feels about receiving assignments and his participation:

"… always when I notice or the phone notifies when there is a new assignment, I check it. I find it fun. It makes one feel a bit important, or like own contribution really is somehow useful then. To me it gives pleasure. .. It is like playful, that you wait for an assignment to carry out and it feels really good." Male, 34

Carrying out assignments was described by the interviewees as fun, playful, and giving challenges and an opportunity to learn and experience something out of the ordinary. Interviewees also mentioned it to be a change to their own photographing and a pleasant additional activity. Interviewees also seemed to interweave the tasks to their everyday activities, such as when cycling to or from work, when walking a dog or when in transition between places during the work day as has been reported earlier for mobile gaming [23].

Feeling of Succeeding. A published photo or being able to influence with a photo gave a feeling of succeeding. Three of the interviewees felt that they had succeeded if their photos were published (3/5). One interviewee described how he had succeeded, if he himself was satisfied with the photo, whether the photo got published or not. For the interviewee with longest experience it was important to influence with his material and get something fixed, to repair a fault or to cause a change. He described:

"And I don't know, whether it creates contentment, that your own photo is in some newspaper, even that, you are used to it. But when you notice, as when I participated to the trial with assignment on potholes due to frost damage, when it [photo] was published in the newspaper, it was fixed within a week. It is, that it has influence." Male, 53

In addition, the expressed interest by others towards their photo was a sign of succeeding. This interest could be the use of the photo for news, as a tip-off or idea for a news story, or someone wanting to use the photo for their other purposes. Furthermore, comments to their photo on the news site were a sign of success, showing that the photo raised interest.

Money as a Motivational Factor. Money and getting an incentive was not the most important motive to participate for the interviewees. However, the little extra income was a nice bonus and acknowledgement for the effort. None of the interviewees mentioned about the reward being too small, rather the other way around. One interviewee was willing to upload news photos to Scoopshot with no expected reward and another would carry out assignments even without any reward if the assignment was interesting enough. Altogether three interviewees expressed that they considered the possibility to influence and cause overall benefit with their photos more important than the reward. One of them commented the Cleaning the environment assignment as follows: *"I think I got 50 euros from it. I was more satisfied, that it was cleaned up."* Male, 53. The active background in photography and interviewee's participation motivations may influence the views on money and its role as an incentive.

Overall Assessment of the Trial. Interviewees found the trial interesting and fun and they were willing to continue participation. It was mentioned that having a clear assignment makes one think and approach it as a challenge to solve and that even competitive aspects came into play.

The most negative issues mentioned by participants were short the validity of assignments, the worn out and much used assignment topics and the impacts of reader participation. As negative impacts of reader participation two interviewees were worried about professional photographers' salary level being reduced or detracting from their jobs. One interviewee feared for the future impact on his work at the rescue services. He explained that if all reader reporters want to snap a photo of a fire, it will distract the rescue workers' operation on the scene of an accident

Comparing Interview and Questionnaire Results. The most notable difference in interview and questionnaire responses is the distance willing to travel for an assignment (Table 3). In the questionnaire all interviewees were ready to travel five kilometers or more for an assignment, but in the interviews three of them preferred assignments on their way or needing to travel only a short distance. This could be the case also for other questionnaire respondents and similarly, the time willing to spend for carrying out tasks might be overestimated in the questionnaire results.

Table 3. Comparison of questionnaire responses and interviewees

Topic	Questionnaire respondents	Interviews	Interviewees in questionnaire
Were the assignments delivered often enough?	No (13/17)	No (5/5)	No (4/5)
Assessment of assignment topics (most imp.)	Communal (7/17), useful (7/17)	N/A	Useful (4/5), communal (3/5)
Willingness to travel for an assignment	5 km or more (14/17)	On the way or short distance (3/5)	5 km or more (5/5)
Suitable assignment validity	2 days (14/17)	Min 1, max 7 days (5/5)	1-2 days (3/5)

6 Discussion

The field trial showed the importance of understanding the participant characteristics, perceived task characteristics, and participation preferences as factors influencing participation. Understanding these factors supports the planning and managing of mobile crowdsourcing activities within the context of hyperlocal news publishing, including the design of assignments and forms of participation. The crowdsourcing platform should support this understanding by providing information on the participant characteristics and participation patterns, as well as accumulate the collected knowledge and make it available for the newsroom. We next discuss the findings based on the trial and provide implications for designing crowdsourcing systems and processes, including assignment characteristics.

Perceptions on the Trial and Its Implementation. The questionnaire respondents were enthusiastic about the crowdsourcing activity and the trial with Scoopshot as the crowdsourcing platform. In the sentence completions the trial was described as interesting, novel and pleasant. Scoopshot was also assessed to be suitable for the mobile assignment based collaboration. However, the actual activity in the trial was low. The following reasons for this can be found in our findings.

First, the themes of the assignments were perceived positively as communal and useful primarily by the respondents from the older age group (over 25 years), whereas the younger age group (15-25 years) assessed the themes primarily as demotivating or boring. Interviewees described the themes as worn out and much used. The relevance of the themes to the participant groups as well as the engaging implementation of the participation needs to be paid attention to when planning the activity and assignment themes. In addition, the characteristics of the participant groups may vary depending on the channel chosen for reader participation and this should be taken into account in the planning of the activity and assignments.

Second, those respondents who answered to the assignments (10/17) reported as top reasons for carrying out the assignments the interestingness of the assignments (7/10), enjoyment of searching for a suitable subject to photograph (5/10), and wanting to earn some money (5/10). The top reasons reported for not carrying out assignments (7/17) were not finding a suitable subject (5/7) and not being in the vicinity of the assignment or suitable location (4/7). In addition to the task relevance and engagingness of the assignments our results emphasize the enjoyment of the activity, monetary incentives and vicinity to a suitable location as important factors similarly to previous studies in mobile crowdsourcing [10, 17-18].

Third, the frequency of the assignments, which was in the trial one assignment per week, was found too low. Assignments were wished more frequently, with a preferred validity of two days, and several assignments were wished to be available at the same time to be able to pick from a set of assignments. This would allow more readers to participate and on the other hand enable choosing those of most interest.

Participation Preferences. Based on earlier research [10, 17-18]. we studied the following factors that influence participation preferences: 1) assignment types, 2) situations (task context), 3) time (temporal context), 4) location (spatial, i.e., physical context), and 5) effort needed in terms of a) time spent and b) distance travelled. Similar to previous studies photo assignments were the most preferred assignment type followed by video and information acquisition tasks [10, 18]. Respondents preferred carrying out assignments on free time and when nothing else to do, as reported previously [10, 18]. Furthermore, participants seem to interweave the assignments with their daily patterns of activities, such as when commuting or in transition to or from work, studies and home, when walking a dog or cycling, and while shopping. Similar findings have been reported previously for ubiquitous gaming [23] and in mobile crowdcourcing [18]. Augmented reality based solutions have been suggested for implementing support for readers when searching for tasks and planning own activity in hyperlocal news making [20]. Mobile client could provide support for making a plan and a route to follow, by being able to choose assignments to a task list and create a route or a map for oneself of the preferred assignments.

Although in this study the respondents were willing to carry out assignments at any time, the more specific reported times by the respondents, that is, weekends and evenings are similar to results reported in previous studies [18, 10]. Questionnaire respondents were willing to carry out assignments if they were close the assignment location by, but specifically when close to workplace, studies or home. This indicates clearly the need for a possibility to search for assignments near the places relevant for the user as reported by Alt et al. [18].

We also studied the amount of effort willing to invest in the crowdsourcing activity, specifically in terms of 1) the time spent on an assignment including travel and submission, and 2) the distance travelled for the assignment. Questionnaire respondents reported willingness to invest relatively high effort in carrying out the tasks compared to earlier findings [10,18]. The results in our study on the willingness to put effort in carrying out the assignments therefore may be affected by other factors, such as the interestingness of the topic, the incentive and be valid only for a maximum effort willing to invest which may be affected by other factors as well. This seems to be the case based on the interview results, in which the vicinity to assignment location was emphasized. In addition, in the open comments of the questionnaire results the balance between the reward and the needed effort was brought up.

Motivations to Participate. The interviews of five participants with photography as a hobby revealed several motivations to participate in news making and crowdsourcing of news. These included the enjoyment of the activity and the variation it brings to their hobby, the opportunity to learn and experience something out of the ordinary, and the challenge and competition in solving the given assignment. Furthermore, the feeling of personal importance and value of one's contribution to the news making as well as the desire to influence something of interest to oneself were raised in the interviews. These motivations reflect the earlier findings on motivations to participate in news making [12, 16-17]. The participation of the interviewees seems to be motivated primarily by intrinsic motivations, although the monetary reward was described as a little extra income that was a nice bonus for the effort invested.

For most of the interviewees the printed photo was an important feedback for success, giving satisfaction of one's accomplishment, as well as the comments on one's photo were a sign of success and a source of satisfaction. On the other hand, one of the interviewees emphasized how he was satisfied, when the photo had influenced the repairing of some issue rather than gaining a personal monetary benefit for the photo. This brings up the wider significance of own activity as an emerging theme. However, interview results are representative only for the special group of hobbyist photographers, who volunteered to be interviewed, rather than for a wider participating crowd.

The results from the questionnaire and interviews seem to indicate that the participant groups have different profiles in relation to the motivations and preferences to participate when mobile assignments are used. Some participants seem to take the tasks more seriously, almost as professional photojournalists, being critical about the quality of their work and wanting to improve their skills. Others seem to be motivated if the tasks touch issues that are important for them as citizens or residents of a certain area. The third group could be those that take part in order to earn some money or have fun, for example. Further studies in this area are therefore needed to support the planning of the crowdsourcing activities and collaborative news making.

Other Issues to Address in the Design of Crowdsourcing Systems and Processes. First, the newsroom needs to gain an understanding on different participant groups and their participation patterns and activity. This should be supported by the crowdsourcing platform or by other means. This calls for being able to collect, analyze and use collected information of the participants and submitted content to create knowledge on the participation. If an external crowdsourcing platform by a third party company is used, this company could provide the information as a service to the customer company. Second, the news publisher should provide multiple channels for participation to ensure participation independent of the socioeconomic background or technology savviness. To ensure equality and enable the readers to voice their views independent of their socioeconomic and demographic background, multiple channels of participation give possibilities for more diverse groups to participate and bring out issues important to them. Further issues that were suggested by participants to the developing the mobile crowdsourcing platform: a) community like features to support viewing of other users' content, b) a possibility to submit without any expected reward set as a price of the photo, c) limiting the number of solvers in the vicinity of the assignment location and d) creating a reputation system for the solvers that can be used by the newsroom staff to choose solvers with a good reputation to ensure the quality of the submissions.

Limitations and Future Research. Although the trial was limited in terms of the number of participants, its length and the number of sent assignments, the findings provide insights that can be taken into account when developing mobile crowdsourcing processes and enabling systems for it. There are several different paths to continue the research: participant characteristics, including socio-demographics and participation motivations in collaborative processes in news making, the use of incentives in this setting, developing the technological support, systems and processes for collaborative news making, ideating new enablers and participation forms for fun and playful experiences for participants, as well as further studies, experiments and trials in real-world on the factors affecting participation to enable successful use and management of crowdsourcing and to enable equal opportunities for readers to voice their views.

7 Conclusions

Findings indicate, that mobile crowdsourcing is a feasible option to be used in hyperlocal news making, that could 1) attract new participant groups to collaboration, 2) support a fun and playful way of participation, 3) give a possibility to influence, and 4) give an opportunity for little extra income. Results indicate that task characteristics in terms of the subjectively perceived significance (e.g. possible impact on important issues in the environment or on community), relevance (related to participant characteristics and participation motivation) and engagingness (fun and motivating) have an effect on the participation. Furthermore, participant characteristics, specifically age, gender, participation motivations and photography as a hobby seem to affect the participation. In addition, participation was influenced by the estimated needed effort vs. the expected benefit (monetary benefit or having a possibility to influence) the vicinity to the assignment location, enjoyment of the activity, and the monetary reward. Hobbyist background in photography seems to be connected to the interviewees' intrinsic participation motivations. As carrying out the tasks seems to interweave into daily activities, mobile user's activity should be supported by enabling the planning of the own participation activity by the crowdsourcing platform.

Crowdsourcing of photo content for news can be seen as a form of crowdsourcing of creativity [24]. Combined with a potential monetary reward the particular form of crowdsourcing studied in this trial combines the intrinsic and extrinsic participation motivations with varying emphasis depending on the participant characteristics.

Our findings show, that it is important for the news publishers that the crowdsourcing processes and systems not only support the assignment related activity but also support understanding the motivations, participant characteristics as well as the underlying determinants and patterns of participation.

Acknowledgements. Special thanks to Tuukka Muhonen and the editorial staff at Sanoma News/Vartti for their cooperation in the study. This research was supported by TEKES as part of the Next Media programme of TIVIT and by the Finnish Doctoral Program on User-Centered Information Technology (UCIT).

References

1. Hänska-Ahy, M., Shapour, R.: Who's Reporting the Protests? Converging Practices of Citizen Journalists and Two BBC World Service Newsrooms, from Iran's Election Protests to the Arab uprisings. Journalism Studies 14(1), 29–45 (2013)
2. Wunsch-Vincent, S., Vickery, G.: Participative Web: User-Created Content [online document]. Organisation for Economic Co-operation and Development (OECD), Directorate for Science, Technology and Industry (2007), http://www.oecd.org/dataoecd/57/14/38393115.pdf
3. Paulussen, S., D'heer, E.: Using citizens for community journalism. Journalism Practice, 1–16 (2013)
4. Diakopoulos, N., De Choudhury, M., Naaman, M.: Finding and assessing social media information sources in the context of journalism. In: Proc. CHI 2012, pp. 2451–2460. ACM (2012)
5. Mortensen, M.: When citizen photojournalism sets the news agenda: Neda Agha Soltan as a Web 2.0 icon of post-election unrest in Iran. Global Media and Communication 7(1), 4–16 (2011)
6. Radcliffe, D.: Here and Now: UK hyperlocal media today. Nesta (2012), http://www.nesta.org.uk/library/documents/Here_and_Now_v17.pdf
7. Lai, S.: Iconic Images and Citizen Journalism. In: Proc. iConference 2011, pp. 702–703 (2011)
8. Wardle, C., Williams, A.: ugc@thebbc: Understanding its impact upon contributors, non-contributors and BBC News [online document], Cardiff, UK, Cardiff School of Journalism, Media and Cultural Studies, http://www.bbc.co.uk/blogs/knowledgeexchange/cardiffone.pdf
9. Estellés-Arolas, E., González-Ladrón-de-Guevara, F.: Towards an integrated crowdsourcing definition. Journal of Information Science 38(2), 189–200 (2012)
10. Väätäjä, H., Vainio, T., Sirkkunen, E.: Location-based crowdsourcing of hyperlocal news: dimensions of participation preferences. In: Proc. GROUP 2012, pp. 85–94. ACM (2012)
11. Väätäjä, H., Egglestone, P.: Briefing news reporting with mobile assignments: perceptions, needs and challenges. In: Proc. CSCW, pp. 485–494 (2012)
12. Fröhlich, R., Quiring, O., Engesser, S.: Between Idiosyncratic Self-Interests and Professional Standards: A Contribution to the Understanding of Participatory Journalism in Web 2.0. Results from an Online Survey in Germany. Journalism 1464884912442282 (2012)
13. Ryan, R., Deci, E.: Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology 25, 54–67 (2000)

14. Nov, O., Naaman, M., Ye, C.: Analysis of participation in an online-photo-sharing community: A multidimensional perspective. J. of the Am. Society for Information Systems and Technology 61(3), 1–12 (2010)
15. Brabham, D.C.: Moving the crowd with iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. First Monday 13(6) (2008)
16. Buehner, T., Jones, J.: Exposing the digital newshound: A study of the values, influences, and characteristics of creators of citizen-journalistic photographic content. In: 2011 Midwinter Conference of the Association for Education in Journ. and Mass Communication (2011)
17. Väätäjä, H.: Readers' motivations to participate in hyperlocal news content creation. In: Proc. GROUP 2012, pp. 309–312. ACM (2012)
18. Alt, F., Shirazi, A.S., Schmidt, A., Kramer, U., Nawaz, Z.: Location-based crowdsourding – Extending crowdsourcing to the real world. In: Proc. NordiCHI 2010, pp. 13–22 (2010)
19. Väätäjä, H., Vainio, T., Sirkkunen, E., Salo, K.: Crowdsourced news reporting: supporting news content creation with mobile phones. In: Proc. MobileHCI 2011, pp. 435–444 (2011)
20. Väätäjä, H., Ahvenainen, M., Jaakola, M., Olsson, T.: Exploring Augmented Reality for User-Generated Hyperlocal News Content. In: CHI 2013 Extended Abstracts, Paris, France (2013)
21. Chesbrough, H.: Open business models – How to thrive in the new innovation landscape. Harvard Business School Press (2006)
22. Wigelius, H., Väätäjä, H.: Dimensions of context affecting user experience in mobile work. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 604–617. Springer, Heidelberg (2009)
23. Bell, M., et al.: Interweaving Mobile Games with Everyday Life. In: Proc. of CHI 2006, pp. 417–426 (2006)
24. Schenk, E., Guittard, C.: Towards a characterization of crowdsourcing practices. Journal of Innovation Economics 1(7), 93–107 (2011)

Nudging People Away from Privacy-Invasive Mobile Apps through Visual Framing

Eun Kyoung Choe[1,2], Jaeyeon Jung[1], Bongshin Lee[1], and Kristie Fisher[3]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
[2] The Information School, University of Washington, Seattle, WA 98195, USA
[3] Microsoft Studios, One Microsoft Way, Redmond, WA 98052, USA
eunky@uw.edu, {jjung,bongshin,kfisher}@microsoft.com

Abstract. Smartphone users visit application marketplaces (or app stores) to search and install applications. However, these app stores are not free from privacy-invasive apps, which collect personal information without sufficient disclosure or people's consent. To nudge people away from privacy-invasive apps, we created a visual representation of the mobile app's privacy rating. Inspired by "*Framing Effects,*" we designed semantically equivalent visuals that are framed in either a positive or negative way. We investigated the effect of the visual privacy rating, framing, and user rating on people's perception of an app (e.g., trustworthiness) through two experiments. In Study 1, participants were able to understand the intended meaning of the visual privacy ratings. In Study 2, we found a strong main effect for visual privacy rating on participants' perception of an app, and framing effects in a low privacy rating app. We discuss implications for designing visual privacy ratings, including the use of *positive visual framing* to nudge people away from privacy-invasive apps.

Keywords: Visual framing, privacy, privacy metrics, rating; nudge, framing effect, valence, positive framing, negative framing, Mechanical Turk.

1 Introduction

Application marketplaces (a.k.a. app stores) have become the mainstream channels to distribute applications onto smartphones. Over 700,000 apps have been published both in the Apple *App Store* and in *Google Play* as of Dec. 2012 [1,4]. However, this abundance of choice comes with consequences; app stores are not free from privacy-invasive apps that collect personal information without sufficient disclosure or user consent. An investigation of 100 popular apps (for iPhone and Android) shows that many of these apps collect personal information such as location and the phone identification numbers, and some apps share these data with third parties without proper user consent [28]. For example, Path, a popular social media app, was recently found to transmit the user's contacts stored on the phone without explicit permission [32].

When shopping for an app in the app store, people have an opportunity to compare different apps that provide similar functionality. For example, searching for "weather" returns more than 1,000 results in Google Play. Clicking an app of interest, people can

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 74–91, 2013.

view detailed information, such as description, screen shots, user ratings, and reviews of the app. In this detailed view, some app stores provide privacy-related information regarding which types of data the app may access; however, prior research has shown that people do not read the permission warnings, and even if they do, they do not understand what the information means, as the terms are vague and confusing [12,16]. Nonetheless, a recent survey reports that 30% of survey respondents had uninstalled apps found to collect personal information that they did not want to share [6].

The goal of this work was to explore novel ways to nudge people away from privacy-invasive apps when they search for and compare apps to install. Specifically, we created *visual representations* of an app's level of privacy protection and investigated how the visual representations influence people's perception of an app. Similar to a movie critics' rating [23], we created visuals for a *privacy critics' rating* of an app, conveying how privacy-preserving or privacy-invasive the app is (Figure 1).

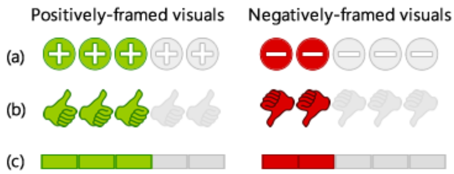


Fig. 1. Visual Framing. Positively-framed visuals of a rating of 3 (left column) is semantically equivalent to negatively-framed visuals of a rating of 2 (right column).

To create influential, persuasive visuals for privacy ratings, we leveraged the well-known "Framing Effects" [30]. The key idea is that people's decisions, in part, depend on the way problems are stated (e.g., positively or negatively). A classic example is how a doctor describes the odds of a grueling operation: many would prefer to choose an operation of which an outcome is "90 out of 100 are *alive* after five years" than "10 out of 100 are *dead* after five years" [21]. Even if these two phrases contain the same information, people—even experts (i.e., doctors)—are systematically subject to framing effects. In addition, framing effects occur without people knowing that they are being affected by it. People are susceptible to the framing as long as they understand the *valence* of an option—whether something is good or bad—without necessarily understanding what makes the option appealing.

Our goal was therefore *not* to make people *understand* the details of how an app's privacy rating is calculated or to help them make an informed decision. Rather, in this work, we investigated whether we can leverage framing effects to nudge people away from privacy invasive apps using visual representations. We used visual elements such as colors and symbols to make the valence information even more salient than text-only valence descriptions. To study the effect of the visual framing of a privacy rating, we first created semantically equivalent visuals for a privacy rating that highlight either the positivity or negativity of the rating, respectively. Then, we measured the effect of the positively- and negatively-framed visuals for privacy ratings on how people perceive smartphone apps with/without a *user rating*. Privacy rating is just one aspect among many other attributes of an app whereas the user rating reflects how general audience holistically thinks about the app based on their individual experiences. Furthermore, people self-reported [8,11,12] that they consider privacy-related information far less important than user ratings in making app choices. Therefore, we considered the effect of the privacy rating in conjunction with the user rating.

Our contributions are threefold. First, we detail visual attributes (e.g., color) and semantics (e.g., valence, sign) that contribute to visual framing. Second, we investigate if framing effects transfer to visual representations of a privacy rating and shift people's perceptions of apps. Third, based on the lessons learned from two experiments, we discuss design implications for visualizing privacy ratings of smartphone apps in such a way to nudge people to avoid privacy-invasive apps.

2 Related Work

2.1 Mobile App Privacy and Permissions

Recent studies reveal the limitations of existing approaches to presenting privacy-related information on mobile phones. Looking into Android's permission interface, two studies show that people pay little attention to permission requests when installing an app and have a poor understanding of what each permission means [12, 16]. Semi-structured interviews of 20 Android users discover that participants do not understand Android permissions, which were described as "confusing, misleading, jargon-filled, and poorly grouped" [16]. An online survey of 308 Android users reports that only 17.5% of respondents reported looking at permissions during their last app installation [12]. Moreover, only 3% of respondents could provide correct answers for permission comprehension questions [12]. On the contrary, the same study reports that 71% of respondents looked at some type of user reviews before installing an app. Similarly, Chin and colleagues reveal that participants often rely on user ratings when deciding an app to install [8].

Realizing these limitations, researchers have made an effort to design simple and easy-to-understand visual representations of privacy-related information. For example, Kelley and colleagues show that presenting online privacy policies using standardized tables improves people' comprehension of the privacy policies [15]. Cranor and colleagues use a symbol (called Privacy Bird) to indicate whether a website's privacy policy matches a user's preference [9]. A follow-up study designs "privacy icons" and demonstrates that when the privacy icons were presented, participants were willing to pay a premium to purchase items from the websites that better protect users' privacy [29]. Similar to these studies, we created simple visual representations for privacy information. However, our work is different from the previous work as our goal was not to improve people's comprehension of the exact details of privacy policies, but rather to create visual representations that can affect how people perceive an app and as a result can nudge them away from privacy-invasive apps.

One study by Lin and colleagues [20] addresses the limitations of Android's permission interface using crowdsourcing. The authors propose a privacy summary that highlights the use of permissions that did not match other people's expectations. To capture people's expectations, the authors conducted online surveys, recruiting respondents from Amazon Mechanical Turk. Such measured expectations can be useful for creating an app's privacy rating, which our work could leverage. In comparison, our work examines *framing effects* in *visual representations* for privacy information with the aim to identify what makes certain visuals more influential and persuasive than others and leverage those properties to influence user perceptions.

2.2 Theoretical Framework: Framing Effects

A few recent studies apply behavioral economics theories in designing persuasive technologies. Lee and colleagues [17] apply three persuasion techniques drawn from behavioral economics—the default option strategy, planning strategy, and asymmetric choice strategy—in designing choices for healthy eating. Inspired by their work, we explore a novel approach to employ *framing effects* to nudge people away from privacy-invasive apps. Humans have been thought to make rational choices, of which the core principles include *invariance* [31]. Invariance requires that two versions of a choice that are recognized to be equivalent should yield the same preference regardless of the manner in which they are described. However, Tversky & Kahneman [30] reveal that presenting the same option can alter people's decisions when varying the framing of acts, contingencies, or outcomes.

Since Tversky & Kahneman first discussed framing effects explaining how valence influences people's willingness to take risks, framings have been studied using *text descriptions* in many domains including privacy domain [13]. To better understand when and why different types of framing will have an effect, Levin and colleagues develop a typology of framings, and distinguish between three different kinds of framings—risky choice, attribute, and goal framing [19]. Our work is particularly inspired by the *attribute framing*, which manipulates the valence of a *single* attribute within a given context. For example, an attribute of ground beef can be labeled as either "75% lean" or "25% fat," and a study shows that people favor the former even though the two labels convey the same information [18]. Framing effects occur because people tend to be somewhat mindless when consuming information in daily life [27]. Therefore, when applied properly, framing can be used as a powerful nudge to influence people to make a better decision [27].

2.3 Nudging for Privacy Decision

Acquisti & Grossklags discuss the relevance of framing to privacy research [2]. In [3], Acquisti coins the term, "nudging privacy," and discusses how existing research in behavioral economics and psychology can help better understand privacy decision making. Furthermore, Balebako and colleagues broadly discuss how their ongoing work in location sharing can leverage nudging interventions to help people make better privacy decisions [5].

In HCI research, the term “nudging” has been used in the broad sense outside of behavioral economics as a means to influence people’s behaviors, such as shopping [14] or exercise habits [25]. When designing a system whose goal is to nudge people toward a certain direction, the system might undermine the decision making power of individuals. Thaler & Sunstein [27] state that a nudge should be “easy and cheap to avoid.” Thus, banning privacy-invasive apps entirely from app stores is not a nudge. We also note that the use of “framing” as a design approach for nudging might impose ethical concerns as discussed in [7]. We share the same vision with the prior research mentioned above in that leveraging framing can be a promising approach to nudge people toward better privacy decisions. However, these techniques should be applied with care so as not to limit people’s freedom to choose.

3 Research Questions and Experimental Setups

We explore the following research questions (RQ) through two experiments:

> RQ1. Can we create complementary visual framings that convey semantically equivalent privacy rating information?
>
> RQ2. Do complementary privacy ratings have similar influence on how people perceive an app? If not, is a negative visual framing more effective in nudging people away from privacy-invasive apps?
>
> RQ3. Do people’s perceptions of an app change if a privacy rating is accompanied by a user rating?

For each study, we created an online experimental setup and recruited participants using Amazon Mechanical Turk (MTurk). We made sure no one could participate more than once by setting a cookie in participants’ browser and by removing answers from duplicate MTurk IDs. The studies were available only to U.S. and Canada residents with at least a 95% approval rate (through a screening option that MTurk provides). We compensated MTurk participants $0.50 USD per survey for Study 1, and $1 USD per survey for Study 2.

4 Study 1: Creating Complementary Visuals

As a first step toward identifying visuals that have higher nudging power, we investigated whether it is viable to create complementary visual framings that convey semantically equivalent privacy rating information. We designed two sets of icons: positively-framed (PF) icons using a green plus sign (+) and negatively-framed (NF) icons using a red minus sign (-) (Figure 1-a). Because most of the ratings (e.g., Amazon star rating) use PF icons (i.e., the more stars, the better), we speculated that people were already familiar with the PF icons but not with the NF icons. We measured the level of comprehension of the PF and NF icons after short training. Similar to prior framing studies, we conducted a between-subjects experiment with two groups: PF icon group and NF icon group.

4.1 Survey Content

We created online surveys for PF and NF conditions. The surveys consisted of a privacy rating description page and two sets of eight icon comparison questions (the first set was for training). On the first page, we introduced a hypothetical privacy metric called "Privacy Critics' Rating." To help participants understand valence of the metric, we provided the following explanation:

> Imagine you are deciding whether to install an app on your smartphone. **Privacy experts** ran a series of tests to assess the app's level of privacy. These tests measured how likely the app is to capture your personal information and to share/sell your information to advertisers or other partners.

We manipulated the description and a screenshot of a fake game app, "Tic Tac Toe," in the following manner:

PF condition: We explained that the privacy rating scale represents the proportion of tests that the app **passed**. Therefore, if an app has more green-colored (PF) icons in the rating, it is more privacy-preserving. The screenshot included a privacy rating of 4 (out of 5).

NF condition: We explained that the privacy rating scale represents the proportion of tests that the app **failed**. Therefore, if an app has more red-colored (NF) icons in the rating, it is more privacy-invasive. The screenshot included a privacy rating of 1 (out of 5).

According to our description, a privacy rating of 4 (out of 5) in the PF condition is equivalent to a privacy rating of 1 (out of 5) in the NF condition. After the description, we asked participants to answer two sets of eight icon comparison questions *as accurately and quickly as possible*. We showed two different privacy ratings and asked which of the two privacy ratings is more privacy-invasive or privacy-preserving (Figure 2). These two questions appeared in a random order. We placed one question per page; participants had to click a "next" button in order to proceed to the next question, which allowed us to measure a task completion time for each question.

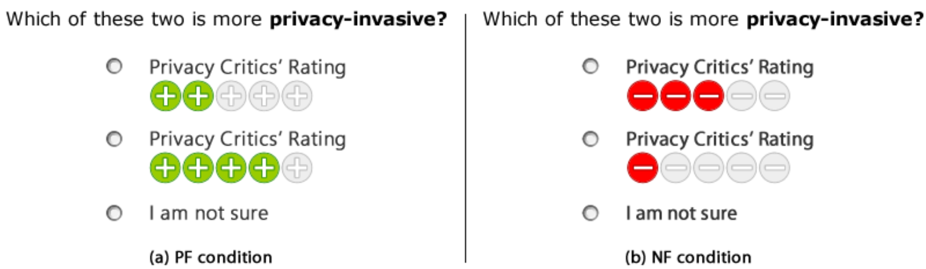


Fig. 2. An example of the icon comparison questions and the complementary visuals for the PF (a) and NF (b) conditions

4.2 Measures

The dependent measures were accuracy (i.e., the number of correct responses) and task completion time (i.e., average response time per question). The first set of eight questions was a training set, thereby using only the second set of eight questions for the analysis. We also collected qualitative feedback in free-form text.

4.3 Participants

We recruited 129 participants and randomly assigned them to either the PF condition (N = 67) or NF condition (N = 62). We limited the study participants to mouse users because different input methods could influence task completion time significantly. Additionally, we filtered MTurkers who did not pay careful attention to the study instructions as indicated by the description reading time; we removed 47 participants who spent less than 20 seconds on the description page (it took the authors more than 20 seconds simply to read the description).

4.4 Results of Study 1

The independent samples t-test showed that in terms of accuracy, there was not enough evidence to suggest that PF icon group (N = 43, M = 7.81, SD = .59) differs from NF icon group (N = 39, M = 7.59, SD = 1.29), $t(80) = 1.03$, $p = .31$. To compare the task completion time (in second), we first removed outliers defined by the mean minus two standard deviations ($M - 2SD$) (i.e., the number of correct responses < 5.7). Then we excluded incorrect responses in calculating the task completion time. The independent samples t-test showed that in terms of task completion time, there was not enough evidence to suggest that PF icon group (N = 39, M = 3.29, SD = .92) differs from NF icon group (N = 36, M = 3.50, SD = 1.12), $t(73) = -.89$, $p = .37$.

4.5 Discussion of Study 1

After reading the description and solving eight icon comparison questions, the majority of participants in both groups were able to comprehend privacy ratings in a similar manner. The PF and NF icons resulted in the comparable level of comprehension and speed by survey participants. Thus, we chose to use this set of PF and NF icons (Figure 1-a) as stimuli for our subsequent framing study.

We note, however, that even if there was no significant difference in the performance, 7 out of 36 (= 19%) participants in the NF icon group mentioned in their written comments that the NF icons were confusing. We suspect that this is due to a strong preconception of "*the more, the better,*" which is what the prevalent PF rating scale conveys. At the same time, other participants commented that the red minus sign we used to convey the negativity was particularly helpful to mark privacy-invasive apps. In Study 2, we explored whether the complementary visuals have similar influence on how people perceive an app.

5 Study 2: Positive vs. Negative Framings

Given that people can comprehend the positive and negative visual framings in a similar manner with some training, we investigated whether there is any effect of *framing* on how people perceive apps. In prior framing studies using text descriptions, researchers report that negatively-framed information tends to influence *more* than the positively-framed information of the same magnitude [24,26]. In Study 2, we tested which visual framing—positive or negative—was more effective in nudging people away from privacy-invasive apps using the icons we designed in Study 1. We also explored how people's perception of an app changes if a privacy rating of the app accompanies a user's overall rating (user rating). This resulted in a 2 (framing: PF icon; NF icon) x 3 (privacy rating: high; medium; low) x 2 (user rating: with a user rating of 3; without a user rating) mixed design with repeated measures; *framing* and *user rating* were between-subjects factors and *privacy rating* was a within-subjects factor, thereby forming four conditions: PF with & without a user rating of 3, and NF with & without a user rating of 3.

5.1 Survey Content

We created online surveys for the four conditions. The surveys consisted of three sections: (1) evaluating four apps (one dummy app followed by three apps of varying degrees of privacy ratings), (2) eight Privacy Critics' Rating icon comparison questions, and (3) demographic questions. At the beginning of the surveys, we introduced a hypothetical privacy metric called "Privacy Critics' Rating" using the same blurb from Study 1. Then, we showed a fake weather app with the following description and a screenshot.

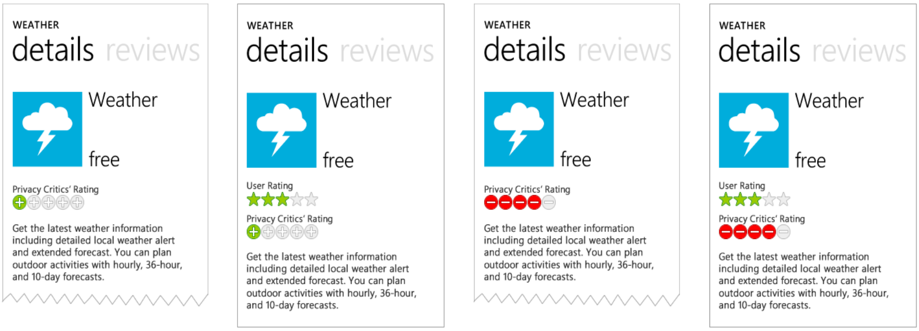
> Imagine that you need a weather app, so you are searching the marketplace on your phone. While looking around many weather apps, you come across the following app page:

The screenshot contained a visual privacy rating, icon, and app description (Figure 3). We manipulated the screenshot of the weather app in the following manner:

Privacy Rating: For each participant, we showed a series of four screenshots of a weather app: one dummy set (user rating: 3.5; privacy rating: 2) and three levels of privacy ratings (high; medium; low) in a random order (Figure 4). The dummy set was included at the beginning to account for a longer reading time and familiarity with the screenshot and the question types.

Positive Framing (PF) vs. Negative Framing (NF): We created a PF and NF condition by manipulating the description of the Privacy Critics' Rating and its icon design as in Study 1 (Figure 4).

With a User Rating of 3 (w/ UR) vs. Without a User Rating (w/o UR): We chose a user rating of three as a representative case for the w/ UR condition because most of the popular free apps in the marketplace have an equal or higher user rating than three. Three green-colored stars (out of five) appeared right above the privacy rating in the w/ UR conditions (Figures 3-b/d). No user rating was provided in the w/o UR conditions (Figures 3-a/c).



(a) "Positive framing without a user rating" condition (b) "Positive framing with a user rating of 3" condition (c) "Negative framing without a user rating" condition (d) "Negative framing with a user rating of 3" condition

Fig. 3. Screenshots of the detailed view. Each screenshot represents the following: (a)—PF & w/o UR, (b)—PF & w/ UR, (c)—NF & w/o UR, and (d)—NF & w/ UR. Below the privacy rating, we provided descriptions of an app.

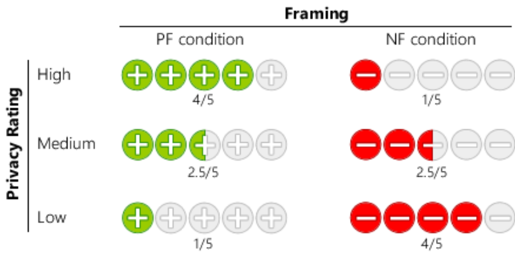


Fig. 4. This illustrates how we manipulated the privacy rating (3 levels) and framing (2 levels)

5.2 Measures

After showing each app, we measured people's perception toward each app by asking the following four questions: (1) trustworthiness of the app (TRUST), (2) likeability of the app (LIKE), (3) willingness to install the app (INST), and (4) willingness to recommend the app to a friend (RCMD). We measured TRUST because we suspected that people would associate an app's privacy rating with its trustworthiness. LIKE is a conventional measure in prior framing studies where a product's attribute is framed differently [19]. We measured INST and RCMD to gauge people's willingness to adopt an app. TRUST and LIKE were measured on a 7-point Likert scale, where 1 = not at all trustworthy / I strongly dislike this app, and 7 = very trustworthy / I strongly like this app. INST and RCMD were measured using Yes/No dichotomous questions. In addition, we suspected that participants' interest level toward the weather app could be a factor related to other dependent measures. So we measured self-reported interest level (INTEREST) toward the weather app on a 7-point Likert scale, where 1 = not at all interested, and 7 = very interested, at the very beginning of the survey as we showed the screenshot *without* the privacy and user rating information.

Within each condition, the privacy rating was the only factor that we varied. It was necessary for participants to understand how the privacy rating worked so that they could answer the evaluation questions based on correct understanding of the stimuli. Therefore, we measured accuracy and task completion time for the eight privacy rating icon comparison questions as in Study 1. However, this time, we placed these questions *after* the app evaluation questions to use them for *filtering* purpose rather than *training* purpose. We made this decision based on our pilot study and previous research indicating that framing effects are susceptible to rational thinking (e.g., asking people to provide a rationale for their choice eliminates or reduces the framing effects [22]).

5.3 Participants

We recruited 332 participants from Amazon Mechanical Turk and randomly assigned them to one of the four conditions: PF & w/o UR ($N = 75$); PF & w/ UR ($N = 95$); NF & w/o UR ($N = 79$); and NF & w/ UR ($N = 83$). We removed 18 duplicate participants (i.e., who participated in Study 1 or pilot studies). Using the same criteria from Study 1, we removed 79 participants who spent little time reading the descriptions. Among the remaining 235 participants, 55% of the participants ($N = 129$) were male, and 89% of the participants ($N = 210$) claimed that they own a smartphone.

5.4 Results of Study 2

After removing 21 participants from the analysis whose number of correct responses to the eight filtering questions was less than $M - 2SD$ (i.e., the number of correct responses < 5.3), we observed that participants' initial interest level toward the weather app was highly related to how much they trust the weather app, $F(1, 209) = 9.02$, $p = .003$, and how much they like the weather app, $F(1, 209) = 23.20$, $p < .001$. Therefore, TRUST and LIKE were analyzed using a mixed-design analysis of covariance (ANCOVA) controlling for the INTEREST as the covariate (Table 1). INST and RCMD were analyzed using a Pearson chi-square test (Table 2).

Table 1. This table shows strong main effects of privacy rating on TRUST and LIKE, and the framing effect (significant interaction effect between privacy rating and framing) on TRUST

	TRUST (F-value)	LIKE (F-value)
Privacy Rating	55.55**	30.42**
Framing	2.33	.91
User Rating	1.83	3.41‡
Privacy Rating X Framing	7.48*	.83
Privacy Rating X User Rating	3.19‡	12.02**

**p<.001, *p<.05, ‡p<.10

"TRUST" Question. After controlling for the effect of INTEREST, we found a significant main effect of privacy rating on TRUST, $F(1.38, 289.13) = 55.55$, $p < .001$. Planned contrasts revealed that a high privacy rating app ($M = 5.15$) was regarded as more trustworthy than a medium privacy rating app ($M = 3.03$), $F(1, 209) = 50.39$, $p < .001$. Also, a medium privacy rating app was regarded as more trustworthy than a low privacy rating app ($M = 1.75$), $F(1, 209) = 29.60$, $p < .001$.

We found a significant interaction effect between privacy rating and framing, $F(1.38, 289.13) = 7.48$, $p = .003$ (Figure 5). To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across PF and NF conditions. We found a significant interaction between a high privacy rating app and a low privacy rating app across PF and NF conditions, $F(1, 209) = 9.09$, $p = .003$, and between a medium privacy rating app and a low privacy rating app across PF and NF conditions, $F(1, 209) = 8.89$, $p = .003$. Also, there was a marginally significant interaction between a high privacy rating app and a medium privacy rating app across PF and NF conditions, $F(1, 209) = 3.50$, $p = .06$. As Figure 5 indicates, the difference between PF ($M = 1.44$) and NF ($M = 2.05$) was significant for a low privacy rating app, $t(211) = 2.42$, $p = .02$. No significant differences were found between PF ($M = 5.27$) and NF ($M = 5.02$) conditions in a high privacy rating app, $t(211) = .34$, $p = .12$, or between PF ($M = 2.97$) and NF ($M = 3.10$) conditions in a medium privacy rating app, $t(211) = 1.58$, $p = .74$.

We found a marginally significant interaction between privacy rating and presence of user rating, $F(1.38, 289.13) = 3.19$, $p = .06$. This indicates that user rating might have different effects on app's trustworthiness at different levels of privacy rating. To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across w/ and w/o UR. We found a significant interaction when comparing a high privacy rating app to a medium privacy rating app across w/ and w/o UR, $F(1, 209) = 4.43$, $p = .04$. Also, we found a marginally significant interaction when comparing a high privacy rating app to a low privacy rating app across w/ and w/o UR, $F(1, 209) = 3.44$, $p = .07$. As Figure 6 indicates, when a user rating of 3 is shown, there is a decline in TRUST for a high privacy rating app while there is an increase in TRUST for a low and medium privacy rating app.

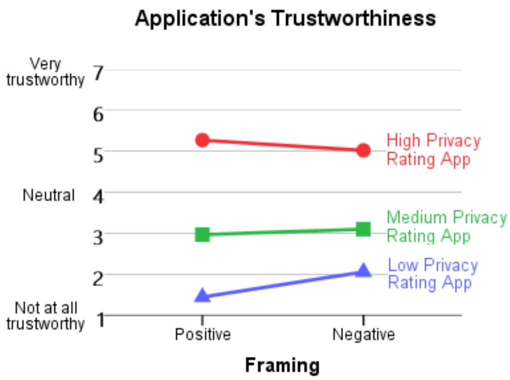


Fig. 5. This shows a significant interaction between privacy rating and framing on TRUST. Participants interpreted the NF icons more positively than the PF icons of the equivalent rating.

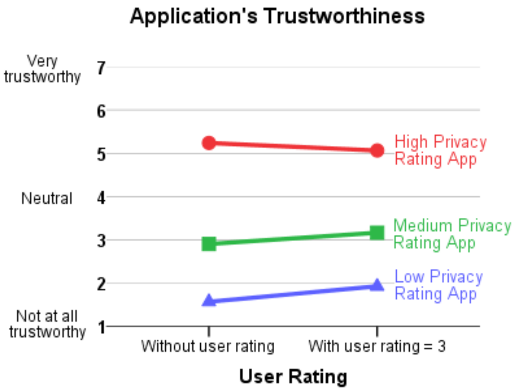


Fig. 6. This shows a marginally significant interaction (p = .06) between privacy rating and user rating on TRUST

"LIKE" Question. After controlling for the effect of INTEREST, we found a significant main effect for privacy rating on LIKE, $F(1.46, 305.42) = 30.42$, $p < .001$. This effect tells us that how much participants liked the weather app was different for high, medium, and low privacy rating apps. Planned contrasts revealed that participants liked the high privacy rating app ($M = 5.01$) significantly more than the medium privacy rating app ($M = 3.45$), $F(1, 209) = 12.41$, $p = .001$. Also, participants liked the medium privacy rating app significantly more than the low privacy rating app ($M = 2.17$), $F(1, 209) = 48.01$, $p < .001$.

We found a marginally significant main effect for user rating, $F(1, 209) = 3.41$, $p = .07$. Participants liked the app w/ UR of 3 ($M = 3.66$) more than the app w/o UR ($M = 3.42$).

We found a significant interaction between privacy rating and user rating, $F(1.46, 305.42) = 12.02$, $p < .001$ (Figure 7). This indicates that user rating had different effects on LIKE depending on different levels of privacy rating. To break down this interaction, planned contrasts were performed comparing each level of privacy rating to one another across w/ UR and w/o UR. We found a significant interaction when comparing a high privacy rating app to a medium privacy rating app across w/ UR and w/o UR, $F(1, 209) = 15.45$, $p < .001$. Also, we found a significant interaction when comparing a high privacy rating app to a low privacy rating app across w/ UR and w/o UR, $F(1, 209) = 14.02$, $p < .001$. The interaction graph (Figure 7) tells us that when a user rating of 3 is shown, there is a slight decline in LIKE for a high privacy rating app while there is an increase in LIKE for a low and medium privacy rating app. The remaining contrasts revealed no significant interaction when comparing a medium privacy rating app to a low privacy rating app in the w/o UR and w/ UR conditions, $F(1, 209) = 2.59$, $p = .11$.

"INSTALL" Question. We found a marginally significant association between framing and participants' choice of installing a low privacy rating app, χ^2 (1, N = 214) = 3.42, p = .06 (Table 2). The odds ratio implies that the odds of participants

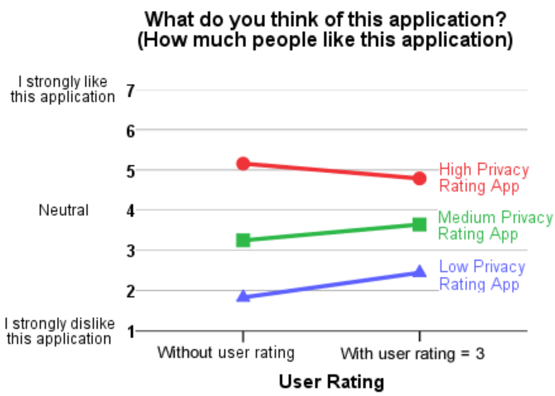


Fig. 7. This shows a significant interaction between privacy rating and user rating on LIKE. A user rating of 3 had a different effect on the high privacy rating app in comparison to medium/low privacy rating apps.

installing a low privacy rating app were 3.36 times higher if the rating were negatively framed than positively framed.

We found a significant association between framing and participants' choice of installing a medium privacy rating app, χ^2 (1, N = 214) = 5.17, p = .02 (Table 2). The odds ratio implies that the odds of participants installing a medium privacy rating app were 2.21 times higher if the rating were negatively framed than positively framed.

We found a marginally significant association between a user rating and participants' choice to install a high privacy rating app, χ^2 (1, N = 214) = 3.20, p = .07 (Table 2). The odds ratio implies that the odds of participants installing a high privacy rating app were 1.80 times higher if there were no user rating than the app accompanying a user rating of 3.

Table 2. More people answered that they would install / recommend a negatively-framed low privacy rating app than the same app that is positively-framed. The number of people who answered that they would install / recommend a high privacy rating app decreased when a user rating of 3 was shown than when no user rating was shown.

Privacy Rating	Answer	INSTALL Question			RECOMMEND Question		
		Positive Framing	Negative Framing	p-value[a]	Positive Framing	Negative Framing	p-value[a]
Low Privacy Rating App	Yes	3 (2.8%)	9 (8.6%)	.06‡	2 (1.8%)	10 (9.5%)	.02*
	No	106 (97.2%)	96 (91.4%)		107 (98.2%)	95 (90.5%)	
Medium Privacy Rating App	Yes	13 (11.9%)	25 (23.8%)	.02*	10 (9.2%)	17 (16.2%)	.12
	No	96 (88.1%)	80 (76.2%)		99 (90.8%)	88 (83.8%)	
		No User Rating	User Rating = 3	p-value[a]	No User Rating	User Rating = 3	p-value[a]
High Privacy Rating App	Yes	88 (82.2%)	77 (72.0%)	.07‡	83 (77.6%)	69 (64.5%)	.04*
	No	19 (17.8%)	13 (28.0%)		24 (22.4%)	38 (35.5%)	

a. A p-value is calculated from a Pearson's Chi-square test. *p<.05, ‡p<.10.

"RECOMMEND" Question. We found a significant association between framing and participants' choice of recommending an app with low privacy rating, χ^2 (1, N = 214) = 5.97, p = .02 (Table 2). The odds ratio implies that the odds of participants recommending a medium privacy rating app to a friend were 5.53 times higher if the ratings were negatively framed than positively framed.

We also found a significant association between user rating and participants' choice of recommending an app with high privacy rating, χ^2 (1, N = 214) = 4.45, p = .04. The odds ratio implies that the odds of participants recommending a high privacy rating app were 1.90 times higher if there were no user rating than a user rating of 3.

5.5 Discussion of Study 2

Study 2 results show a strong effect of privacy rating on all dependent measures. This indicates that when a privacy rating of a given app is disclosed *visually*, people are influenced by the privacy rating. The influence of the privacy rating appears to decline (although still significant) when we showed a user rating of 3 (Figures 6/7). This suggests that people are susceptible to both privacy rating and user rating.

The effect of framing was subtle. First, we observed framing effects on TRUST in a low privacy rating app. Participants expressed a lower level of trustworthiness of an app when its privacy rating was positively framed than negatively framed. For medium and high privacy rating apps, framing effect did not occur. A similar trend was observed for INST and RCMD—a low privacy rating app was a common denominator for the framing effects to be observed, and when observed, it was always the negatively framed icons that people interpreted more positively. However, there was no framing effect on LIKE; after controlling for people's app interest level, privacy rating and user rating dominantly influenced LIKE. As we suspected, it appears that participants associated the privacy ratings with TRUST more than LIKE.

Prior framing studies using text descriptions consistently show that positive framing leads to more favorable evaluations than negative framing [e.g., 18,21]. Researchers demonstrate that describing an option in a negative light (e.g., "mortality rate") focuses attention on the unfavorable possibilities associated with this option, rendering it less acceptable to the decision-maker [19]. Therefore, we initially suspected that emphasizing negativity (e.g., privacy-invasiveness) would nudge people away from privacy invasive apps with a low privacy rating. However, our study results suggest this is not the case.

On the contrary, PF icons were more effective in making a low privacy rating app look more unfavorable. We suspect that people have strong connotations of "the more, the better" in the rating context. Because a negatively framed privacy invasive app has *more* signs in the rating than the equivalent PF icons (i.e., four minus signs vs. one plus sign), it is plausible that the higher number of ratings, regardless of its meaning, could have contributed to how people perceive the PF/NF icons.

Our results also suggest that there was no framing effect in the high and medium privacy rating apps. Therefore, the use of PF icons for depicting privacy ratings is a better choice for nudging people away from privacy invasive apps while not affecting high and medium privacy rating apps.

6 Discussion and Future Work

Our results suggest that visual representations of privacy information of apps can influence installation decisions by smartphone users. When disclosed visually on the detailed view of an app, the majority of participants commented that they found the privacy rating very helpful in deciding whether to install an app. In this section, we discuss design recommendations for a visual privacy rating as we answer our research questions. We also discuss limitations of this work and areas for future research.

6.1 Design Recommendations

To design complementary visuals that represent the same privacy rating in a positive or negative way (RQ1), we suggest leveraging visual attributes (e.g., colors) and semantics (e.g., valence, signs, symbols) that are prevalent throughout the culture for conveying valence. In text, people make favorable and unfavorable associations with positively and negatively phrased attribute labels (e.g., "fat" is associated with "bad" and "lean" with "good"). Similarly in visuals, we can take advantage of people's preconceptions built on their life experiences.

In designing PF and NF privacy ratings, we tried out various symbolic figures in red and green and chose the most promising one based on pilot tests measuring task completion time and error rate. For example, plus and minus signs, or thumbs-up and thumbs-down are well-known symbols in the US (Figures 1-a/b). When colors (red for negative and green for positive) are added on top of these, the valence becomes more pronounced. However, we ruled out thumbs-up and thumbs-down icons because they are already being used to convey different meanings elsewhere (e.g., "thumbs-up" is associated with "like" in Facebook). We also ruled out neutral figures (e.g., rectangles in Figure 1-c)—providing extra information (e.g., a legend) shall be necessary to frame neutral figures.

Study 2 results show that the complementary visuals do not have the same influence on how people perceive an app in certain conditions (RQ2). If the goal is to nudge people away from privacy-invasive apps, we suggest using a *positive visual framing* for privacy rating. However, this might not be the case outside the realm of "rating scale" design; when "the more, the better" connotation no longer exists, positive visual framing might lead to more favorable evaluations and vice versa, similar to prior framing effects shown in texts. Therefore, future research is warranted to study the influence of positive and negative visual framings in other contexts.

The results of Study 2 also show that people's perception of an app changes if a privacy rating is accompanied by a user rating but that a privacy rating still has a strong main effect (RQ3). This is promising in that people do understand and consider the level of privacy protection when it is presented to them visually. Prior research shows that textual privacy-related information is easily discarded [12,16]. This again suggests the need for *visual privacy ratings*. This work, however, does not address whether the main effect of visual privacy rating is due to the privacy information shown in a rating scale, due to the visuals, or due to the combination of both. Therefore, future research is needed to tease out these factors.

6.2 Limitation and Future Work

In running online experiments with participants recruited from the MTurk site, we made a conservative assumption of filtering mindless Turkers defined by those who spent less than 20 seconds on the description page where we explained the privacy rating and what it represents. Because we introduced a privacy rating with which most people are unfamiliar, eliminating answers from an arbitrary guess was necessary. This is analogous to avoiding recruits who do not understand what “fat” and “lean” mean in the ground beef framing study [18]. By the same token, we eliminated outliers whose number of correct responses to the icon comparison questions was less than $M - 2SD$. However, this does not guarantee that those who spent longer than 20 seconds on the description page correctly interpreted the privacy rating, which is beyond our control when running an experimental study online.

In this work, it was not our main interest to investigate the relative influence of privacy-related information in comparison to user ratings (although we touched upon the interaction between a privacy rating and user rating in Study 2). However, our study results suggest that understanding this topic would inform how much emphasis should be placed visually on the privacy ratings in comparison to user ratings. To answer this question, we would need to understand how people’s perceptions of an app change depending on different designs and the relative weight of a visual privacy rating in return.

We would also like to further investigate how visual framings can be applied outside the privacy domain as a method for nudging. Health information feedback of consumer health devices, for example, is designed with a specific goal in mind—promoting health-enhancing behaviors. One promising direction is to design visual framings that represent the outcome of a healthy or unhealthy behavior stressing either the positive or negative consequences as we provide health information feedback. Framing research in texts has addressed when framing effects are *reduced* or *eliminated* [10,22], which helps HCI researchers select application areas for employing the visual framing.

7 Conclusion

We created visuals for a mobile app’s privacy rating by leveraging the well-known “Framing Effects.” In Study 1, we showed that it is viable to create semantically equivalent privacy ratings framed in either a positive or negative light. In Study 2, we showed a strong main effect for visual privacy rating on people’s perception of an app. Our study results suggest that when an app’s privacy property is provided in the form of a visual privacy rating, people are able to understand and are heavily influenced by it. We also examined how the visual framings shift people’s perception of an app, and observed the occurrence of framing effects in a low privacy rating app. In designing a visual privacy rating to nudge people away from privacy-invasive apps, we recommend using positive visual framing, and leveraging visual attributes and semantics that are prevalent throughout the culture for conveying valence. Investigating the relative influence of a visual privacy rating in comparison to a user rating warrants future research efforts. In closing, this work provides empirical guidance for creating influential, persuasive visual framing that can nudge people away from privacy invasive apps.

Acknowledgment. We would like to thank Tim Paek (Microsoft Research) for his thoughtful comments on earlier drafts of this paper. We also thank Stuart Schechter (Microsoft Research) for providing suggestions on the study design.

References

1. 148Apps.biz, App Store Metrics (2012), `http://148apps.biz/app-store-metrics` (retrieved December 31, 2012)
2. Acquisti, A., Grossklags, J.: What can behavioral economics teach us about privacy? In: Acquisti, A., Gritzalis, S., Di Vimercati, S., Lambrinoudakis, C. (eds.) Digital Privacy: Theory, Technologies, and Practices, pp. 363–379. Auerbach Publications (2007)
3. Acquisti, A.: Nudging privacy: behavioral economics of personal information. Security & Privacy 7(6), 82–85 (2009)
4. AndroLib (2012), `http://www.androlib.com/appstats.aspx` (retrieved December 31, 2012)
5. Balebako, R., Leon, P.G., Almuhimedi, H., Kelley, P.G., Mugan, J., Acquisti, A., Cranor, L.F., Sadeh, N.: Nudging users towards privacy on mobile devices. In: Proc. CHI 2011 Workshop on Persuasion, Nudge, Influence and Coercion (2011)
6. Boyles, J.L., Smith, A., Madden, M.: Privacy and data management on mobile devices. Pew Internet & American Life Project (2012), `http://pewinternet.org/Reports/2012/Mobile-Privacy/Key-Findings.aspx?view=all` (retrieved December 31, 2012)
7. Camerer, C., Issacharoff, S., Samuel, C., Loewenstein, G.: Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism". University of Pennsylvania Law Review 151, 1211–1254 (2003)
8. Chin, E., Felt, A., Sekar, V., Wagner, D.: Measuring user confidence in smartphone security and privacy. In: Proc. SOUPS 2012 (2012)
9. Cranor, L.F., Guduru, P., Arjula, M.: User interfaces for privacy agents. ACM TOCHI 13(2), 135–178 (2006)
10. Druckman, J.: Using credible advice to overcome framing effects. J. Law, Economics, and Organization 17, 62–82 (2001)
11. Egelman, S., Felt, A.P., Wagner, D.: Choice architecture and smartphone privacy: there's a price for that. In: Proc. WEIS 2012 (2012)
12. Felt, A., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android permissions: user attention, comprehension, and behavior. In: Proc. SOUPS 2012 (2012)
13. Johnson, E.J., Bellman, S., Lohse, G.L.: Defaults, framing and privacy: why opting in-opting out. Marketing Letters 13(1), 5–15 (2002)
14. Kalnikaitė, V., Rogers, Y., Bird, J., Villar, N., Bachour, K., Payne, S., Todd, P.M., Schöning, J., Krüger, A.: How to nudge in situ: designing lambent devices to deliver information salience in supermarkets. In: Proc. UbiComp 2011, pp. 11–20 (2011)
15. Kelley, P.G., Cesca, L., Bresee, J., Cranor, L.F.: Standardizing privacy notices: an online study of the nutrition label approach. In: Proc. CHI 2010, pp. 1573–1582 (2010)

16. Kelley, P.G., Consolvo, S., Cranor, L., Jung, J., Sadeh, N., Wetherall, D.: A conundrum of permissions: installing apps on an android smartphone. In: Proc. FCDS 2012, pp. 68–97 (2012)
17. Lee, M.K., Kiesler, S., Forlizzi, J.: Mining behavioral economics to design persuasive technology for healthy choices. In: Proc. CHI 2011, pp. 325–334 (2011)
18. Levin, I.P., Gaeth, G.J.: Framing of attribute information before and after consuming the product. J. Consumer Research 15, 374–378 (1988)
19. Levin, I.P., Schneider, S.L., Gaeth, G.J.: All frames are not created equal: a typology and critical analysis of framing effects. Organizational Behavior and Human Decision Processes 76(2), 149–188 (1998)
20. Lin, J., Amini, S., Hong, J., Sadeh, N., Lindqvist, J., Zhang, J.: Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In: Proc. UbiComp 2012, pp. 501–510 (2012)
21. Marteau, T.M.: Framing of information: its influence upon decisions of doctors and patients. British Journal of Social Psychology 28, 89–94 (1989)
22. Miller, P.M., Fagley, N.S.: The effects of framing, problem variations, and providing rationale on choice. Personality and Social Psychology Bulletin 17, 517–522 (1991)
23. MSN Movies, X-Men (2012), `http://movies.msn.com/movies/movie/x-men` (retrieved December 31, 2012)
24. Peeters, G., Czapinski, J.: Positive-negative asymmetry in evaluations: the distinction between affective and informational negativity effects. European Review of Social Psychology 1, 33–60 (1990)
25. Rogers, Y., Hazlewood, W.R., Marshall, P., Dalton, N., Hertrich, S.: Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? In: Proc. UbiComp 2010, pp. 261–270 (2010)
26. Taylor, S.E.: Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. Psychological Bulletin 110, 67–85 (1991)
27. Thaler, R.H., Sunstein, C.R.: Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven (2008)
28. The Wall Street Journal, What they know (2012), `http://blogs.wsj.com/wtk-mobile` (retrieved December 31, 2012)
29. Tsai, S., Egelman, L., Cranor, L.F., Acquisti, A.: The effect of online privacy information on purchasing behavior: an experimental study. In: Proc. WEIS 2007 (2007)
30. Tversky, A., Kahneman, D.: The framing of decisions and the psychology of choice. Science 211(4481), 453–458 (1981)
31. Tversky, A., Kahneman, D.: Rational choice and the framing of decisions. J. Business 59, 251–278 (1986)
32. Wired, Apple says grabbing address book data is an iOS policy violation (2012), `http://www.wired.com/gadgetlab/2012/02/apple-responds-to-path` (retrieved December 31, 2012)

The Impact of Encumbrance on Mobile Interactions

Alexander Ng, Stephen A. Brewster, and John Williamson

School of Computing Science, University of Glasgow, Glasgow, United Kingdom
a.ng.1@research.gla.ac.uk, stephen.brewster@glasgow.ac.uk,
jhw@dcs.gla.ac.uk

Abstract. This paper investigates the effects of encumbrance (holding different types of objects while using mobile devices) to understand the interaction difficulties that it causes. An experiment was conducted where participants performed a target acquisition task on a touchscreen mobile phone while carrying different types of bags and boxes. Mobility was also evaluated since people carry items from one place to another. Motion capture hardware was used to track hand and arm postures to examine how holding the different types of objects caused excessive movement and instability therefore resulting in performance to decline. The results showed encumbrance and mobility caused target accuracy to decrease although input while holding the box under the non-dominant arm was more accurate and exerted quicker targeting times than holding no objects. Encumbrance affected the dominant hand more than the non-dominant hand as targeting error significantly increased and caused greater hand instability. The issues caused by encumbrance suggest the topic requires more attention from researchers and users would benefit greatly if better interaction techniques and applications are developed to counteract the problems.

Keywords: Encumbrance, Mobility, Mobile interactions, Target acquisition.

1 Introduction

Mobile devices such as smartphones play a vital role in our everyday activities as they allow users to perform common tasks such as talking to friends, emailing documents to colleagues and searching for nearby services while on the move.

As a result, mobile devices are being used in a wide range of different contexts and it is important to examine if users experience interaction and usability difficulties when faced with potentially demanding multitasking situations. One context that has not been explored in great depth is studying users when they are encumbered: holding typical objects as such handbags, umbrellas, shopping bags and boxes while engaging with their devices simultaneously. People frequently carry these objects while walking from one place to another and as a result, users often need to use their phones at the same time to send messages, look at maps or refer to other services. Therefore, encumbrance and mobility is closely linked with each other. Using a mobile phone in these kinds of situations while physically hampered is challenging as it can be awkward to see the screen and make input. It would be valuable to understand how

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 92–109, 2013.

encumbrance affects interactions with touchscreen mobile devices so that better input techniques and error detection could be developed to negate the issues and problems it causes. Also, carrying different types of objects may have a different impact on the user thus it is beneficial to be able to categorize the different types of encumbrances by the way they affect the user's performance when interacting with mobile devices.

Since users are likely to carry objects between places, it is also imperative to examine how walking and encumbrance together affect input performance on touchscreen mobile devices. The physical motion required to walk naturally causes the user's arm to move and swing in phase with each foot step. However, this natural stance is disrupted when attention from the mobile device is required as the user attempts at steadying their constantly moving arm and hand to input. The effects of encumbrance is likely to cause interaction to become even more physically demanding and put significant pressure on the user's mental ability to multitask as they walk and navigate the surrounding area while avoiding nearby obstacles. The issues caused by walking alone on mobile interactions have previously been investigated in great detail as researchers have developed various solutions to assist the user to input more accurately [3] and [4]. However, it is unclear what further issues the introduction of encumbrance causes on interaction when the user is walking. Therefore, it is important to understand if performance is further worsened by encumbrance and to detect new usability problems that previously have been overlooked. It would also be interesting to see if encumbrance and mobility causes distinctive interaction issues therefore better applications could be developed to assist the user to maintain a good control of their devices in realistic multitasking situations.

To examine the effects of encumbrance and mobility on mobile interactions, we defined a set of common objects that users frequently carry to be assessed in our experiment. There is a large number of possible objects that people hold in their daily activities and the issue is further complicated the object's characteristics such as its size, shape, heaviness and quantity. Also, a particular object could be held in various different ways which normally depend on personal preference and the context the person is currently in. As a result, it is important to focus on the typical strategies that users perform while encumbered in order to simultaneously use their mobile device. An observational study was carried out to find the most common types of objects and the way they were being carried in a range of public settings. Based on the results collected from the study, the encumbrance experiment examined two types of objects: holding a bag in hand and carrying a box underarm. A target acquisition task was used on a touchscreen mobile phone to evaluate the effects of the two types of encumbrances and mobility. The participants performed target selections on a touchscreen mobile phone and holding the different types of objects at the same time while either standing still or walking to simulate realistic encumbrance scenarios. To investigate how holding the different types of bags and boxes cause targeting difficulties, motion capture cameras were used to track hand and arm movements during interaction. The assumption was that carrying different objects while walking will cause more physical movements and instability to the user's hand and arm which makes it more challenging for the user to maintain a steady position to input accurately on the touchscreen device.

2 Background

This section of the paper will review the related literature with the first part discussing research that has examined encumbrance and the second part reviewing studies on mobile interactions while walking.

2.1 The Effects of Encumbrance

There has been little research that specifically examined the interaction problems caused by encumbrance due to holding common objects.

Ng *et al.* [10] studied the topic of encumbrance by examining wrist rotation gestures as a novel hands-free and eyes-free interaction technique to reduce the issues caused by holding a bag and box when interacting with a mobile phone. A small bag and a rectangular box were evaluated to replicate some of the effects of carrying cumbersome objects. The main motivation behind the study was to illustrate that holding common objects does have an impact on mobile interactions and different types of objects have a different impact on the user's performance. The findings from their Fitts' Law targeting experiment suggested holding the bag caused users to become less accurate while the performance of carrying the box underarm was similar to unencumbered. The swinging motion caused when holding the bag made it difficult to steady the arm for input while the users were able to stabilize their forearm when gripping the box in place which helped them to perform wrist gestures more easily. Target movement times between the two types of encumbrances were not significant. The study also showed that sensor components such as gyroscopes that are found in most modern mobile devices could be used to detect precise wrist orientation and movements and therefore support the user when they are physically hindered.

Oulasvirta and Bergstrom-Lehtovirta [13] studied the relationship between holding a group of smaller-sized objects and input accuracy on different computing devices. Twelve different multitasking situations were examined which covered a range of hand grips and arm postures while holding a variety of everyday items such as a writing pen and a beverage cup. Participant performed the various types of encumbrances while selecting targets on a laptop computer via mouse and trackpad input and text entry by different forms of keyboards on a mobile phone. The results showed that holding objects such as a pen while performing the target selection task on the laptop by using the trackpad caused performance to decline more than mouse input. One-handed text entry on the mobile device while holding an object (such as a pair of scissors) caused a decrease in performance when compared to two-handed input especially with single-handed stylus targeting since it normally requires both hands to input. Interestingly, text entry via a virtual keyboard exerted better results than the physical equivalent as more finger pressing pressure is required. The study examined a set of interesting manual tasks and included activities that require a pushing action rather than the common holding in hand grip. One final point that the study makes is the notion of safety as there are situations where it may not be possible to prevent hindrance when using mobile devices (for example, holding small children). Better context detection systems and more effective input techniques are required to help users during interaction when confronted with these challenging situations. Wolf [17] assessed how people performed manual tasks by examining hand grips and positions in order to explore the areas that are free to perform secondary interaction activities.

Mainwaring *et al.* [8] conducted an ethnographic study across three major cities to examine the personal connection between the items that people carried and how these items were used in their daily context with surrounding people and interfaces. Items were classified into various categories such as those which distracted the user from the environment (music players, phones and books), personal belongings (wallets, keys, make-up) and professional tools (laptop, PDAs). The findings from the study suggested different items have its own unique personal value and as a result different objects may have a varied impact on interaction with mobile devices. The issue of a particular object's personal value to the user creates an unusual viewpoint on encumbrance. For example, dropping a wallet could be more disastrous and frustrating to the user than a mobile phone while on the move. It would be worthwhile to investigate if this is the case and if objects can also be grouped by personal value rather than the standard categories of the object's size, shape or how it obstructs the user.

Tamminen *et al.* [16] observed how outdoor environments constantly compete for the user's attention and discovered repeated instances where the user's hands were busy performing activities ranging from holding a newspaper while travelling between locations to clutching a cigarette packet while searching for money placed in the person's pockets. Performing several activities at the same time is likely to cause the user difficulty in dividing its visual attention to complete each individual task successfully as discussed by [12]. There needs to be more efficient interaction methods and alternative techniques to assist the user when they are physically impaired and visually distracted. People carry personal belongings and objects from one place to another therefore it is also important to examine the impact mobility has on mobile interactions and its relationship with encumbrance.

2.2 The Effects of Mobility

The interaction difficulties caused by walking while using mobile devices have been well documented.

Bergstrom-Lehtovirta *et al.* [2] examined the relationship between target selections on a touchscreen mobile phone and the user's preferred walking speed (PWS) on a treadmill. The results showed that when users walked approximately between 40 - 80% of their PWS, performance began to level as users were able to stabilize the dominant hand more successfully to input more accurately on the touchscreen mobile phone. Mizobuchi *et al.* [9] recorded an average walking speed of 1.77 km/h when examining the relationship between mobility and button size. This finding is much lower than the walking speed of an average adult human being [5]. It is important to see if holding different types of objects cause the user to reduce their walking speed even further and if the slowdown in pace meant that users were able to keep a reasonable level of input accuracy and performance.

Kane *et al.* [4] attempted to compensate some of the problems caused by mobility by developing Walking User Interfaces (WUI) which dynamically change the characteristics of interface widgets to support the user to input more effectively while walking. The results showed a trade-off between button size and the amount of effort and time required to scroll the screen to make the appropriate selection. This was also discussed by Schildbach and Rukzio [15]. The performance of their WUI prototype was comparable to an equivalent static interface although it was not as effective as a fixed layout with larger sized buttons. Goel *et al.*[3] and Nicolau and Jorge[11] have also studied the issues of walking and developed better text entry systems to help typing on touchscreen mobile devices. It would be useful to examine if these applications and similar interfaces are still as effective and could solve some of the interaction problems caused by encumbrance.

Brewster [1] showed that button pressing on a PDA was more accurate and subjectively easier when the user was sitting down compared to walking outdoors. A drop in performance of approximately 30% was recorded and one possible cause could have been due to the increased mental attention required to walk and navigate the environment while engaging with the mobile device at the same time. The effects of encumbrance may result in error rate to increase further between standing and walking. The other aspect that is worth considering is evaluating walking-based experiments in laboratory settings and outdoor environments. Our experiment was conducted inside a quiet room due to the restrictions of the motion capture hardware. Consequently, this may have had an undesired effect on targeting performance when carrying the different types of objects since the indoor environment is much calmer than a real world setting therefore making the context less challenging for the user. However, Lin *et al.* [7] suggested using an artificial route with obstacles to increase the user's cognitive workload to a level similar to walking in outdoor settings.

Moving away from HCI literature to examine how walking affects the user's mental performance, Lajoie *et al.* [6] discussed attentional demands for static and dynamic tasks. The dual-task procedure was used where subjects reacted to auditory stimuli in three different mobility positions (sitting, standing and walking). The results showed walking required more cognitive attention than standing and sitting down due to the additional motion of balance needed to walk. Pellecchia [14] examined the relationship between mental demands and muscle motor movements and indicated that by increasing the difficulty of the cognitive task resulted in greater body movements in terms of postural sway. This suggests there is a close relation between the person's cognitive abilities and human motor performance. If walking alone can cause increased mental stresses on the user, it is important to examine if encumbrance creates similar cognitive difficulties therefore affecting the user's ability to engage with mobile devices effectively when interaction is required.

The limited research on encumbrance suggests the topic is at its early stages and a better understanding into the usability problems it causes can be beneficial to mobile device users. Although many studies have examined the issues of walking during interaction and solutions have been developed to enhance the user experience, the effects of mobility combined with encumbrance may result in much greater problems therefore it is crucial to examine if such issues do occur. The next section of the paper will discuss the methodology of our experiment which was conducted to investigate the impact of carrying different types of objects on targeting performance on a touchscreen mobile phone.

3 Methodology

The section will be split into three parts to describe the procedure taken for the encumbrance experiment.

The first part will discuss the initial observational study conducted to classify the most common objects that users were seen to carry and to choose two types of those objects to be evaluated in the experiment. The second part will describe the use of a motion capture system to detect hand postures and body movements and how users interacted with the mobile phone when encumbered and walking. The third part will describe the target acquisition task and the design of the experiment.

3.1 Common Types of Encumbrance Objects

The main purpose of the observational study was to examine the objects that users were seen to carry regularly while using a mobile device and to group those objects into suitable categories.

Since there is a vast amount of possible objects that could encumber the user, it is essential to identify and concentrate on the most common types of objects that are likely to be held during interaction. In order to define a set of encumbrance objects to be assessed in the experiment, three different types of public locations were observed (main street, transport station and supermarket) to examine the wide variety of objects that people held and carried. The experimenter would observe the general public during peak times for two hours at each type of location (for example, early commuting hours between 8am to 10am at a railway station and lunch period between 12pm and 2pm in a supermarket) since there will be a great influx of people and the probability of seeing a range of different objects is increased. Two different sites for each type of location were observed which resulted in six set of data. Each object seen being held or carried was noted down in terms of the following characteristics: *type*, *shape (rectangular or round?), size (length, width & thickness), quantity, input hand (non-dominant, dominant or both), hand action and grip required, and arm posture*.

Once the observational study was completed, the data collected was firstly grouped by object type and then sorted by how often it was noted down during the study. The results showed that different types of bags were the most frequently held objects as it account for 49% of all the items recorded. Boxes were the second most common object with 35% while the remaining 16% of objects documented ranged from beverage cups, umbrellas, specialized equipment to children, prams and pet on leashes. The objects recorded were also separately categorized based on the arm movement and hand actions required by the user. There were four main categories: 1. *swinging* – objects such as bags and holding a child's hand were placed into this group as it caused the arm to swing somewhat unpredictably; 2. *bulky* - boxes were grouped into this class as people were seen to hold different types of boxes normally underarm which required an awkward but assuring grip from the arm to prevent the object from falling to the ground; 3. *push and pull* - objects such as prams and trolleys were put into the *push* class while people were seen *pulling* wheeled luggage; and 4. *complex* – objects in this category included keys, wallets and hot beverages which require more careful and intricate finger action and grip.

Based on our observations of people in the public environment, different types of bags and boxes were chosen as the encumbrance objects to be evaluated in the experiment. Since there was a great variety of bags and boxes noted during the observation, the decision was made to evaluate two different types of each object based on its size and shape. Therefore, the experiment assessed two bags (small and medium sized) and two boxes (thin and thick broadness). The small bag represented a handbag while the medium bag simulated people carrying a shopping bag. The dimensions (width x height x depth) of the small and medium bags were 35 x 25 x 17 cm and 45 x 55 x 25 cm respectively. The thin and thick boxes measured 37 x 30 x 15 cm and 39 x 30 x 29 cm respectively. The bags were held in hand while the boxes were carried underarm as people were seen to adopt these strategies during the observational study. All bags and boxes weighed 3kg each to keep the object's heaviness consistent. The weight of 3kg was chosen to simulate the effects of carrying realistic objects that would make interaction physically difficult yet limit the amount of physical straining on the participants. The objects and the method that they were held during the experiment are shown in Figure 1.

Fig. 1. The four objects evaluated in the experiment. From left to right: small bag, medium bag, thin box and broader box. The images show how each object was held in the non-dominant hand or arm.

3.2 Tracking Hand and Arm Motions

To understand how holding the different objects encumbered the user and to analyze the level of hand and arm instability, a Vicon motion capture camera system was used (http://www.vicon.com). Twelve infrared cameras sampling at 120Hz recorded body movements (to a thousandth of a millimetre) in three-dimensions by tracking reflective markers which were firmly attached to each participant at specific body locations to track their movements while performing the experimental task in the capturing volume. The reflective markers were placed on the front and back of the upper torso and the hand and arm areas. A total of 15 markers were attached to each participant and Figure 2 illustrates their location.

The marker on the neck was used to calculate the total distance walked and the average walking speed for each participant. The right and left shoulder and left thumb markers (all participants were right handed therefore the mobile device was held in the left, non-dominant hand) were used to determine the relative position between the device and the user to calculate the amount of hand movement along each dimension. One marker was attached to the right index finger (intermediate phalanx section) to track the motion of the input finger. It would have been more appropriate to place the marker on the tip of the index finger but due to the size of the markers it would have obscured part of the touchscreen and made targeting more difficult. The remaining markers were used to define sections of the body. Participants were asked to avoid wearing loose clothing and long hair was tied up in a head cap provided to avoid excessive marker movement and prevent the markers from being occluded.

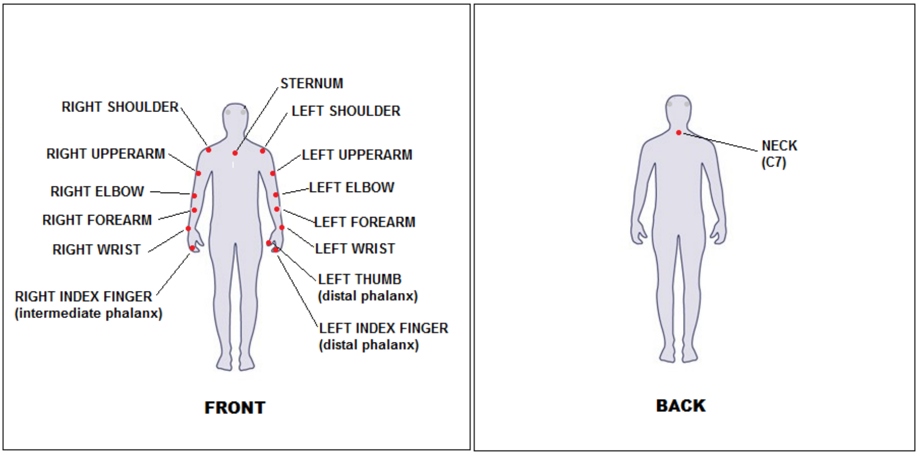


Fig. 2. The location of the reflective markers (red dots) placed on each participant

3.3 Experimental Task

The experimental task was to select a series of crosshair targets one at a time on a touchscreen mobile phone. The participants had to select the current target crosshair (colored green) as quickly and as accurately as possible. The screen also showed the location of the next target (colored red) so that the participants always knew where to input next until they reached the last target. There were one hundred targets aligned in a 10 x 10 grid across the screen which were randomly ordered for each condition. Each target border measured (width x height) 40 by 60 pixels with the central crosshair measuring 10 pixels in each direction (1 pixel = 0.1mm). A gap was created between the last row of the targets and bottom of the touchscreen to prevent the participants accidentally tapping the soft keys. The task ran on a Google Nexus One Android 3.1 smartphone as shown in Figure 3.

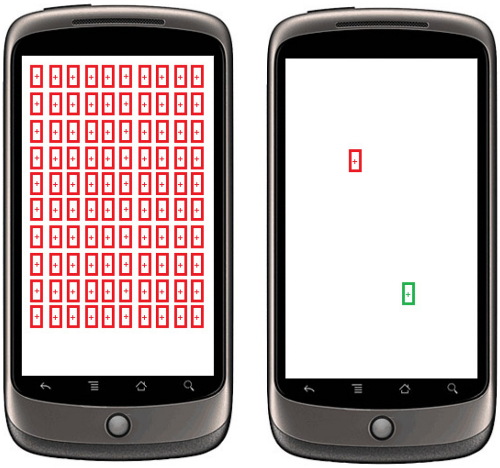


Fig. 3. The image on the left side illustrates the layout of the targets while the experimental task is shown in the right side image

3.4 Experimental Design

A within-subjects design was used for the experiment and each participant completed the target acquisition task while unencumbered (holding nothing) and carrying each of the four objects in either the dominant or non-dominant side.

As a result, there were nine encumbrance levels and each level was evaluated either standing still or walking which gave 18 conditions in total. The participants stood at the centre of the capturing volume for the standing conditions and navigated a pre-defined rectangular route (the outer and inner borders were 2.8 x 3 meters and 1 x 1.4 meters respectively) for the walking conditions as shown in Figure 4. The dimensions of the route were limited to the position of the cameras. Participants were instructed to keep within the path and walked in a clockwise direction. Each participant's preferred walking speed (PWS) was calculated before the experiment began. The participants were asked to navigate the route for two minutes at a pace that they normally would if they were walking on a quiet street.

The conditions were randomly ordered for each participant to reduce learning effects. 18 participants (4 males, 14 females) aged between 19 - 38 years and all right handed were recruited from the university to take part in the experiment. The Independent Variables were encumbrance and mobility. The Dependent Variables were targeting accuracy, targeting speed and the level of movement in the non-dominant hand holding the device. The hypotheses were:

H1. Targeting accuracy significantly decreases when the user is encumbered and walking compared to unencumbered and standing still.

H2. Targeting speed is significantly slower when the user is encumbered and walking compared to unencumbered and standing still

H3. Targeting is less accurate when the dominant hand or arm is encumbered compared to the non-dominant hand or arm carrying the objects.

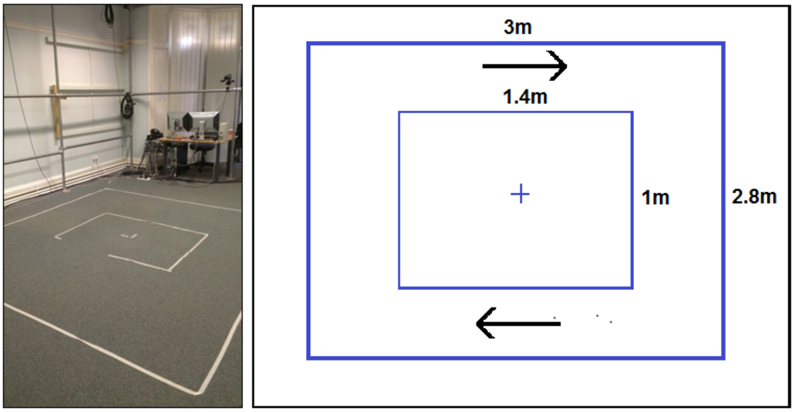


Fig. 4. The left image shows the actual capturing area. The right image illustrates the dimensions of the pre-defined path. Participants stood at the centre for the standing conditions.

4 Results

To eliminate unintentional target selections, the recorded target positions that were greater than 70 pixels horizontally and 110 pixels vertically from the centre of the target crosshair were removed from the final data analysis.

The reason for choosing the specific limit was to permit one target size margin of error. Target accuracy was measured as the distance (in pixels) from the centre of the target crosshair to the position recorded on the touchscreen. Speed of input (in seconds) was the time taken to select the current target.

4.1 Target Accuracy and Target Speed

A two-factor repeated-measures ANOVA with type of encumbrance and mobility as factors was calculated to examine both target accuracy and target speed.

The x-axis and y-axis were analysed independently for accuracy to assess if there was more error in a particular direction. For target accuracy on the x-axis, there was a significant main effect for stance, $F(1,17) = 69.358$, $p < 0.05$ and for encumbrance, $F(8,136) = 7.131$, $p < 0.05$. The interaction was also significant $F(8,136) = 2.658$,

$p < 0.05$. A pairwise comparison for encumbrance with Bonferroni corrections showed that unencumbered was more accurate than holding the bags in either hands and carrying the thin and thick boxes under the dominant arm. However, carrying either the thin or thick box under the non-dominant arm was not significantly less accurate than holding no objects. Also, carrying the thin box under the non-dominant arm was more accurate than holding the wider box under the dominant arm. Table 1 illustrates the pairwise comparisons that were significant for encumbrance on targeting accuracy along the x-axis.

Table 1. The table shows the pairwise comparisons for encumbrance that were significantly different for accuracy on the x-axis. *Adjustment for multiple comparisons: Bonferroni.

Comparison		**Mean Diff.**	**Std. Err.**	**Sig.***
Unencumbered	Small bag (non)	-2.718	0.690	0.038
Unencumbered	Small bag (dom)	-3.490	0.874	0.034
Unencumbered	Medium bag (non)	-3.366	0.461	0.000
Unencumbered	Medium bag (dom)	-5.248	0.973	0.002
Unencumbered	Thin box (dom)	-4.463	1.048	0.019
Unencumbered	Thick box (dom)	-4.952	0.928	0.002
Thin box (non)	Thick box (dom)	-4.063	0.966	0.021

Similarly for input accuracy along y-axis of the screen, there was a significant main effect for mobility, $F(1,17) = 25.901$, $p < 0.05$ and for encumbrance, $F(8,136) = 4.022$, $p < 0.05$. The interaction between the two factors was also significant $F(8,136) = 2.334$, $p < 0.05$. The participants were more accurate while standing than walking. A pairwise comparison with Bonferroni adjustment showed that unencumbered was only significantly more accurate in the vertical direction than holding the medium bag in the dominant hand (mean difference = -3.927, std. error = 0.933 and $p = 0.021$). All other encumbrance comparisons were not significantly different ($p > 0.05$). Figures 5a and 5b illustrate the mean target accuracy for the x-axis and y-axis of the touchscreen respectively. Based on the results for accuracy, hypothesis H1 is rejected since holding the boxes in the non-dominant hand was not significantly less accurate than unencumbered. Also, carrying the medium bag in the dominant hand was the only hindrance that caused accuracy to be significantly worse than holding no objects in the y-axis. Standing was significantly more accurate than walking.

The results for targeting speed from conducting an ANOVA indicated a significant main effect for encumbrance; $F(8,136) = 13.239$, $p < 0.05$ and mobility; $F(1,17) = 12.230$, $p < 0.05$. The interaction was also significant $F(8,136) = 3.257$, $p < 0.05$. Unexpectedly, holding the medium bag and both boxes in the non-dominant side caused selection times to be significantly lower than unencumbered. Hypothesis H2 is rejected based on this finding. Also, targets took significantly less time to select when carrying each of the four objects in the non-dominant hand or arm than the dominant side. Figure 6 illustrates the mean targeting times for each condition.

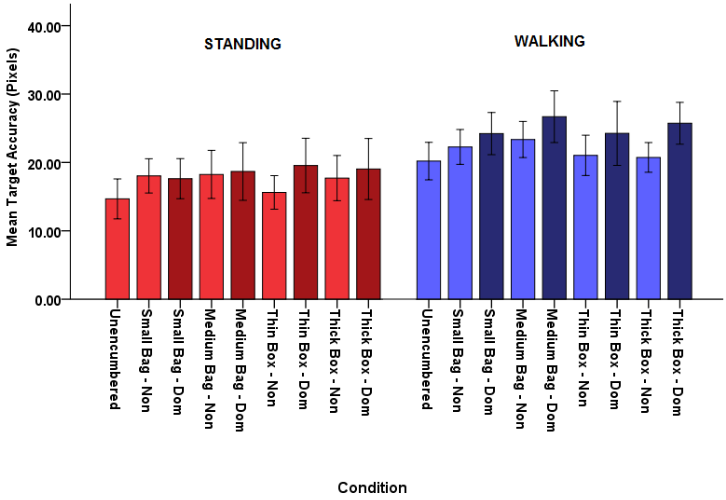


Fig. 5a. Mean target accuracy along the x-axis of the touchscreen (95% CI). The red and blue bars represent the standing and walking conditions respectively. The lighter and darker shade of each color indicates holding the objects in the non-dominant and dominant sides respectively.

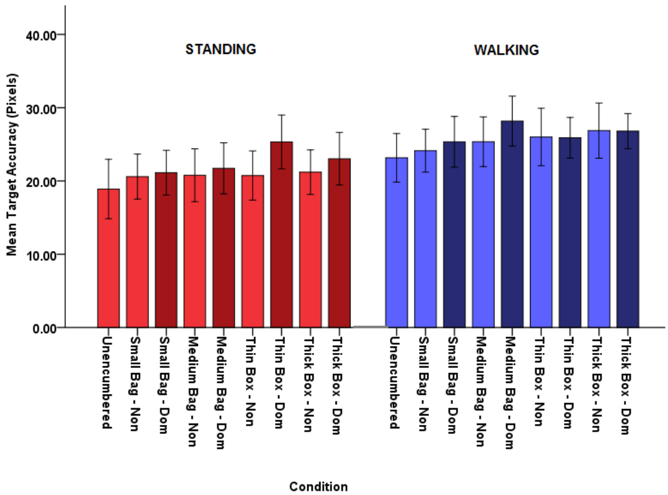


Fig. 5b. Mean target accuracy along the y-axis of the touchscreen (95% CI). The color representation for the conditions is the same as Fig. 5a.

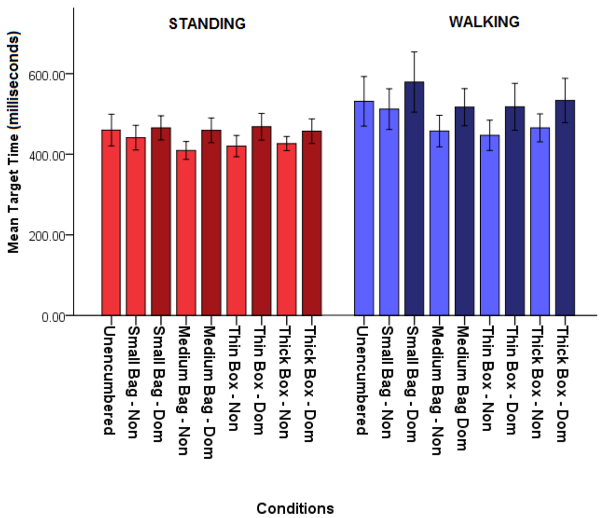


Fig. 6. The mean targeting speed (seconds) for each condition (95% CI). The color representation for the conditions is the same as Fig. 5a.

4.2 Preferred Walking Speed (PWS)

Table 2 shows the drop in the participants' preferred walking speed (PWS) for each type of encumbrance while performing the task compared to walking alone and no interaction.

Holding the broader box under the non-dominant arm caused the highest decline in PWS of approximately 41%. Each encumbered condition's drop in PWS was greater than the results reported by [2], who found a 24% decline in PWS when performing a similar pointing task to ours on a treadmill. This suggests that carrying different types of objects during interaction may further affect the user's walking pace.

Table 2. The decrease in PWS (%) while holding the different types of bags and boxes compared to no interaction

Condition	Drop in PWS (%)	
Unencumbered	16.68	SD = 9.46
Small bag (non-dominant)	25.17	SD = 11.96
Small bag (dominant)	27.64	SD = 12.09
Medium bag (non-dominant)	30.48	SD = 13.17
Medium bag (dominant)	31.49	SD = 13.17
Thin box (non-dominant)	36.88	SD = 13.45
Thin box (dominant)	33.48	SD = 14.76
Thick box (non-dominant)	41.06	SD = 11.64
Thick box (dominant)	36.48	SD = 11.67

4.3 Analysing Hand Movements

The motion capture data was processed to examine the movements of the non-dominant hand which held the mobile phone to see if carrying the different types of objects chosen for our experiment caused the user difficulty to steady their hands for input.

The markers placed on the non-dominant hand (left thumb and left index finger) and both shoulders were used to calculate the relative position between the device and the participant to determine the change in movement on each axis. The x-axis, y-axis and z-axis of motion represents left – right, forward – backward and upwards – downwards movements respectively. These notations will be used when discussing hand movements to avoid confusion with describing target accuracy.

Encumbrance was grouped by the hand or arm carrying the object (i.e. dominant vs. non-dominant) rather than the actual object itself when analyzing hand movements. This was a more appropriate method to understand how holding different objects impacts handedness and to identify where the problem is occurring. A two-way ANOVA with mobility and handedness (non-dominant vs. dominant) as factors was calculated to examine its effect on input accuracy. The results showed there was a significant main effect for mobility; $F(1,71) = 133.369$, $p < 0.05$ and for handedness; $F(1,71) = 19.309$, $p < 0.05$. The interaction was also significant $F(1,71) = 12.753$, $p < 0.05$. Pairwise comparisons with Bonferroni adjustment showed the participants were less accurate when the dominant hand was encumbered compared to the non-dominant hand. The mean target accuracies for the dominant and non-dominant sides while standing and walking are shown in Table 3.

Table 3. The mean target accuracy (pixels) when the non-dominant and dominant hand/arm was encumbered

	Target Accuracy (pixels)	
Non-dominant. – Standing	17.402	SD = 5.965
Dominant – Standing	18.723	SD = 7.784
Non-dominant. – Walking	21.846	SD = 5.201
Dominant – Walking	25.221	SD = 7.370

The amount of movement of the non-dominant hand holding the mobile phone was assessed in each direction separately to see if encumbrance caused more movement and instability on a particular axis. The mean hand movement (mm) was calculated by processing the amount of motion between three tenths of a second prior to a target being selected and the instance the tap onscreen was recorded.

A two-factor repeated-measures ANOVA with mobility and handedness as factors was conducted in each direction to assess if there was a significant difference in mean movement on the hand or arm carrying the bag and the box. For left and right direction of movement, there was a significant main effect for mobility; $F(1,27) = 791.591$, $p < 0.05$ and encumbered hand; $F(1,27) = 240.090$, $p < 0.05$. The interaction was significant $F(1,27) = 51.422$, $p < 0.05$. For forward and backward movement, ANOVA showed a significant main effect for mobility; $F(1,27) = 855.252$, $p < 0.05$

and encumbered hand; $F(1,27) = 188.013$, $p < 0.05$. The interaction was significant $F(1,27) = 24.424$, $p < 0.05$. The results for upward and downward movement indicated a significant main effect for stance; $F(1,27) = 1255.995$, $p < 0.05$ and encumbered hand; $F(1,27) = 61.908$, $p < 0.05$. The interaction was significant $F(1,27) = 459.970$, $p < 0.05$. Table 4 shows the mean movement in each direction of the non-dominant hand while encumbered and either standing still and walking the route.

Table 4. The mean movement of the left hand holding the mobile phone in each direction while encumbered. (S) and (W) represent standing and walking respectively.

Encumbered Hand/Arm	Movement Direction	Mean Movement (mm)	Std. Dev.
Non-Dominant (S)	Left/Right	1.411	0.262
Dominant (S)	Left/Right	4.781	1.11
Non-Dominant (W)	Left/Right	5.338	0.922
Dominant (W)	Left/Right	6.648	0.611
Non-Dominant (S)	Forward / Backward	0.697	0.232
Dominant (S)	Forward / Backward	2.505	0.414
Non-Dominant (W)	Forward / Backward	4.044	0.667
Dominant (W)	Forward / Backward	4.862	0.760
Non-Dominant (S)	Upward/Downward	0.991	0.273
Dominant (S)	Upward/Downward	3.286	0.499
Non-Dominant (W)	Upward/Downward	5.939	0.550
Dominant (W)	Upward/Downward	6.365	0.664

When the participants were standing still, carrying the objects in the dominant hand caused the other hand (holding the device) to move significantly more in each of the three directions when compared to the non-dominant hand being encumbered. The difference in mean hand movement between the non-dominant and dominant side being encumbered was very similar while walking. However, walking caused a significant increase of movement in both the non-dominant and dominant sides when compared to standing still. The results support hypothesis H3 as holding the bags and boxes in the dominant side resulted in significantly more movement in the hand holding the device and caused input to become less accurate when compared to the non-dominant side being hindered.

5 Discussion

The results from the experiment have shown that holding different types of bags in hand or carrying boxes underarm caused a negative impact on targeting performance on a touchscreen mobile phone.

The participant's input performance has illustrated that holding the objects chosen for our experiment caused targeting to become less accurate particularly when the object is being held in the dominant hand which was also targeting at the touchscreen to input at the same time. Interestingly, there was no significant effect for accuracy between unencumbered and carrying the two differently sized boxes under the non-dominant arm. Furthermore, performing the experimental task while unencumbered took significantly longer to input than holding the medium bag and both types of boxes. [10] reported the performance of wrist gestures while carrying a box under the dominant arm was similar to unencumbered since users were actually able to use the box to assist their input. In terms of assessing the effects of mobility, walking while encumbered in general caused input to become less accurate and required more time to select the targets onscreen. The non-dominant hand which held the mobile phone was tracked to examine the level of movement caused by carrying the different types of objects. There was significantly more movement and instability when the dominant hand or arm was hampered compared to the non-dominant side. The difference in the amount of movement between encumbering the dominant and non-dominant sides was very similar when walking compared to standing still which may suggest walking alone could have an overwhelming effect on interaction.

Observing how the participants held the bags and boxes provided a valuable insight of some of the physical difficulties that users may experience when interacting with mobile devices while encumbered. The expectation was that carrying the thinner box underarm would be physically less challenging than the wider box. However, watching the participants suggested that the thinner box caused more problems to hold in place to avoid it being dropped to the ground especially when walking the route. Comments from the participants revealed the thinner box kept slipping down from their arms and it was difficult to find a comfortable gripping position while performing target selections, especially when carrying the box under the dominant arm. Surprisingly, the broader box did not cause as much physical issues as the thin box once a secure carrying posture was found. But, a few participants did find it uncomfortable to carry the wider box over a long period of time due to the length of their arms which may have caused selecting the targets more difficult and performance to decline.

The majority of participants had little trouble holding the two types of bags during interaction as the physical issues were more due to fatigue and tiredness. A number of participants commented that it was physically unpleasant to carry the bags due to the handles causing discomfort to their hands which may have caused targeting performance to decline. Holding the medium bag required a more restrictive upright arm position to prevent the bag from touching the floor. The participants mentioned that it was more demanding to hold both type of bags in the dominant hand and input at the same time compared to carrying the objects in the non-dominant hand. Furthermore, when standing still, it was evident that the dominant hand started to move downwards as the arm began to tire which would have made viewing the touchscreen more awkward and caused input to become more challenging. Remarks from the participants suggested that holding both types of bags in the dominant hand while walking the route required considerably more effort to control and maintain the forearm in a steady posture to target at the screen as the bag unpredictably swings from side to side. Also, it was fascinating to examine the different strategies adopted by the participants to carrying the different types of bags and boxes. Although we instructed the participants to carry the boxes underarm and the bags in hand, each participant's method of carrying the objects was slightly different and unique to the individual. This was more apparent when holding the box underarm as the participants with shorter arm lengths found it more challenging to grip the box and had to adjust their arm position at regular intervals to make input easier.

A simple recommendation that may improve input performance of two handed interaction while encumbered is to avoid carrying objects in the dominant hand or arm. However, in realistic situations, it may not be possible to switch holding the objects to the non-dominant hand only as the user could be occupied by carrying multiple objects. We need to design and develop effective techniques and applications to help users input more accurately when encumbering the dominant hand is unavoidable. One possibility is to increase target size to give the user a bigger margin of error. However, this would require careful interface design as it limits onscreen space for other widgets as discussed by [4]. We also propose that future studies should revisit previous research which have developed enhanced methods to assist users to input more efficiently while on the move. For example, Goel *et al.* [3]'s WalkType application has shown to improve text entry while the user is walking. A repeated study that examined both encumbrance and mobility would indicate if the application is still as effective and could be used to reduce some of the issues caused by encumbrance. A final suggestion on future studies is to assess the impact of carrying a wider range of objects and include different hand gripping positions such as those that require a pushing action to extend our knowledge into the effects of encumbrance on mobile interactions.

6 Conclusions

The main purpose of the study was to examine the impact of encumbrance and make researchers aware of the usability issues and difficulties when users are physically hampered during interaction with mobile devices.

An observational study was conducted to investigate the typical objects that people frequently held and carried in their everyday activities. Based on the results, we evaluated different types of bags and boxes to simulate realistic cumbersome and physically challenging situations that users are likely to encounter. The results from our main experiment showed that encumbrance caused targeting accuracy on a touchscreen phone to decline compared to holding no objects. However, targeting while carrying the thinner box under the non-dominant arm was significantly quicker and more accurate than unencumbered or holding the box in the dominant side. Moreover, targeting was less accurate when the dominant hand or arm was physically hampered during interaction and tracking the motion of the left hand (which held the mobile phone) showed there was significantly more hand movements compared to the non-dominant side carrying the objects. This suggests that we should focus on assisting the user to input more accurately especially when the dominant hand and arm is encumbered.

Encumbrance also affected the user's preferred walking speed and although input was significantly less precise when the user was walking compared to standing, the performance difference was not as great as expected. Further investigation is required to see if there is a tradeoff between walking speed and encumbrance and assess input performance when the user has to maintain their PWS while encumbered. We hope the study presented in this paper has motivated researchers to develop better interaction techniques to aid users to input more effectively on touchscreen mobile devices.

References

1. Brewster, S.A.: Overcoming the lack of screen space on mobile computer. Personal and Ubiquitous Computing 6, 188–205 (2002)
2. Bergstrom-Lehtovirta, J., Oulasvirta, A., Brewster, S.: The Effects of Walking Speed on Target Acquisition on a Touchscreen Interface. In: Proc. MobileHCI 2011, pp. 143–146. ACM Press (2011)
3. Goel, M., FindLater, L., Wobbrock, J.: WalkType: using accelerometer data to accommodate situational impairments in mobile touchscreen text entry. In: Proc. CHI 2012, pp. 2687–2696. ACM Press (2012)
4. Kane, S.K., Wobbrock, J.O., Smith, I.E.: Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. In: Proc. MobileHCI 2008, pp. 109–118. ACM Press (2008)
5. Knoblauch, R., Pietrucha, M., Nitzburg, M.: Transportation research Record 1538(-1), 27–38 (1996)
6. Lajoie, Y., Teasdale, N., Bard, C., Fleury, M.: Attentional demands for static and dynamic equilibrium. Experimental Brain Research 97(1), 139–144 (1993)
7. Lin, M., Goldman, R., Price, K.J., Sears, A., Jacko, J.: How do people tap when walking? An empirical investigation of nomadic data entry. Intl. J. of Human Computer Studies 65(9), 759–769 (2007)
8. Mainwaring, S., Anderson, K., Chang, M.: Living for the global city: mobile kits, urban interfaces, and ubicomp. In: Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 269–286. Springer, Heidelberg (2005)
9. Mizobuchi, S., Chignell, M., Newton, D.: Mobile text entry: relationship between walking speed and text input task difficulty. In: Proc. MobileHCI 2005, pp. 122–128. ACM Press (2005)
10. Ng, A., Brewster, S., Crossan, A.: The Effects of Encumbrance on Mobile Gesture Interactions. In: Proc. of Body, Movement, Gestures & Tactility in Interaction with Mobile Devices, Workshop at MobileHCI 2011, Stockholm, Sweden (2011)
11. Nicolau, H., Jorge, J.: Touch typing using thumbs: und understanding the effect of mobility and hand posture. In: Proc. CHI 2012, pp. 2683–2686. ACM Press (2012)
12. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: Proc. CHI 2005, pp. 919–928. ACM Press (2005)
13. Oulasvirta, A., Bergstrom-Lehtovirta, J.: Ease of juggling: studying the effects of manual multitasking. In: Proc. CHI 2011, pp. 3103–3112. ACM Press (2011)
14. Pellecchia, G.: Postural sway increases with attentional demands of concurrent cognitive task. Gait Posture 18(1), 29–34 (2003)
15. Schildbach, B., Rukzio, E.: Investigating selection and reading performance on a mobile phone while walking. In: Proc. MobileHCI 2010, pp. 93–102. ACM Press (2010)
16. Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A.: Understanding mobile contexts. Personal Ubiquitous Computing 8, 135–143 (2004)
17. Wolf, K.: Microinteractions beside ongoing manual tasks. In: Proc. TEI 2011, pp. 447–448. ACM Press (2011)

Conception of Ambiguous Mapping and Transformation Models

Christopher Martin, Matthias Freund, Henning Hager, and Annerose Braune

Institute of Automation, Technische Universität Dresden, 01062 Dresden
{christopher.martin,matthias.freund,
annerose.braune}@tu-dresden.de,
henning.hager@monkey-works.de

Abstract. Model transformations are the linking element between the different levels of abstraction in the model-based user interface development. They map source elements onto target elements and define rules for the execution of these mappings. Approaches for the reuse of transformation rules use formal transformation models, which only specify the mappings and abstract from the implementation. Current solutions are usually only able to describe unambiguous (1-on-1) mappings. In general, however, there are ambiguous (1-on-n) mappings from which the unambiguous mappings are only chosen during the design process. The knowledge which source element can be mapped onto which target elements is to date not being formalized. This paper therefore presents a proposal for an ambiguous mapping and transformation model and describes its usage in an iterative development process.

Keywords: Model-based User Interface Design, Model Transformations, Mapping Model, Transformation Model, Iterative Development.

1 Introduction

Developing user interfaces is a cost- and time-consuming task, especially if multiple platforms in terms of hard- and software have to be supported. Model-driven development processes for user interfaces (UIs) try to handle this problem by allowing a designer to specify the functionality of a user interface by means of models. These models abstract from implementation details and can ideally be used to generate executable user interfaces for different target platforms. In general, the models used in such model-driven development processes are defined on different layers of abstraction. The generally accepted *Cameleon Reference Framework* (*CRF*) constitutes four layers of abstraction [1, 2]: The *Task & Concepts* (*T&C*) level allows defining the tasks that the user shall execute with the user interface. An *Abstract User Interface* (*AUI*) model describes the UI in a modality-independent way. In contrast, a *Concrete User Interface* (*CUI*) model is already bound to a specific modality but still independent of an underlying computing platform. A *Final User Interface* (*FUI*) is an executable UI dependent on a specific (class of) computing platform(s).

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 110–125, 2013.

Model transformations either convert more abstract user interface models into increasingly concrete models (cf. the vertical transformations in Fig. 1) or they transfer models of a certain level of abstraction to another context of use (cf. the horizontal transformations in Fig. 1). In any case, transformations constitute the connection between the different layers. Therefore, their quality and reusability is a key factor for the efficiency of the entire model-based approach.

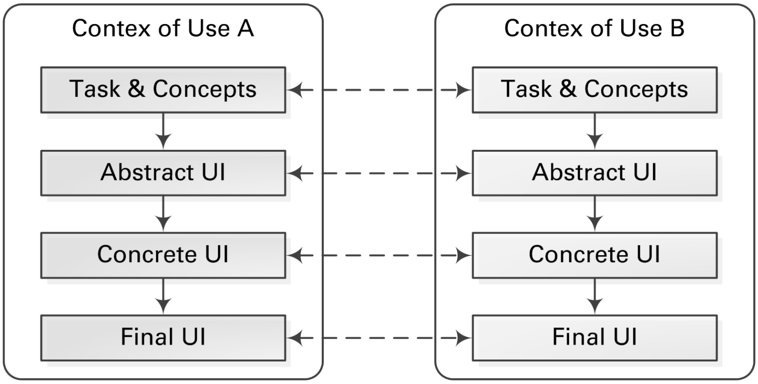


Fig. 1. Simplified version of the Cameleon Reference Framework [1] as introduced in [2]

Model transformation rules incorporate two basic tasks (comp. [3]): (1) the description of mappings from source elements onto target elements and (2) the description of the execution of these mappings including the generation of the target model(s). While standard (hard coded) model transformations usually do not separate the mappings from their execution but combine them in the transformation code, there are already some approaches that favor the definition of formal mapping respectively transformation models. These models describe which model element or meta-model element of the source language can or shall be mapped onto which element of the target language. Thereby, they are independent of the technology that is used afterwards for the implementation of the transformation execution. Consequently, as shown in Fig. 2, they separate the functionality of the transformation (the description of mappings) from the implementation details (the execution of these mappings). The advantage of mapping models is that the correlations between elements of the source and the target (meta-)model can be configured instead of having to be programmed. In this case, the creation of a model transformation only requires knowledge about the application domain and not about implementation language details.

Existing mapping model specifications essentially only allow unambiguous (1-on-1) mappings that map one source element onto exactly one target element. For example, a mapping model from AUI onto CUI elements would describe that – in case of a graphical user interface – every modeled AUI-level *"selection element"* has to be mapped to e.g. the drop-down list element at CUI-level.

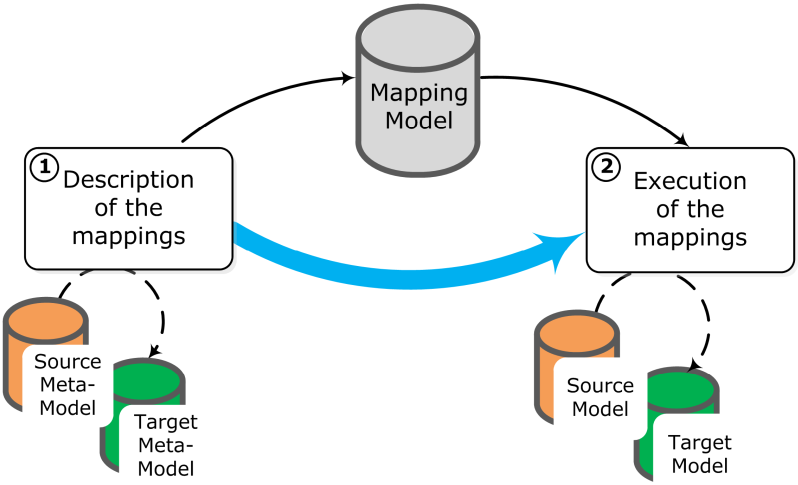


Fig. 2. Separation between description and execution of mappings by means of a mapping model

However, Puerta and Eisenstein show that not unambiguous (1-on-1) but ambiguous (1-on-n) mappings are frequent for the model-driven development of user interfaces [4]. That means, that the ”selection element” instruction at AUI-level in general could be visualized by means of a drop-down list as well as by a set of radio buttons or a vocal selection. Consequently, mapping models for user interface design require multiple instead of unique target model elements. During the design process these ambiguities have to be resolved to allow the transformation of the model. Which of the (ambiguous) mappings is used depends, among others, on the target platform, the developer's preferences and on company guidelines. It is important to note that the resolving of the ambiguities has to be performed for each element of the source model as, in general, not every instance of a source meta-model element shall be mapped to an instance of the same element of the target meta-model. For example, depending on the number of the possible choices of a “selection element”, either a drop-down list or radio buttons could be reasonable.

The model-based user interface development has to support iterative development processes to be widely accepted by UI designers because user interfaces are usually not developed in a single linear development process but rather in multiple development cycles. Therefore, the proposed mapping model should not only be able to specify ambiguous mappings but also save the information about the chosen mappings for each element. This allows for an easy reuse of transformations and the chosen mappings for every iteration step of the development process.

Hence, this paper presents a meta-model that (a) supports the definition of ambiguous mappings for the transformation of UI models and (b) stores the chosen mappings of the executed transformation.

2 Related Work

In regular transformations, both the mapping of a source element onto a target element as well as the generation of the target model are hard coded for one specific pair of UI description languages in an implementation technology such as ATL [5], QVT [6] or XSLT [7]. The occurring possible (ambiguous) mappings are usually either not defined so that only one of the possible target elements is used or only programmed directly into the transformation in an informal way [8].

Bézivin et al. show in a fundamental work that, in contrast to the hard coded transformations, it is also possible to formally specify the mappings in so called *Transformation Models* [9]. Therefore, the mappings of the meta-model elements are described by formally assigning source and target elements to each other. This approach is used by the UI description language UsiXML, which specifies a transformation model that allows the description of a model transformation based on graph grammar [10]. The MARIA[1] description language family [11] also allows the definition of unambiguous mappings with the help of the graphical editor *MARIAE* [12]. However, in both cases the unambiguous mappings apply to every element of the same type, i.e. every instance of the same meta-model element. This means that e.g. every "*selection element"* can only be mapped to a drop-down list for the entire user interface. An alternative mapping to a radio button or a vocal selection is only possible in another transformation for a different context or by manually changing the target model. Consequently, there is no reusable information about the ambiguous mappings.

First approaches for ambiguous mapping models are made by Hager et al. [8] and by Aquino et al. [13]. Hager addresses this problem by implicitly describing all possible ambiguous mappings for one pair of description languages, namely *DISL*[2] [14] at the AUI level and *Movisa* [15] at the CUI level, in (hard coded) transformation rules. The UI designer can choose an unambiguous mapping in an interactive process. With this approach, a UI designer can map a DISL element type (AUI level) onto various Movisa element types (CUI level) even in a single user interface. The choices made are then stored in a *Persistent Transformation Mapping Model* (*Petmap*) and are available for further iteration steps. However, the ambiguous mappings for this pair of UI description languages are also not formalized as they are implicitly included in (hard-coded into) the transformation rules.

In [13], Aquino et al. propose *Transformation Templates* that extend the UsiXML Transformation Model defined in [10]. They allow defining (possibly ambiguous) mappings by the specification of a selector that is applied to the target meta-model. Furthermore, weighting factors can be specified that reference the UsiXML context model and that are used if multiple target elements for one source element are possible. However, both [8] and [13] specify their transformation models imperatively: They include implementation details of the transformation. Because of the imperative description of the mappings in this approach it is possible to use variables, conditions and loop constructs. For example, the specification of the selectors requires deep

[1] Modelbased lAnguage foR Interactive Applications.

[2] Dialog and Interface Specification Language.

knowledge of the syntax and semantics of the transformation templates and thus is not applicable by UI designers but only by transformation language experts.

In contrast to imperative modeling, declarative modeling of transformations simply states which elements shall be mapped onto each other but not how these mappings are implemented by the transformations: It describes the “what” and not the “how”. The advantage of a declarative modeling is the simple and clear notation suitable also for a UI designer without detailed knowledge of implementation technologies. The disadvantage of this approach is the limited language scope that e.g. does not allow expressing dependencies between different mappings.

Nevertheless, the creation of imperative transformation models requires advanced knowledge about the syntax and semantics of the implementation technology and therefore a transformation expert. Consequently, the approach presented in the following makes use of a declarative mapping model.

3 Conception of a Meta-model for the Description of Ambiguous Mappings

This section introduces the concepts used to realize the requirements for ambiguous mappings for transformations of UI models and for storing chosen concrete mappings derived in Section 1. The realization of these concepts as the *Persistent Ambiguous Mapping and Transformation Model* (*PAMTraM*) will also be described. The workflow of this solution including all involved models and meta-models as well as the role of the PAMTraM is shown in Fig. 3.

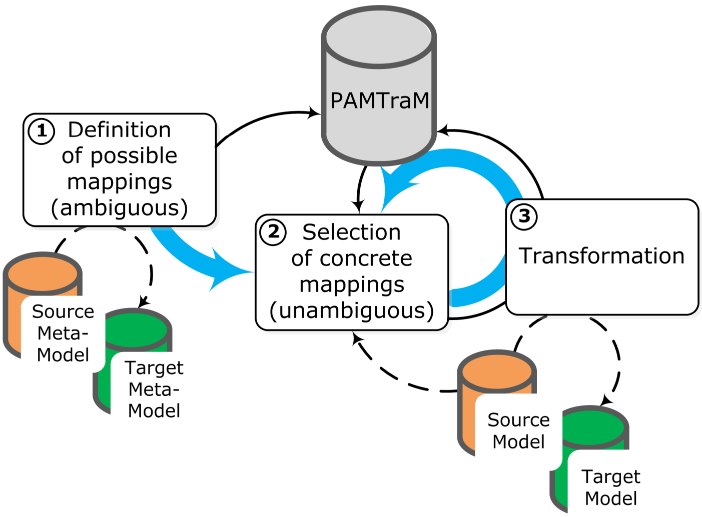


Fig. 3. Proposed workflow and involvement of the PAMTraM

During step 1, all ambiguous mappings from elements of the source meta-model onto elements of the target meta-model have to be defined. These mappings are stored in the PAMTraM. This step has only to be performed once for every language pair.

Afterwards, step 2 of the workflow consists of choosing the concrete mapping for every element of the source model. Therefore, the ambiguous mappings defined in step 1 are read from the PAMTraM and evaluated by a selector. This selector can e.g. be a UI designer or additional constraints (e.g. platform constraints) that allow an automatic evaluation.

Finally, step 3 constitutes the actual execution of the transformation. The execution uses the unambiguous mappings defined in step 2 to create the target model from the source model. Furthermore, the performed transformation including the selected unambiguous mappings is stored in the PAMTraM so that it can be reused for the next iteration cycles. In addition, the information about chosen mappings can e.g. be evaluated if statistical selection mechanisms are to be applied during step 2.

Steps 2 and 3 can be repeated to allow an iterative development process. The following sections describe the realization of the two parts of the PAMTraM by means of its meta-model.

3.1 Definition of Ambiguous Mappings

The mapping rules are defined on the meta-model level [3] because the allowed mappings of elements are mainly based on the UI description languages. They specify mappings from a source to one or more target meta-model elements. The mapping model must therefore allow the definition of those mappings and thus reference the source and the target meta-model. To ensure a comfortable creation of the mapping models, a declarative description will be implemented, such as e.g. already used in MARIAE's transformation editor for the definition of the unambiguous mappings. This approach allows the definition of mappings at a high level of abstraction. This way, a domain expert can create the mapping model and a transformation expert is not required.

The declarative description of the mappings, compared to an imperative approach, requires additional effort for the implementation of the mapping model into an executable transformation rule. But this additional effort is put into perspective, since such an implementation is only required once per UI description language pair. Another positive effect of the declarative description is the capability to easily define ambiguous mappings that assign one source element to *multiple* target elements. Fig. 4(a) shows the structure of a declarative, ambiguous mapping (*Mapping*). It enables the assignment of a source meta-model element to all eligible target meta-model elements.

As it is not reasonable to define multiple *Mapping* elements (each in turn defining multiple target meta-model elements) for one source meta-model element, additional constraints or validation steps should ensure that only one mapping is defined for every element of the source meta-model.

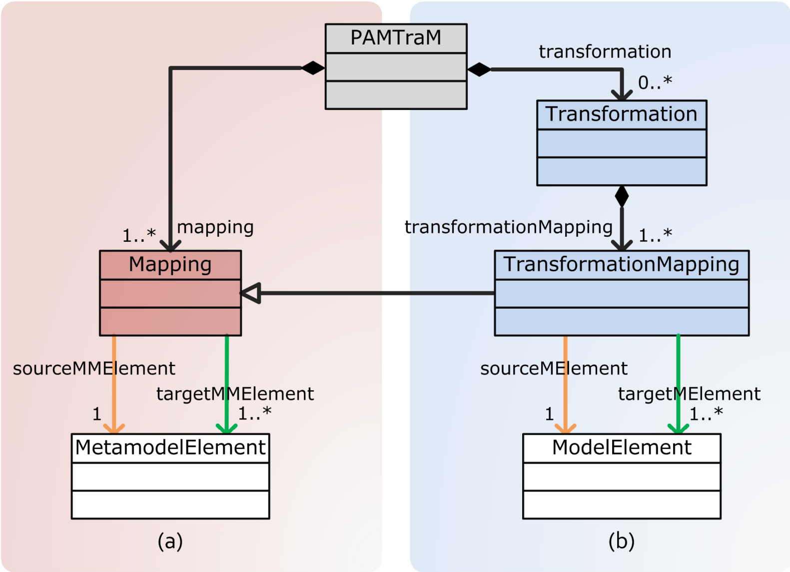


Fig. 4. Resulting implementation of the PAMTraM for the storage of ambiguous mappings (a) and performed transformations (b)

It is important to note that the element *MetamodelElement* can be *any* element of any meta-model – it is not limited to classes (if a UML-like structure is used for the meta-models) but can also be a class attribute. This is necessary because – depending on the meta-models used – it cannot be ensured that merely classes are mapped onto each other. This is best explained by a simple example: In DISL (see Section 2), the different widgets that can be used to realize a user interface are not represented by different classes of the meta-model but by different enumeration literals of the enumeration *GenericWidgetType*. Consequently, DISL uses a simple meta-model class *Widget* with an attribute *genericWidget* of the type *GenericWidgetType*. Fig. 5 shows the relevant excerpt of the DISL meta-model.

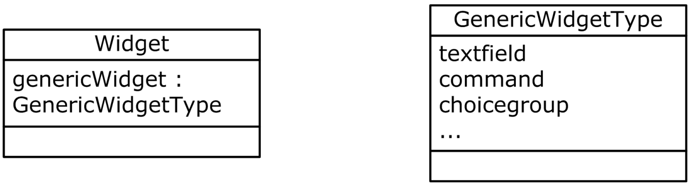


Fig. 5. Excerpt of the DISL meta-model: realization of widget types

Consequently, when using DISL as source or target meta-model, the mappings do not have to be specified for the class *Widget* but for the different enumeration literals like *textfield* or *command*.

3.2 Storage of Concrete Mappings

Modeling ambiguous mappings allows the explicit definition of all possible mappings of source to target element types. Before a concrete transformation (as an instance of the ambiguous mappings) can be executed, the ambiguities need to be resolved. This resolution depends on numerous factors such as the specific context of a user interface or industry- or company-specific guidelines. Furthermore, the ambiguities usually have to be resolved for every element individually, because it cannot be assumed that a source element type should always be mapped to the same target element type (cf. [16]). This process may be time-consuming – especially if a user is involved in the resolving of ambiguities – and has to be repeated for every iteration cycle even if only minor changes have been applied to the UI model. Consequently, the design decisions, which are made during a transformation, should be saved, so that they can be reused for iterative improvements of the user interface: In that case, an unambiguous mapping has only to be determined for the model elements that have been added or modified during the last iteration cycle whereas the stored concrete mappings of a previous transformation can be reused for the unaltered elements.

For this purpose, the meta-model element *Transformation* (see Fig. 4(b)) represents a concrete transformation that was executed for a specific user interface in a specific context, which can be saved in the PAMTraM. Gruhn et al. define a transformation as an instance of a mapping on the model level [3]. Therefore the unambiguous mappings that are determined for each model element during the transformation are saved as *TransformationMappings* (see Fig. 4(b)). Analogous to the *Mappings*, a declarative structure is used again. In comparison to the *Mappings*, *TransformationMappings*, however, map concrete (model) elements onto each other. Otherwise they have similar characteristics like the involved element types. Hence, an inheritance relationship between *TransformationMapping* and *Mapping* was used.

It should be noted that not necessarily only a single target element must be selected: Especially in multimodal user interfaces, it is often useful to map an AUI element to multiple CUI target elements, so as to combine a graphical output of an alarm with an auditory output of the same alarm. This relationship is represented by the *TransformationMapping* that allows the mapping onto multiple target model elements.

Those two (basically independent) requirements – modeling of ambiguous mappings and saving of performed transformations – were jointly implemented in a meta-model, since they are both specific for each pair of meta-models. In addition, when re-transforming a user interface, both, the basic mappings as well as the specific performed transformations, are needed. A common storage of all performed transformations in one model also allows for a good starting point for further developments on the reduction of the ambiguities like the statistical analysis of the determined mappings.

The resulting meta-model PAMTraM (cf. Fig. 4) is applicable for any combination of target and source meta-models, because there are no restrictions or statements made about their structure and only general concepts were implemented. Beyond that, it allows the use of different methods to resolve the ambiguities as it only stores the resulting mappings.

3.3 Tool Support for the Application of the PAMTraM

In order to make the workflow (depicted in Fig. 3) – and therewith the concepts introduced in Sections 3.1 and 3.2 – applicable, tools supporting the UI designer in performing the three steps of the workflow should be created. Therefore, we have implemented the PAMTraM by using the *Eclipse Modeling Framework* (*EMF*) [17].

During the first step of the workflow, the UI designer has to be supported in the definition of the ambiguous mappings. Although the EMF automatically creates simple, generic editors (e.g. a tree editor), the usability of these editors is not very high as they are not tailored towards the needs of the user. Instead, the design of a specialized *concrete syntax*[3] supporting the modeler to think in his domain is necessary [18].

In general, concrete syntaxes can be divided into *textual* and *graphical* representations, both with different advantages and disadvantages. For example, whereas a textual concrete syntax may enable a faster specification of models, a graphical concrete syntax may help in visualizing relationships. As multiple concrete syntaxes can be used simultaneously, the advantages of both types can be combined. As a first prototype, we have implemented a textual concrete syntax for the PAMTraM by means of the tool *EMFText* [19]. In order to assist the modeler, a content assistant and code completion are integrated.

Regarding the second step of the workflow, the necessary tool support can be divided into two parts: the implementation of the selection mechanism and the reuse of previous transformations in case of an iterative development.

As our approach does not dictate the used mechanism for the selection of concrete mappings during the second step of the workflow, the required tool support cannot be generalized. For example, if the UI designer shall choose the concrete mappings, a tool supporting this user interaction has to be implemented. This could e.g. be a dialogue that presents the different possibilities to the user and asks him to select one or more of the alternatives. If, however, statistical algorithms or platform constraints shall be applied, these mechanisms have to be implemented.

In case of an iterative development, a previous version of the user interface may already have been transformed once or repeatedly. In this case, the chosen (unambiguous) mappings for those parts of the source model that have not been altered during the iteration should usually be reused to keep the resulting parts of the target model that are not involved in the iterative changes consistent. Consequently, the selection mechanism should only be applied to those parts of the source model that have been changed during the iteration step. The described behavior has been implemented by means of a tool that reads previous transformations from the PAMTraM. Afterwards, the user is asked if one of the previous transformations (and with those the results of the used selection mechanism) shall be reused or if the selection mechanism shall be applied for every model element.

The tool support necessary for the third step of the workflow consists of the transformation executing the selected concrete mappings (and thereby generating the target model) and storing them in the PAMTraM for further iteration cycles (cf. Section 3.2). However, as this transformation is dependent on the used source and target meta-model, its implementation cannot be generalized either.

[3] The term *concrete syntax* denotes the concepts used for the representation of models.

An exemplary implementation of the tools supporting the workflow is presented in the following case study.

4 Case Study

This section demonstrates the application of the presented approach based on the development of a user interface for the process plant shown in Fig. 6. The process plant consists of three tanks that are connected via pipes. For this case study, the UI shall enable the user to select one of the three tanks resulting in the illustration of the fill level of the selected tank. Two possible final user interfaces are presented in Fig. 7: The UI on the left makes use of a set of radio buttons to select the tank and of an animated image to display its fill level. In contrast, the UI on the right uses a drop-down list for the selection of a tank and a textual representation of its fill level.

The starting point of the case study is an AUI model expressed in DISL. By application of the PAMTraM, a concrete model of the UI, based on MARIA, shall be generated. Therefore, the XML schemata of both languages have been converted to EMF-based meta-models for easier processing in the EMF environment used for the implementation of the tools described in Section 3.3. The (model-to-text) transformation generating the final UI has only been implemented prototypically to demonstrate possible representations of the resulting UI.

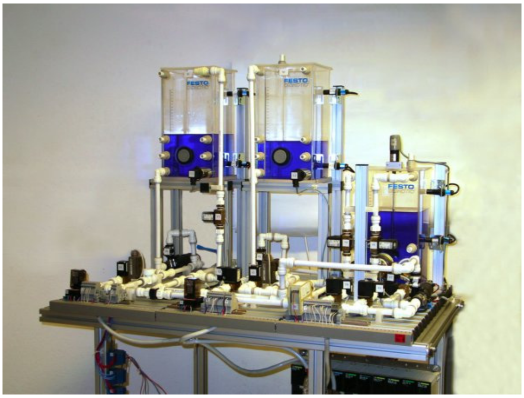


Fig. 6. Depiction of the process plant used in the case study

DISL is a dialogue descriptions language, which is situated on the AUI level of the Cameleon Reference Framework [14]. DISL offers ten *Widgets*, represented by *EnumerationLiterals* (cf. Fig. 5), for the description of user interfaces. In the DISL model (AUI model) of the case study, the structure of the realized process visualization is described. For this purpose, a widget to display the currently shown tank and its level (*textfield*), a selection element for the three tanks (*choicegroup*) and a widget to confirm that choice (*command*) are modeled.

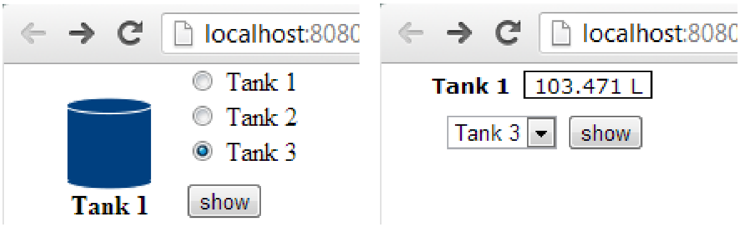


Fig. 7. Two versions of the final UI based on different results of the selection of unambiguous mappings (cf. Fig. 3)

MARIA is a description language for graphical, vocal and multimodal user interfaces, which covers the top three abstraction levels of the CRF [11]. For this case study, only the CUI level is used, in which MARIA, among others, defines the *MultimodalDesktopCUI* that is used hereinafter. Table 1 defines possible mappings of the *Widgets* that are used in the case study onto elements (*Class*es) of the MultimodalDesktopCUI of MARIA. These constitute the basis for the formulation of the *Mappings* of the *PAMTraM*.

Table 1. Excerpt of the possible mappings from DISL elements onto elements of the MARIA MultimodalDesktopCUI

		DISL – AUI		
		textfield	command	choicegroup
MARIA – MultimodalDesktopCUI	audio	x		
	button		x	
	check_box			x
	choice_element	x		
	drop_down_list			x
	image	x		
	link		x	
	list_box	x		x
	radio_button			x
	text	x		
	text_area	x		
	text_field	x		
	video	x		
	vocal_activator		x	
	vocal_output	x		
	vocal_selection			x

Table 1 shows (in excerpts) the occurrence of ambiguities, for example for the possibilities to define an output variable (*textfield*). For further steps, these (ambiguous) mappings were stored in an instance of the PAMTraM that was presented in Section 3.

Fig. 8 displays the resulting code for the definition of the mappings that has been created with the textual model editor (comp. Section 3.3). By means of the editor capabilities (e.g. the content assistant), a UI designer can define the mapping model without deep knowledge about the underlying meta-model of the PAMTraM.

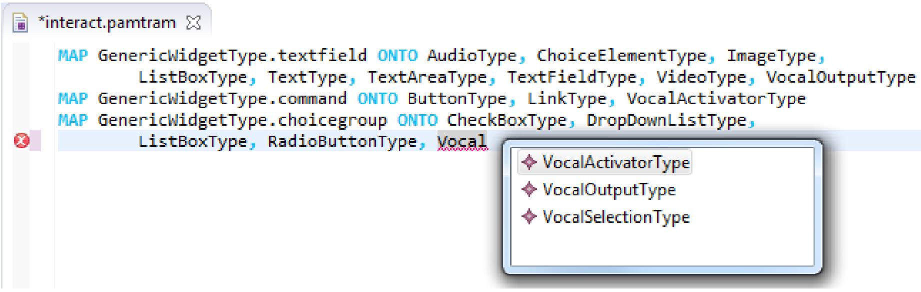


Fig. 8. Definition of the mappings with the textual editor

An interactive transformation was used in this case study to solve the occurring ambiguities. For each AUI element, it presents all possible target elements to the user in a graphical interface. The user can then choose the mapping to be used in the transformation by selecting one or more target elements. This transformation was implemented with help of the Epsilon transformation framework [20] and EMF.

Fig. 9 shows the prompt for user interaction during the initial transformation of the *textfield*-Widget that displays the fill level. The possible target elements correspond to those that were specified in Table 1 and defined in the ambiguous mapping model. It should be noted that the alternative(s) that was/were chosen by the user are stored as *TransformationMappings* as described in Section 3.2.

Fig. 10 shows the corresponding part of an instance of the PAMTraM (cf. Fig. 4(b)): The upper half displays a stored transformation consisting of several transformation mappings for the various UI elements. In the bottom half, one of the stored transformation mappings is presented in more detail showing that the source model element "fill_level" as instance of the meta-model element "textfield" has been mapped to an instance of the target meta-model element "image".

When re-transforming the UI model, e.g. in case of an iterative development, the user can select one of the stored, previously performed transformations. In this way, the corresponding unambiguous mappings are reused and a user interaction is only necessary for the transformation of elements that were added or changed during the last iteration step (comp. Sections 3.2 and 3.3).

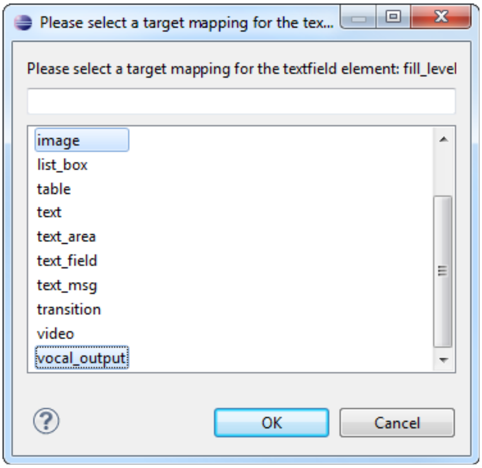


Fig. 9. User interaction for selecting the desired mapping

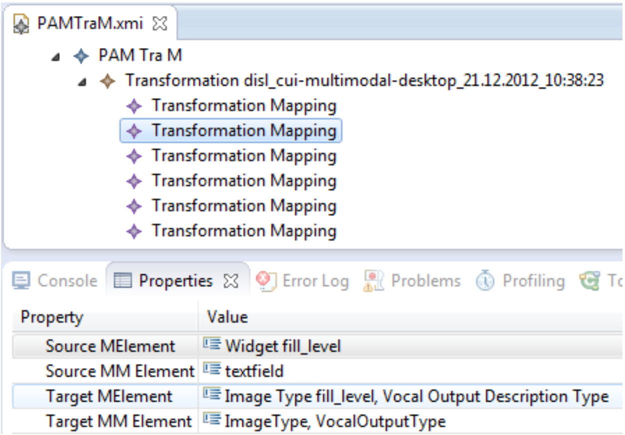


Fig. 10. Automatic storing of the selected mappings as TransformationMappings in the PAM-TraM

As a designer can make different choices when selecting the mappings, different UIs can be the result of the transformation. Two possible user interfaces for the process plant that were created using the PAMTraM approach have already been shown in Fig. 7. They are the result of a model-to-text transformation applied to the CUI model resulting from the previous steps. In the UI on the left, an image was chosen to represent the fill level of the tank, while textual output was chosen on the right. Furthermore, a group of radio buttons respectively a drop-down list was chosen to select the tank, whose fill level should be shown. These decisions could e.g. be made based on platform restrictions as the second UI is more compact and thus fits better onto smaller screens.

5 Conclusions

The proposed *Persistent Ambiguous Mapping and Transformation Model* meets the requirements that were developed in Section 1. It implements (a) the definition of ambiguous mappings and (b) the storing of executed transformations in one combined meta-model. By making use of a declarative structure, domain experts without any special knowledge about the execution of a transformation should be enabled to define ambiguous mappings. Indeed, this has yet to be proven by an empirical evaluation In order to hide the structure of the underlying meta-model from the domain expert, a textual editor has been created that supports the user in the definition of mappings. However, a graphical editor, like the one MARIAE offers, could be useful to handle the often very large models and to support the user's actions. Furthermore, it has to be elaborated if mappings between groups of elements (instead of single elements) need to be supported and how this support could be realized.

The structure of the defined meta-model allows the integration of different methods to resolve ambiguities: In the case study, we have demonstrated an interactive transformation. Furthermore, the utilization of the context of use or an integration of statistical algorithms (e.g. the evaluation of previous user choices based on stored transformations) is possible.

PAMTraM supports the iterative development of user interfaces by storing the chosen, unambiguous mappings. By reusing these stored mappings, the selection mechanism (e.g. a user interaction) only has to be applied to elements that have changed since the last iteration cycle.

Moreover, as PAMTraM uses universal structures, it is not limited to transformations between AUI and CUI (as demonstrated in the case study) but can be applied to all model(-to-model) transformations within the CRF. This extends to a transformation that makes only use of unambiguous mappings. Thus, a continuous methodology including a continuous tool support can be applied. However, as PAMTraM makes use of models respectively meta-models of the source and the target language, it cannot be applied to languages that are not represented by a formal meta-model but e.g. by a description in natural language.

The transformation execution that evaluates the ambiguous mappings defined in the PAMTraM and stores the chosen, unambiguous mappings has to be implemented for every language pair. However, the achieved separation between the domain knowledge stored in the PAMTraM and the actual transformation reduces the necessary effort for this implementation: The transformation can be implemented by an expert with in-depth knowledge of the creation of transformations who does not need to be a domain expert. Furthermore, the once implemented transformation can be used for every user interface that is described using the same UI description language pair.

Future work will cover the integration of the context of use into the decision process to reduce ambiguities. Therefore, possibilities to define conditions have to be created. Besides that, the user has to be able to attach these conditions to ambiguous mappings. By combining this with a subsequent user interaction, it can be ensured that the user may only choose between alternatives that are reasonable (for a certain context of use and a certain user interface). Compared to the exclusive user interaction realized in Section 4 the user is relieved of useless target elements.

Finally, an extensive empirical study has to be conducted in order to prove and classify the advantage of our approach. Therefore, the specification of transformations for various pairs of UI description languages as well as the development of specific user interfaces by means of the specified transformation has to be evaluated and compared to existing approaches concerning its effectiveness and usability.

Acknowledgments. The IGF proposal 16606 BG of the research association ”Gesellschaft zur Förderung angewandter Informatik e.V.” (GFaI) is funded via the AiF within the scope of the ”Program for the promotion of industrial cooperational research” (IGF) by the German Federal Ministry of Economics and Technology (BMWi) according to a resolution of the German Bundestag.

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, 289–308 (2003)
2. Vanderdonckt, J.: A MDA-compliant environment for developing user interfaces of information systems. In: Pastor, Ó., Falcão e Cunha, J. (eds.) CAiSE 2005. LNCS, vol. 3520, pp. 16–31. Springer, Heidelberg (2005)
3. Gruhn, V., Pieper, D., Röttgers, C.: MDA - Effektives Softwareengineering mit UML2 und Eclipse. Springer, Heidelberg (2006)
4. Puerta, A., Eisenstein, J.: Towards a general computational framework for model-based interface development systems. Knowledge-Based Systems 12(8), 433–442 (1999)
5. The Eclipse Foundation: ATL – a model transformation technology, http://www.eclipse.org/atl/
6. Object Management Group: Meta Object Facility (MOF) 2.0 Query/View/Transformation, V1.1 (2011), http://www.omg.org/spec/QVT/1.1/
7. W3C: XSL Transformations (XSLT), Version 1.0 (1999), http://www.w3.org/TR/xslt
8. Hager, H., Hennig, S., Seißler, M., Braune, A.: Modelltransformationen in nutzerzentrierten Entwurfsprozessen der Automation. i-com 10(3), 19–25 (2011)
9. Bézivin, J., Büttner, F., Gogolla, M., Jouault, F., Kurtev, I., Lindow, A.: Model transformations? Transformation models! In: Wang, J., Whittle, J., Harel, D., Reggio, G. (eds.) MoDELS 2006. LNCS, vol. 4199, pp. 440–453. Springer, Heidelberg (2006)
10. 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)
11. Paternò, F., Santoro, C., Spano, L.D.: MARIA: a universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM Transactions on Computer-Human Interaction 16(4), 1–30 (2009)

12. Lisai, M., Paternò, F., Santoro, C., Spano, L.D.: Supporting transformations across user interface descriptions at various abstraction levels. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 608–611. Springer, Heidelberg (2011)
13. Aquino, N., Vanderdonckt, J., Pastor, O.: Transformation templates: adding flexibility to model-driven engineering of user interfaces. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1195–1202 (2010)
14. Schäfer, R.: Model-based Development of Multimodal and Multi-device User Interfaces in Context-aware Environments. C-LAB Publication, Dissertation, Universität Paderborn (2007)
15. Hennig, S., Braune, A.: Sustainable Visualization Solutions in Industrial Automation with Movisa – a Case Study. In: 9th IEEE International Conference on Industrial Informatics, pp. 634–639 (2011)
16. Hennig, S., Van den Bergh, J., Luyten, K., Braune, A.: User Driven Evolution of User Interface Models – The FLEPR Approach. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 610–627. Springer, Heidelberg (2011)
17. Eclipse Modeling Framework Project, `http://www.eclipse.org/modeling/emf/`
18. Stahl, T., Völter, M., Bettin, J., Haase, A., Helsen, S., Czarnecki, K., von Stockfleth, B.: Model-Driven Software Development: Technology, Engineering, Management. John Wiley & Sons, Chichester (2006)
19. DevBoost: EMFText, `http://www.emftext.org`
20. Epsilon, `http://www.eclipse.org/epsilon/`

Model Assisted Creativity Sessions for the Design of Mixed Interactive Systems: A Protocol Analysis

Christophe Bortolaso[1] and Emmanuel Dubois[2]

[1] School of Computing, Queen's University, Kingston, Ontario, Canada K7L 3N6
[2] IRIT - ELIPSE, University of Toulouse, 31000 Toulouse, France
Christophe.Bortolaso@queensu.ca,
Emmanuel.Dubois@irit.fr

Abstract. To help designers face the complexity of mixed interaction and identifying original and adapted solutions, we developed and evaluated an original approach to interaction design. This approach, called Model Assisted Creativity Sessions (MACS), aims to combine the best elements of both a model of mixed interaction, and a collaborative and creative session. The objective is twofold: to support the exploration of the design space, and to establish a common language between participants. To assess the viability of this approach, we relied on a protocol analysis of the verbal recordings of two existing design situations. Results show that the model impacts the generation of ideas and that participants use the model concepts to share their thoughts during the session.

Keywords: Design, Method, Mixed Interaction, Model, Creativity.

1 Introduction

In the last ten years, we have seen an incredible amount of new interactive systems combining physical and digital entities. The emergence of augmented reality in games such as *EyePet*[1], the development of Kinect based interaction techniques such as *Omnitouch* [15], the commercialization of *Sifteos'* cubes[2], etc. constitute relevant instances of the huge potential offered by the combination of physical and digital worlds. These interaction forms are now commonly known under several terms such as mixed interaction, augmented reality, tangible user interfaces or even pervasive systems. To refer to all these advanced forms of interaction, we will use in this paper the general term of Mixed Interactive Systems (MIS).

With regards to traditional interactive systems (i.e. WIMP) designing MIS implies *many additional considerations*. For instance, physicality and topological properties of the involved artifacts, mediums enabling the communication between physical and digital worlds, spatial and semantic relationships between physical and digital artifacts or the quasi-infinite possibilities in terms of input/output technologies are dimensions that designers do not need to consider when designing traditional interfaces.

---

[1] EyePet™, http://www.eyepet.com/
[2] Sifteo, https://www.sifteo.com/

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 126–143, 2013.

Due to these new forms of inputs, these systems are growing in complexity, and the amount of possibilities in terms of interaction techniques is exploding. Indeed, during the development of a MIS, designers are required to manage not only the general complexity of HCI, but also all these specific considerations. The design space is therefore broader, richer and harder to explore. This implies many conceptual and *methodological problems* [26] and resulting designs often stay focused on archetypical solutions.

Based on this assessment, we wondered which support would help designers to better face, explore and take advantage of this complex design space. How do we help designers to look past these trivial and mostly technology driven design solutions? Providing such a support would also contribute to the diversification of MIS by helping designers identify original and adapted solutions.

Our contribution in this paper consists in revisiting the ideation step in the design process of MIS. Interaction design being intrinsically interdisciplinary, we focus in this paper on the collaborative and explorative part of the design process. In the following sections, we first provide the necessary background related to explorative design. We then elaborate on existing design approaches and resources in the domain of HCI and mixed interaction. We then introduce and illustrate a design method called MACS, which stands for Model Assisted Creativity Session. We continue by showing the viability of this approach through the analysis of the use of MACS in two real design situations. We finally conclude on the benefits of this approach and the perspectives it offers in terms of design and research.

2 Toward a New Form of Creative Design

First, as already pointed out by Visser [30], the design activity can be considered as a *construction of representations*. Through this paradigm, design has to be viewed as an iterative task involving the externalization of ideas, the creation of representative artifacts, and the reinterpretation of the produced representations. Through this process designers infer properties of the design that were not perceptible before. It ultimately leads to new insights that contribute to the discovery of a satisficing solution [27]. This process, widely described by Schön [25] as *reflection-in-action*, constitutes the basics of any design activity.

Furthermore, it is now commonly accepted that the design process can also be understood as the exploration of a space whose limits are blurry and ill-defined [27]. As defined by Gero [12], this exploration takes place in a space where variables (i.e. properties of the designed artifact) are adjusted. This space can be subdivided into three different sub-spaces [12]:

- The *routine design space* defined as a space of potential designs, in which all the variables and their applicable ranges of values are directly extracted from existing solutions.
- The *innovative design space* defined as a space of potential designs in which all the instantiated variables are known but which values are new or unfamiliar.
- The *creative design space* defined as a space that uses new variables. This generally leads to a paradigm switch.

Accessing the Innovative and Creative space requires the knowledge of existing variables (see Fig. 1). For example, when designing a MIS, considering the spatial and semantic relationships between physical and digital artifacts, necessarily leads to more intelligible designs.

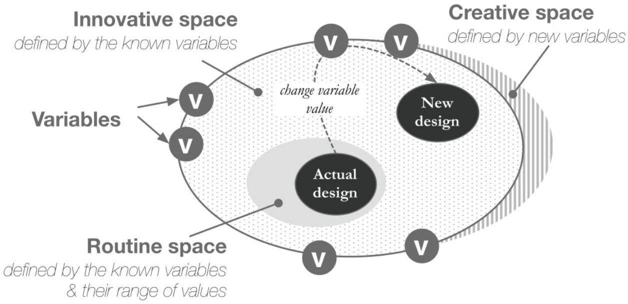


Fig. 1. Access to the Innovative space through the manipulation of variables

In the field of HCI, a set of widely accepted representations is used as a support to the instantiation of variables related to interaction. Scenarios, sketches, storyboards, mock-ups, etc. have proven to be efficient and flexible representations. In addition, these informal representations have the huge advantage of being understandable by most of the stakeholders in the design process such as end-users. Therefore these representations play the role of boundary objects that allows "*people from different areas of expertise to bridge their separate knowledge domains, to create a shared understanding, and improve decision making*" [7].

However, their contribution to the exploration of the design space is limited to the manipulation of the variables that the designer is already familiar with. As a result, the expressive and generative power of these representations highly depends on the knowledge and level of skillfulness of the designer [31] itself: the access to the Innovative and Creative spaces are then highly influenced by the designer's experience. The solution to this limitation could rely on the use of a specific language. Indeed, as stated by Löwgren and Stolterman [22], in interaction design "*a language is necessary in making ideas and thoughts more precise and well-crafted*".

While many design methods focus on the use of traditional artifacts, we chose to explore how abstract concepts, theories and models can assist the design activity. Our research can therefore be seen as an instance of modern design theories such as *C-K design theory* [16] (i.e. in which the design activity is understood as an interplay between existing Concepts and domain Knowledge) or even as the stream recently highlighted by Stolterman and Wiberg in [28] and so-called *Concept-Driven Interaction Design*. These attempts result in a set of new methods, in which design takes inspiration and restriction from existing paradigms, theories and languages. To our knowledge, only a few HCI methodologies have tried to implement this approach. We synthesize a set of existing methods adopting this approach in the next sections.

3 Concept-Driven Interaction Design in Practice

In the field of HCI, the literature mentions a set of design methods that can be considered as Concept-Driven approaches. They mostly differ in the level of abstraction of involved representations.

First, some of them are focused on the use of concrete representations. One of the first methods implementing this paradigm is the PICTIVE method [24]. Proposed by Muller in 1991, it consists in asking end-users to create the interface of their dreams through the manipulation of a paper-based and predefined set of buttons, windows, lists, icons, etc. The proposed set of widgets is limited to the existing ones (e.g. buttons, lists, text fields, etc.), which therefore constrains the exploration in the innovative design space. However, PICTIVE remains dedicated to the design of WIMP interfaces. More recently, Koleva proposed a similar approach [21] but in order to explore the possibilities in terms of augmented reality for museum visits. In this experience, curators and engineers collaboratively manipulate a diagram representing component-based software assembly. The manipulated software components are instantiated on the fly from an existing list of available components. This list therefore represents a set of existing possibilities in terms of augmented reality solutions, but the creation of links between the different components remains the responsibility of the curators and engineers. However, approaches like PICTIVE or Koleva's, are mainly technologically driven, and the exploration consequently takes place in a space that is too concrete and narrowed.

To overcome such limitations, a set of approaches suggests a concept-based exploration at a higher level of abstraction. For example, the Rich Picture Method [23] suggests that a facilitator and several participants collaborate to draw a picture depicting stakeholders, relationships and their concerns. The Rich Picture notation provides all the concepts (i.e. the variables of the design space) that need to be expressed and link them in a diagrammatic form. Here the design space is not explored through a set concrete interface elements, but at a higher conceptual level representing the concepts of the application domain. Finally, in the domain of Tangible User Interfaces (TUIs), Hornecker proposed an original method called the Card Brainstorming Game [17]. This game mainly consists of a brainstorming session in which participants are stimulated by provocative cards. Each card is related to a dimension expressed in the Tangible Interaction Framework [18], a conceptual paradigm of TUIs. As observed by the author, "*it thus ensures that all the central tenets have been taken into account*" at the end of the session. However, this last approach remains poorly generic and does not provide any systematic strategy to explore the design space.

From our point of view, these Concept-Driven Interaction Design instances converge on two aspects: 1) the manipulation of a defined and limited number of concepts and 2) the collaborative elaboration of a dedicated representation. This sort of abstract based exploration constitutes a very promising approach to interaction design and needs to be further investigated, evaluated and used in real design situations.

4 Formalism and Creative Design: A Promising Combination

The challenge of our research is to gain a better understanding of the benefits of exploring the design space through the use of a model of MIS. During the last five years, we tested and refined an original concept-driven method dedicated to the design of MIS. The approach we developed is called *Model Assisted Creativity Session* (MACS). The goal of a MACS is to identify and describe a set of alternative solutions to a design problem. Such creative sessions usually involve between 5 and 7 participants including a facilitator. The core principle of MACS is to collaboratively take advantage of the concepts expressed and characterized in an existing model of mixed interaction. The participants thus generate ideas and encode them on the fly in the model's notation. For now, this method has been successfully used [6] with two different interaction models: ASUR [11] and MIM [9]. However, the literature mentioned a lot of other models that are candidates for use in a MACS: each one is describing different aspects of MIS and adopting a different paradigm, using different concepts and represented through various notations [18, 26, 29].

Beyond its role of providing a support to the description of generated solutions, the model is also intended to play a major role in the creative and collaborative process:

1. *Support the exploration of the design space through the manipulation of a limited number of concepts:* it enables participants to primarily focus on relevant aspects of the system to design. Limiting the participants to the exploration and recombination of these relevant aspects, thus stimulates the creativity of the participants by arousing unexpected combinations. Indeed, it is now common knowledge that a design activity always takes place in a constrained cognitive space [4] and that creativity highly benefits from these constraints [8].
2. *Support the co-construction of an external representation:* the model allows us to represent the generated solutions and creates a common language between disciplines. It thus contributes to the elaboration of a shared understanding of the problem. This phenomenon, known as cognitive synchronization, has proven to be one of the main criteria to the success of collaborative design activities [10].

These two objectives will constitute the basis of our further analysis and evaluation that we present in the next sections.

4.1 The MACS Method Definition

To reach the objectives that we defined in the previous section, the MACS method is composed of a set of steps and principles that the group should follow. More details about the core principle of this method are provided in [6] and examples of its use in different context have been illustrated in [5, 14]. However, to ease the understanding of this paper, we briefly describe the five main steps of the approach below:

Step 1. The facilitator introduces the model and its notation to all the participants. This introduction should not take longer than 15 minutes. The presentation of the concepts that will be manipulated and their associated representations contributes to the cognitive synchronization between participants of different disciplines and enables the use of the model. The facilitator needs therefore to be an expert of the model. At the end of this step, we recommend to provide the participants with a concept reminder such as a legend of the model.

Step 2. The facilitator introduces the design situation and the problem to solve. As for step 1, this step should take no longer than 15 minutes. Any design artifact that describes a part of the design problem can be used at this point (e.g. mock-ups, videos, sketches, requirements list, etc.).

Step 3. This is the main step of the method in which participants will explore the design space. In practice, participants generate ideas to answer the interaction design problem and encode them on-the-fly into the model notation. The ideas take the form of model-concepts and/or links between them. The model notation also enables the group to keep track of the generated ideas and to iterate around the produced solutions. The facilitator must manage the group dynamics and can also help participants with the model's syntax.

Step 4. This step is conducted by the facilitator after the session and aims at extracting the best solutions from the idea generation. To do so, the facilitator characterizes the produced models with regards to the initial requirements composing the design problem. The facilitator therefore has to select solutions that best fit the requirements. He can also alternatively combine (part of) different solutions, each of them partially covering the requirements, in order to obtain complete solutions satisfying the requirements. At the end of this step, the best solutions are candidate to prototyping and reintroduced into the design process.

Step 5. As for step 4, the facilitator also conducts this step after the session. It is aimed at identifying the characteristics of the generated solution that need further refinement. This can happen if some of the model dimensions have not been discussed in step 3 or if the model does not express certain characteristics. In the first case it may lead to a new MACS using the same model or a different one more appropriate for the identified problem. In the second case it will require the use of another model or dedicated representation. In both cases, these issues open the design to further iterations.

This set of five steps constitutes the basis to conduct a session. To complete the definition of the method, the following section illustrates how the manipulation of the model performed during step 3 can contribute to the exploration of the design space.

4.2 How Can a Model Help to Generate Interactive Solutions?

In this section, we show how to generate systematically alternatives by using an interaction model. To illustrate this method, we chose a famous TUI: the Reactable [20]. This system allows users to compose electronic music by manipulating little bricks on an interactive tabletop. Each brick represents either a sample of music or sound effects that can be applied to it.

To demonstrate the power this approach, we will rely on a model of MIS called ASUR [11], which stands for Adapter, System, User and Real objects. ASUR defines a mixed interactive situation as a set of entities linked by information flows. More details about the ASUR model are available in [11].

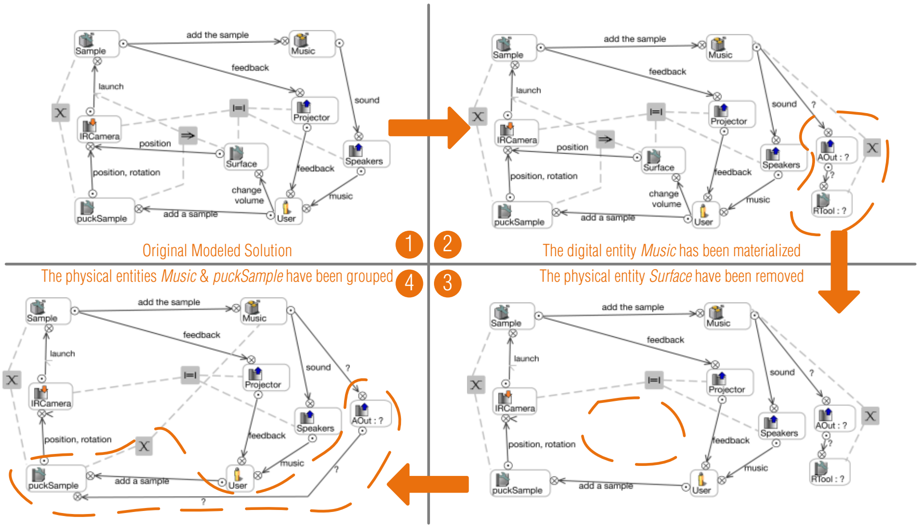


Fig. 2. 1) ASUR model of the task "adding a sample" with the Reactable; 2) Model 1 after materializing the digital entity `Music`, 3) Model 2 after removing the physical entity `Surface`, 4) Model 3 after grouping of `Music` and `puckSample` physical artifact

For example, Fig. 2 n°1 (top-left corner model) shows the ASUR description of the Reactable system. This model describes the task "Adding a sample on the table". It shows a `User` handling a physical entity called `puckSample` whose form and position are recognized by an adapter `IRCamera`, which is physically associated with a physical entity `Surface`. The adapter `IRCamera` notifies a digital concept called `Sample`, which acts on the main digital concept `Music`. Both `Music` and `Sample` digital concepts are returned to the `User` through the adapters `Projector` and `Speakers`, which respectively enable the visual feedbacks on the surface and the diffusion of the sound.

To generate alternative designs to the Reactable, we propose to modify consecutively three times the original model (Fig. 2 Model n°1) describing the existing Reactable system.

To begin, we propose to materialize the digital entity `Music` shown on model n°1 in Fig. 2. This results in creating a physical object, whose specific characteristics will change in function of the variation of the sound. The model n°2 of Fig. 2 shows the changes it impacts on the model. This solution leads to a new system in which the user sees/touches/feels physical variations in the music.

On this new basis, we suggest to modify this second model by removing an entity. To highly impact the system, we removed one of the most prominent objects in the interaction: the `Surface`. This leads to the model n°3 shown on Fig. 2. Here, the user can no longer put the puck on the surface and has to manipulate it directly. As a result, several possibilities are now opened. For example, this puck could take the form of a cube whose position and orientation in the space act on the `Music`.

Finally, we can go even further by performing a third modification and grouping several entities together. Based on model n°3, we can group the physical `puckSample` and the materialized `Music`, the two objects we have created previously. It results in a fourth solution (see model n°4 on Fig. 2) in which the user manipulates an object whose characteristics vary in function of the sound. For example, it could take form of a malleable object whose form varies with the frequency of the `Music` and which user can move, shake, and orient to act on the system. This is obviously an imaginary solution, but tomorrow's materials could easily drive us to such interactive systems.

To conclude, we have shown that by applying basic modifications on the original modeled solution, it is possible to explore the design space and discover unexpected solutions. The techniques we used to modify the model (i.e. materialization, removing and grouping) are applicable infinitely, in any order and could be completed by symmetrical techniques (i.e. dematerialization, adding, splitting).

This list of technique is definitely non-exhaustive, but provides a good overview of how simple manipulations of a model can help in the finding of alternatives. Further work needs to be conducted to identify a larger set of generative techniques and to study their completeness.

In addition, these techniques should not be seen as an automatic way to generate new solutions. Indeed, each time an entity is impacted by the use of a technique, we replaced it by a new entity with an undefined semantic (cf. each "?" on the entities of models 2, 3, 4 on Fig. 2). This means that the use of such manipulation is intended to force designers to reinterpret the situation and to reconsider the initial problem. Such manipulations do not generate ready-to-use solutions. These techniques therefore need to be combined with traditional creative techniques such as divergent or analogous thinking. When used in combination with the ASUR model, the MACS method supports the combination of creativity and modeling, thus taking advantage of both.

5 Analysis

We showed that the use of a model during collaborative and creative session could constitute a promising and alternative way of designing MIS. As underlined in the introduction, the aim of such an approach is twofold.

First, the model is supposed to assist the exploration of MIS design space through the manipulation of a limited number of concepts. While traditional representations such as scenarios, sketches or mock-ups are known to impact collaborative and creative thinking, this remains uncertain with representations such as an interaction models. Therefore, we came up with the following hypothesis:

Hypothesis 1: In a MACS, the model impacts the idea generation, i.e. the exploration of the design space. To assess if the model has an impact on the generation of idea, we first need to verify if the model is used by the participants to support the creative process. To do so, our following analysis will quantify the links between the generation of ideas during the session and the use of model concepts during a session.

Secondly, the model is aimed at providing a language to the group that collaboratively manipulates the concepts. Computer science engineers, designers, researchers and end-users will use the words and the representations associated to the introduced concepts to express their ideas during the sessions. Therefore, the model should establish a common ground between participants from different backgrounds and disciplines. Consequently, we hypothesized about this aspect as follows:

Hypothesis 2: In a MACS, the model acts as a language of reference, i.e. support the collaboration during the session. To assess this second hypothesis, we need to verify if the model's concepts are actually used to express and discuss ideas during the sessions. To do so, our following analysis will focus on the use of model concepts during the session.

To verify these hypotheses, we performed a verbal protocol analysis of two case studies that we detail hereafter.

5.1 Case Studies

Over the last five years we used the MACS method eight times with different case studies, application domains, design teams and type of problems. Among these several applications, we selected two occurrences hereafter referred to as $MACS_a$ and $MACS_b$ to be candidates for a detailed analysis. These two instances were concerned with a project focusing on the development of MIS dedicated to cultural centers and involved highly motivated interdisciplinary partners (computer scientists, ergonomists, curators, dancers, etc.). These authentic design situations raised design questions related to existing and concrete problems. Participants of these two instances of MACS were thus intrinsically motivated by a clear and unique objective, which constitutes a major aspect in the creative process [2]. Prior to these two sessions, participants (except the facilitator) had no knowledge about the ASUR model.

MACSa: An Interactive Exhibit Dedicated to the Understanding of Cladistics. In this session, participants were asked to identify input interactive techniques to manipulate a 3D tree representing a classification of species. The application, called CladiBubble, aims at explaining why some well-known groups of species, such as fishes or reptiles, are no longer valid in terms of cladistics. Cladistics is a scientific method that classifies species regarding to their phylogenetic criteria and is the main focus of the Museum of Toulouse's exhibits. With CladiBubble, Visitors are invited to manipulate a 3D cladogram in order to bring species spatially close to each other. They then bring them into a bubble to create a group. Finally, the application validates the group in terms of cladistics and provides cues to constitute a group that is correct.

At the beginning of the MACS, during step 2, we used a video prototype to demonstrate the application. The design problem was to find mixed interaction techniques to allow the manipulation of the cladogram and grouping of species with the bubble.

The design space was therefore opened to many input interaction techniques and ways of rendering the application in the museum. The session included five participants: a facilitator, a technical engineer, a HCI specialist, an ergonomist and a MIS specialist. The models were edited on a computer assisted graphical editor and projected on a wall that was visible by all the participants (see Fig. 3).

Fig. 3. The $MACS_a$ session. A participant is showing an element on the model, which is projected on the wall.

Fig. 4. The $MACS_b$ session. Participants are collaboratively editing a model with the paper based support.

MACSb: An Augmented Ballet to Perceive Dancer Emotions. This session was related to the design of an augmented ballet. The main objective of this project was to emphasize the dancers' emotions and to express it to the audience. To do so, a dancer's sequences of moves were sensed through a motion capture suit. Each sequence of movement was then linked to a predefined emotion or set of emotions. The challenge of this MACS was to identify alternative ways of making the dancer's emotions perceptible to the audience. As a result, the design space was mainly open to output interaction techniques and representations of emotions.

The session included seven participants: a facilitator, an emotion specialist, a MIS specialist, a 3D specialist, a usability expert, a dancer and a final user (i.e. a person used to watching ballet dancing). Here, the models were edited with a tangible and collaborative editor (see Fig. 4) composed of large sheets of paper, pre-printed and colored post-it notes representing ASUR entities, and colored pens.

5.2 Protocol Analysis

To verify our two hypotheses and assess the viability of the MACS methods, we relied on the protocol analysis technique [13], a largely accepted methodology in the field of design research. This technique consists of coding a set of recordings of a defined task by following a specific scheme. In our case, we collected audio/video recordings of the two sessions and analyze them by using a specific scheme dedicated to the analysis of collaborative design sessions.

Coding Scheme. The coding scheme we used, is based on an existing method named COMET [10] which has been designed to analyze activities of designers during meetings. Here, we relied on the part of the method that is specifically dedicated to the analysis of dialogues. To do so, COMET sequentially relies on 1) the textual transcription of the dialogues occurring in a design session, 2) the segmentation of each turn-taking into one or several segments conform to the COMET predicate format (`MOD(ACT/ARG)`) and 3) the coding of each segment.

Table 1. Our coding scheme. Gray lines represent the original solution code in the COMET [10] scheme. We divided it into 5 different codes in order to differentiate the different types of solutions that are relative to the model's concepts, or out of the model scope.

CLASS	CODE	DEFINITION
MOD	AS	Assertion
	RE	Request
ACT	GE	Generate: Idea generation
	EV	Evaluation: Judgment, criticizing
	IT	Interpretation: Rephrasing, reformulation
	IF	Information: Completing, adding, specifying
ARG	MS	Model Metaphor: General idea, a whole model
	ME	Model Entity: Physical/digital/Adapter/User artifact
	MA	Model Attribute: Characteristic of an entity
	MC	Model Channel: Communication between entities
	OT	Other Solution: Idea out of the model scope
	PD	Problem Data: Part of the design problem
	DO	Domain Objects: Concept of the application domain
	DP	Domain Procedures: Constraint of the application domain
	GO	Goals: main objective of the application
	TA	Tasks: atomic task in the use of the application

This scheme is divided into three classes (see Table 1). The first class, called Modulation, is relative to the type of intervention (`MOD=assertion | request`). The second class, called Action, is relative to the type of action (`ACT=idea generation | evaluation | information | interpretation`). The third class, called Argument, is relative to the subject that is currently referenced (`ARG=solution | problem data | domain object | domain procedure | goal | task`). As a result, each segment of the protocol will be coded as being a Modulation, an Action and an Argument.

To further adapt the use of COMET to our context, we added a level of precision to the solution code in the argument class (see gray lines on Table 1). Initially in COMET, the solution code should be attributed to segments in which a participant elicits a solution answering the design problem. With our adaptation, solutions can be 1) relative to the model's concepts, or 2) out of the model scope. In order to analyze and understand the use of the model's concepts during the session, four possible values, representative of the concepts expressed in the model used, have been considered for the solution codes: metaphor, entity, attribute and channel. Table 1 above summarizes all the codes composing our final coding scheme.

Corpus Features & Protocol Elaboration Method. First, as shown on Table 2, both sessions combined, our analysis has been conducted on more than 2700 segments, which represent more than 26000 words.

Table 2. Corpus features and Kappa agreement coefficients between the coders and protocols

$MACS_a$			$MACS_b$		
Duration = 1:40:46 nbSeg = 974 segments (9896 words)			Duration = 2:02:32 nbSeg = 1815 segments (16861 words)		
C1&C2 K=0.59	C1&AP K=0.61	C2&AP K=0.71	C1&C2 K=0.69	C1&AP K=0.70	C2&AP K=0.77

C1&C2: Agreement between coder 1 and coder 2.
C1&AP: Agreement between coder 1 and arbitrated protocol.
C2&AP: Agreement between coder 2 and arbitrated protocol.

However, even with such a large sample, the protocol analysis method is known to be influenced by the interpretation of the individual that performs the coding. To reduce this interpretation effect and tend to a form of objectivity, we relied on a specific coding process, as described in [13]. Therefore, the protocols of both case studies have been coded by two different individuals and then arbitrated within more than ten days after the coding. We then computed a level of agreement between the different protocols based on Cohen' Kappa coefficient. Table 2 shows the Kappa coefficient we obtained. The coefficients between 0.61 and 0.8 indicate a substantial and good agreement between the coders. Given our context, the data can be considered as reliable. All the following analysis has been realized on the arbitrated protocol (AP).

6 Results

Both sessions have been conducted successfully and produced modeled solutions that answer the design problem.

$MACS_a$ has produced 17 models with various forms of MIS solutions. For example, participants have envisioned the use of a tangible pump to enlarge the bubble group in the game. Differently, another interesting solution was related to the use of several tangible bricks representing species that could be associated with a multi-function tangible selector. The display of the application has also been revisited through the use of an augmented shadow that allows visitors to point and select elements on the screen. However, due to the amount of ideas, we do not report all the modeled solutions here.

$MACS_b$ has produced 5 different models. During this session, the team grabbed the opportunity to reconsider several aspects of the project, which was ongoing for three years before this session. As a result, unanticipated solutions emerged. For example, a proposition was to relay all the data points, captured by a motion suits sensors, to the audience, rather than combining this information into a single representation. Participants therefore explored different representations and modalities such as numbers, colors, sounds, curves, etc. As a result, the emotion could be interpreted by the spectator himself. Another original solution suggested using the dancer's movements could be used to sculpt a virtual 3D object. The resulting 3D object would be returned to the audience through different projectors or augmented reality glasses.

In the following paragraphs, we present the data we collected from the protocol analysis. Even if the differences between these two case studies make them not rigorously comparable, this double analysis aims at showing the general trends we observed in the use of the MACS method with the ASUR model.

6.1 Overall Observations on the Code Distribution

Before we focus on our hypotheses, we made a set of observations regarding the code distribution of our protocol analysis. First of all, as shown in Fig. 5, protocol analysis reveals that the code distribution is very similar over the two sessions. This encourages us to think that the method is stabilized and has predictable behavior, whatever the team, design problem or the support used to edit the model (PC based or paper based).

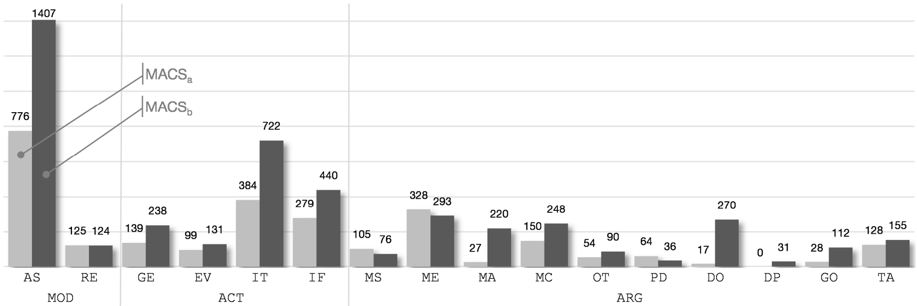


Fig. 5. Code distribution for each class (modulation, action, argument) and for both sessions

However, we observed some meaningful differences between the two sessions. For example, the $MACS_b$ (ballet) shows a bigger interest in terms of model attributes (220 MA segments) and channel (248 MC segments) while $MACS_a$ (museum) remains mainly focused on the reference to model entities (328 ME segments). This difference is due to the design problem itself. Indeed, as elicited in the case study presentation, the question relative to ballet augmentation ($MACS_b$) was more prone to the exploration of emotions rendering, which implies the use of attributes and channels in the model. In contrast, the interaction technique problem of CladiBubble ($MACS_a$) was more focused on the identification of artifacts to be manipulated by the user.

Another interesting difference is relative to the reference to domain objects (DO). $MACS_b$ was highly more focused on objects relative to the ballet and to the dancing while participants of the $MACS_a$ made almost no reference to the domain objects (concepts relative to cladistics and museums). The level of definition of the design problem can explain this difference. Indeed, $MACS_a$ occurred while the CladiBubble prototype was in a real concrete state, while when $MACS_b$ occurred, the ballet augmentation project was essentially starting from scratch.

These two observations tend to show that 1) participants are able to use the dimensions of the model that are the most relevant to a specific design problem and 2) the MACS method is adapted for design problems with different levels of definition.

6.2 Hypothesis Assessment

Hypothesis 1: The Model Impacts the Idea Generation. To assess this first hypothesis, we focused on the solutions to the problem generated by the participants and on the actions of generation. As a result, we first looked at all the segments which argument was coded as a solution to the design problem (ARG=MS|ME|MA|MC|OT) and then at the links between action of generation (ACT=GE) and the references to a model concept (ARG=MS|ME|MA|MC).

On this basis, we computed the ratio R1 (1) of solution relating to a model concept on the total amount of argument related to a solution.

$$R1= \frac{\sum_{i=0}^{nbSeg}\{MS,ME,MA,MC\}}{\sum_{i=0}^{nbSeg}\{MS,ME,MA,MC,OT\}} \quad (1)$$

R1=92% for of $MACS_a$ and R1=90% for $MACS_b$. This means that when participants are talking about solutions to the design problem, the very large majority of their verbal intervention is referring to a model concept.

To go further in the dynamic of the session, we also analyzed the sequencings between actions of generation and argumentation about the model concepts. We therefore counted the number of times a participant referencing a model concept (ARG=MS|ME|MA|MC) was followed by a participant generating an idea (ACT=GE). Table 3 illustrates the kind of situation we considered. This analysis deeply relies on the situated point of view of design. Indeed, design can be seen as an activity that is sequentially and contextually anchored [10]. Through this point of view, a MACS is a constant exchange between participants themselves and between participants and the model. Each turn taking can therefore influence the next one.

Table 3. Example of considered sequences. In this sample, extracted from the $MACS_a$, the Facilitator's segment 232 triggers the idea generation action in the Technician's segment 233.

Seg.	Time	Participant	Transcription	MOD	ACT	ARG	Solution
231	0:58:05	Ergonomist	This is more precise than a camera!	AS	IT	ME	
232	0:58:09	Facilitator	But what's the interest in using accelerometers and gyros?	RE	IT	ME	
233	0:58:25	Technician	This could be used for gestures when you move the nodes?	AS	GE	MC	Model 3: Gestures to move nodes
234	0:58:30	Facilitator	Oh yes! We could use gestures!	AS	IT	MC	

We observed that 64% of actions of generation were preceded by a reference to a model concept for the $MACS_a$ and 62% for $MACS_b$. In other words, the majority of the idea generations occurred directly after a participant would mention an idea in the scope of the model's concepts. As a result, we conclude that the majority of the generative actions are stimulated by a model concept.

Given that the large majority of the topics discussed during both sessions are related to the model framework and that the model concepts stimulate the idea generation process, we validate our first hypothesis.

Hypothesis 2: The Model Acts as a Language of Reference. To assess this second hypothesis, we focused our analysis on amount of references to the model concepts without considering the type of action (`ACT=GE|EV|IT|IF`) and on salient behaviors in the video recording of the sessions.

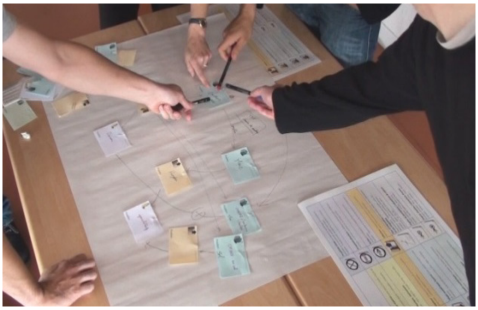


Fig. 6. An example of the collaborative use of the model

First of all, we observed from the protocol analysis of both sessions that all the participants make reference to the model. In addition, we observed in the MACSb (in which a paper-based collaborative editor was used) that all the participants have contributed to the edition of the models. Finally, the model is also regularly used as a support to the referencing of already generated elements. For example, as shown on Fig. 6, participants point directly at an element on the model to support their assertions.

We then computed the ratio R2 (2) of references to the model concepts on the total amount of segments during the two sessions.

$$R2= \frac{\sum_{i=0}^{nbSeg}\{MS,ME,MA,MC\}}{\sum_{i=0}^{nbSeg}\{MS,ME,MA,MC,OT,PD,DO,DP,GO,TA\}} \quad (2)$$

R2=68% for of $MACS_a$ and R2=55% for $MACS_b$. This means that for both sessions the majority of segments were related to a model concept (`ARG=MS|ME|MA|MC`). This also means that the discussion is mainly driven by the use of the model to generate, evaluate, interpret and inform ideas.

We also considered the temporal aspect of the sessions. We computed the average occurrence frequency of a reference to a model concept. It appears that over the total duration of the session, a reference to the model occurs each 9.11 seconds (*SD*=16.78) for the $MACS_a$ and each 8.75 seconds (*SD*=26.38) for the $MACS_b$. Even if the standard deviation is high in regard to the time, we can conclude that references to the models concepts occur very frequently during the sessions.

These set of qualitative observations and computations lead us to conclude that the model effectively plays the role of the language of reference during the sessions, which consequently validates our second hypothesis.

7 Discussion and Perspectives

In this paper, we first summarized the principles of a design method combining the best elements of both a model of mixed interaction, and a collaborative and creative session. We then studied the viability of this approach in order to understand if a design team can benefit from a formal representation to collaboratively explore the design space of mixed interaction. Regarding to previous evaluations of this method which were mainly based on qualitative results and informal observations [6], we demonstrated here the viability of the MACS method through a quantitative analysis based on verbal protocols. The results presented here show that in a MACS, the model impacts the generation of ideas (H1) and acts as a language of reference (H2).

Participants also perceived these two advantages. Indeed, during the sessions' debriefings, participants reported some fruitful comments. For example, at the end of the $MACS_b$, the ballet dancer argued that during the session "*the model really allows to turn around an idea to find variation or combinations*". A few minutes later he added: "*…suddenly, I feel I can speak at levels in which I usually do not say anything, because I know nothing about it!*". During the same debriefing, the emotion specialist pointed out that "*the model has provided a common basis for support the communication between computer scientists and others*". These interventions clearly indicate that participants felt the help provided by the model to explore the design space and to support the communication between people from different backgrounds.

However, to strengthen the value of our approach, it would be interesting to quantify to which extent the introduction of the model impacts the idea generation. To do so, it is required to compare the MACS with other creative methods such as a brainstorming. This could allow us to quantify the gain of the MACS in terms of design space exploration or the impact of the model manipulations on participants' cognitive load for example. Seeing the results we obtained to show the viability of our approach in this paper, we think that the MACS method is now mature enough to perform such a comparison. Techniques like linkography constitute relevant analytic tools to perform such a comparison.

In addition, the analysis we presented here could be extended to a behavioral analysis. In fact, the analysis of participants' physical and perceptual actions could 1) contribute to a better understanding of the model's impact during these sessions and 2) provide clues to the development of interactive tool dedicated to creative groups.

It is also important to underline that we learned from our experience the MACS approach is compatible with traditional HCI development process methods. In fact, during the development of the two case-studies presented here, we successfully used the MACS approach in parallel to other traditional participatory design techniques. In fact, the solutions modeled during the session can be used as a starting point for subsequent design steps, in particular when it comes to prototyping and implementation.

Indeed, model transformations are useful tools to infer (parts of) the running prototype from a source model such as those used in a MACS.

In addition, a MACS is supposed to work with any model. As a result, we could envision the use of other frameworks such as RBI [19] or TAC [29]. However, several reasons drove our choice to the use of ASUR. First, ASUR model is composed of a small number of concepts. Participants can thus roughly represent an interaction technique by manipulating less than seven types of entities, which make the model easy to handle and to learn. Secondly, ASUR is supported by a graphical representation. Several participants reported at the end of sessions, that they used the diagram's form to remember and to come back on some generated ideas. It therefore appears that the model's number of concepts and notation play an important role in the success of these sessions. An analysis based on the cognitive dimensions of notation [3] could be helpful to drive the choice of a model. For example, the viscosity and the level of abstraction of the notation appear to be relevant dimensions to focus on.

To conclude, we strongly believe that concept driven interaction design approaches are promising approaches in the field of HCI and more specifically for the design of advanced interactive systems. Indeed, these types of approaches have proven to be efficient in more established disciplines. For example, in the domain of physics, engineers can rely on a 40 dimensions matrix called TRIZ [1], to explore the design space and find original solutions to problems. This approach has proven to be efficient and is now used by many practitioners in the domain of physics all over the world. The research we conducted here is thus a step into the development of such a systematic approach dedicated to mixed interaction design.

References

1. Altshuller, G.: Innovation Algorithm: TRIZ, systematic innovation and technical creativity. Technical Innovation Ctr. Retrieved from, p. 312 (1999)
2. Amabile, T.M.: Creativity in context. Boulder, p. xviii, 317. Westview Press (1996)
3. Blackwell, A., Green, T.R.G.: Notational Systems - the Cognitive Dimensions of Notations framework. In: John, M.C. (ed.) HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science (2002)
4. Bonnardel, N.: Towards understanding and supporting creativity in design: analogies in a constrained cognitive environment. Knowledge-Based Systems 13(7-8), 505–513 (2000)
5. Bortolaso, C., Bach, C., Dubois, E.: MACS: combination of a formal mixed interaction model with an informal creative session. In: EICS 2011, Pisa, Italy, pp. 63–72. ACM, New York (2011a)
6. Bortolaso, C., Bach, C., Dubois, E.: Co-Design of Interactive Museographic Exhibits: the MIME case study. In: ReThinking Technology in Museums, pp. 37–48. University of Limerick, Limerick (2011b)
7. Bresciani, S., Blackwell, A., Eppler, M.: A Collaborative Dimensions Framework: Understanding the Mediating Role of Conceptual Visualizations in Collaborative Knowledge Work. In: Hawaii International Conference on System Sciences, p. 364. IEEE Computer Society, Los Alamitos (2008)
8. Costello, F.J., Keane, M.T.: Efficient Creativity: Constraint-Guided Conceptual Combination. Cognitive Science 24, 299–349 (2000)
9. Coutrix, C., Nigay, L.: Balancing physical and digital properties in mixed objects. In: AVI 2008, Napoli, Italy, pp. 305–308. ACM, NY (2008)

10. Darses, F., Détienne, F., Falzon, P., Visser, W.: A Method for Analysing Collective Design Processes. In: ECCE 2010, Linköping, Sweden (2010)
11. Dubois, E., Gray, P.: A Design-Oriented Information-Flow Refinement of the ASUR Interaction Model. In: Gulliksen, J., Harning, M.B., van der Veer, G.C., Wesson, J. (eds.) EIS 2007. LNCS, vol. 4940, pp. 465–482. Springer, Heidelberg (2008)
12. Gero, J.S.: Design prototypes: a knowledge representation schema for design. AI Magazine 11(4), 26 (1990)
13. Gero, J.S., Mc Neill, T.: An approach to the analysis of design protocols. Design Studies 19(1), 21–61 (1998)
14. Graham, N., Dubois, E., Bortolaso, C., Wolfe, C.: Scenarchitectures: The Use of Domain-Specific Architectures to Bridge Design and Implementation. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 341–358. Springer, Heidelberg (2011)
15. Harrison, C., Benko, H., Wilson, A.D.: OmniTouch: wearable multitouch interaction everywhere. In: UIST 2011, p. 441. ACM Press, New York (2011)
16. Hatchuel, A., Weil, B.: C-K design theory: an advanced formulation. Research in Engineering Design 19(4), 181–192 (2008)
17. Hornecker, E.: Creative idea exploration within the structure of a guiding framework: the card brainstorming game. In: TEI 2010, pp. 101–108. ACM Press (2010)
18. Hornecker, E., Buur, J.: Getting a grip on tangible interaction: a framework on physical space and social interaction. In: CHI 2006, pp. 437–446. ACM, Montréal (2006)
19. Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-based interaction: a framework for post-WIMP interfaces. In: CHI 2008, p. 201. ACM Press (2008)
20. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: TEI 2007, pp. 139–146. ACM (2007)
21. Koleva, B., Egglestone, S.R., Schnädelbach, H., Glover, K., Greenhalgh, C., Rodden, T., Dade-Robertson, M.: Supporting the creation of hybrid museum experiences. In: CHI 2009, p. 1973. ACM Press, New York (2009)
22. Löwgren, J., Stolterman, E.: Thoughtful interaction design. MIT Press, Cambridge (2004)
23. Monk, A., Howard, S.: Methods & tools: the rich picture: a tool for reasoning about work context. Interactions 5(2), 21–30 (1998)
24. Muller, M.J.: PICTIVE - an exploration in participatory design. In: CHI 1991, pp. 225–231. ACM, New Orleans (1991)
25. Schön, D.A.: The Reflective Practitioner: How Professionals Think In Action, 1st edn., pp. 305–308. Basic Books (2008)
26. Shaer, O., Jacob, R.J.K.: A specification paradigm for the design and implementation of tangible user interfaces. TOCHI 16(4), 1–39 (2009)
27. Simon, H.A.: Les Sciences de l'Artificiel, 3ème edn., p. 215. Folio (1969)
28. Stolterman, E., Wiberg, M.: Concept-Driven Interaction Design Research. Human-Computer Interaction 25(2), 95–118 (2010)
29. Ullmer, B., Ishii, H., Jacob, R.J.K.: Token+constraint systems for tangible interaction with digital information. TOCHI 12(1), 81–118 (2005)
30. Visser, W.: Designing as construction of representations: a dynamic viewpoint in cognitive design research. Human Computer Interaction 21(1), 103–152 (2006)
31. Youn-kyung, L., Sang-Su, L., Da-jung, K., Lim, Y., Lee, S.-S., Kim, D.: Interactivity Attributes for Expression-oriented Interaction Design. IJDesign 5(3), 113–128 (2011)

Model-Based Self-explanatory UIs for Free, but Are They Valuable?

Alfonso García Frey[1], Gaëlle Calvary[1], Sophie Dupuy-Chessa[2], and Nadine Mandran[3]

[1] Grenoble INP
[2] UPMF
[3] CNRS, LIG
41 Rue Des Mathématiques,
38400 Saint Martin d'Hères, France
{Alfonso.Garcia-Frey,Gaelle.Calvary,
Sophie.Dupuy,Nadine.Mandran}@imag.fr

Abstract. Model-Driven Engineering (MDE) has been extensively used for generating User Interfaces (UIs) from models. As long as these models are kept alive at runtime, the UIs are capable of adapting to variations of the context of use. This paper investigates a potentially powerful side effect: the possibility of enriching the UIs with explanations directly generated from these models. This paper first describes a software infrastructure that supports this generation of explanations. It then reports on a user study that evaluates the added value of such model based self-explanations.

Keywords: Self-Explanatory User Interfaces, Model-Driven Engineering, Models at runtime, Self-Explanation.

1 Introduction

Many works ([9, 10, 14]) have reported on the benefits of supporting users through explanations in interactive systems. These explanations address specific questions that users ask about the User Interface (UI). For instance, *how* a task can be accomplished, *why* a feature is not enabled, or *where* an option is. Classical approaches [7], which are based on predefined information such as static documentation, FAQs, and guides, specify this information at design time. Their scope is therefore limited because users can have questions about the UI that are not covered by these kinds of supports. Moreover, this static documentation is not only a time consuming task but, additionally, it requires manual updates when the program specification changes. To overcome this limitation, some researches [4] have recently proposed Model-Driven Engineering (MDE) as a means for supporting users at runtime. Model-Driven UIs use the models created at design time as their knowledge-base at runtime, exploiting the models and the relationships between them to find answers to the users' questions. These kinds of UIs with support facilities based on their own models are also known as Self-Explanatory UIs. Their main advantages are that answers are generated at runtime, and they evolve with the program specification automatically.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 144–161, 2013.

This paper firstly describes a self-explanatory system, explaining how to compute different types of explanations based on the underlying models. This system is representative of what explanations a model-based UI can provide. The paper then reports on an evaluation about the suitability of model-driven engineering for explanation purposes. We conducted a user study on such a self-explanatory UI. This UI reproduces an existent website, where all the necessary models have been generated by a reverse engineering process. The resultant model-driven UI has been extended with support facilities based on such models, to allow users to ask questions about the UI at runtime.

The paper is structured as follows. The next section provides related work on model-based explanations. Then, the paper describes the working hypothesis. Later, it presents the infrastructure of the self-explanation system that we used in our experiment. We also explain the algorithms that we used to, based on the models available at runtime, generate the questions and answers according to the model-driven approach. The paper then covers the user study, describing the experiment and the evaluation protocol. The final section analyzes and discusses the results of the qualitative analysis we did on the collected data. Finally, the paper ends with a conclusion and a brief discussion about the future perspectives.

2 Related Work

This section briefly explains the role of MDE in HCI and how its different types of models are used in different works for supporting users with specific types of explanations. A different type of explanation is provided regarding the nature of the question asked by the user. For instance, [15] describe five categories of questions: goal-oriented (*What can I do with this program?*), descriptive (*What is this? What does it do?*), procedural (*How do I do this*?), interpretive (*Why did this happen*?) and navigational (*Where am I? Where is it?*).

Other authors describe similar categories but with different terms. For instance in [17] we find conceptual explanations (*What is this?, What is the meaning of this?),* Why-explanations describing causes and justifications for facts, How-explanations for describing processes, Purpose-explanations (*What is this for?* or *What is the purpose of this?*), and cognitive explanations, which “explain or predict the behavior of ‘intelligent systems’ on the basis of known goals, beliefs, constraints, and rationality assumptions” [17].

Some of these types of explanations have been provided through different models in model-based approaches. We firstly explain the model-driven approach of UIs before presenting such model-based explanations.

2.1 MDE for HCI

Model-Driven Engineering (MDE) has been recently applied to the engineering of UIs. It consists in describing different features of UIs (for instance, tasks, domain, context of use) in models from which a UI is produced according to a forward engineering process [19]. An example of a MDE-compliant approach is the Cameleon Re-ference Framework [2], where a task model is transformed into an Abstract User Interface (AUI) model, which is in turn transformed into a Concrete User Interface (CUI) model representing the interactors or widgets, from which the Final User Interface (FUI), i.e., the code of the UI itself, is derived. The experimentation presented later is based on this approach.

Some models of the Cameleon Reference Framework existed before MDE and they have been already used separately for explanation purposes in several works. For instance, task models have been extensively used for automatically generate procedural information in different forms.

2.2 Model-Based Explanations

An early example that employs a task model (in the form of user's actions) for explanation purposes is Cartoonist [18]. Cartoonist generates GUI animated tutorials to show a user how to accomplish a task, exploiting the model for providing run-time guidance.

Pangoli and Paterno [13] allow users to ask questions such as *How can I perform this task?* or *What tasks can I perform now?* by exploiting a task model described in CTT. Contrary to Cartoonist, answers are provided in [13] in natural language. Tasks modeled in the form of Petri Nets are used for similar purposes by Palanque et al. in [12], answering questions such as *What can I do now?* or *How can I make that action available again?*

Other works report on the usage of task models as a means for creating collaborative agents that help the user [3].

Behavioral models, presented in different forms, have been also used to support *Why* and *Why not* questions in user interfaces. In [12] *Why* questions are answered using the same approach based on Petri Nets that is exploited for procedural questions. By analysing the net it is possible to answer questions such as *Why is this interaction not available?*

The Crystal application framework proposed by Myers et al. [10] uses a “Command Object model” that provides developers with an architecture and a set of interaction techniques for answering *Why* and *Why not* questions in UIs. Crystal improves users’ understanding of the UI and help them in determining how to fix unwanted behavior.

Lim et al. [8, 9] observed that *why* and *why not* questions improve users' understanding and confidence of context-aware systems.

Vermeulen et al. [20] propose a behavior model based on the Event-Condition-Action (ECA) paradigm, extending it with inverse actions ($ECAA^{-1}$) for asking and answering *why* and *why not* questions in pervasive computing environments.

These researches show explanations based on individual models. We aim to evaluate the suitability and added value of model-based approaches of UI, that can use one or more different models at the same time. In particular, we want to see whether these model-based approaches can generate more powerful explanations or have any extra added-value with regard to the previous isolated solutions. This is what our working hypothesis describes in the next section.

3 Working Hypothesis

In our previous work, we tried to combine different models for explanation purposes. We explored [5] how to let designers to describe the task model behind a UI, which is provided in the form of a Final UI (the model representing the UI itself), so procedural explanations could be provided based on this task model. We also explored how to support designers through design questions based on design rationale notations [6], and an architecture for the unification of several help facilities in [4], showing how to employ CUI models for *Where* questions, or annotated task models for providing descriptions at runtime.

These last works, that combine and integrate different models to compute several types of explanations, suggest that the model-driven approach of UIs, which is based on similar design models to those presented in the related works, should provide users with explanations that are at least of the same interest, usefulness and relevance than those provided by each model individually. Moreover, as model-based UIs can rely on a greater number of models or even include new ones, the power of explanation could be potentially better as well.

To verify the added-value of the model-driven approach of UIs for explanation purposes, we conducted an experimentation based on a model-driven prototype, which is described in the next section. The prototype was built according to the Cameleon Reference Framework, and it includes a self-explanatory facility in the form of a dialog to let users ask different types of questions about the UI. This system is representative of what explanations a model-based UI can provide.

4 Prototype Description and Supporting Infrastructure

This section firstly describes the prototype and the infrastructure of the help facility. Then it details the algorithms used for computing the questions and answers provided to the users, detailing how the models are used at runtime.

4.1 Prototype Description

The prototype consists in a cars shopping website called UsiCars. This website is inspired by a real site from a real car manufacturer. We reproduced only the part of the website that is devoted to the selection and configuration of the vehicles, keeping the options and the structure of the original website.

This website was chosen for two main reasons. The first one is that we needed to use a system that contains knowledge that is understandable and accessible by all the participants, but complex enough for not being easy to use. A website for configuring cars covered this point as all the participants understand many of the car related concepts, but at the same time there are enough specific options with domain related concepts to create complex tasks that are non trivial to perform. The second reason is that we found the original website difficult to use by real users in different forums.

The reproduction of the website was done by a reverse engineering process. The first step was to explore all the different tasks that the user can perform to select and configure a vehicle. We created a task model according to this information. Secondly, we created a transformation to obtain an Abstract UI model that conforms to the structure of the original website. Thirdly, we wrote another transformation to generate the Concrete UI model from the Abstract UI model. This transformation produces all the widgets that we find in the original website. We also used the same images and we respected the same sizes for all the widgets from the original site, to ensure that we obtain the same usability properties. Finally, we wrote another transformation to generate the Java code and produce the resulting site.

In each one of the model to model transformations, we generated not only the target model but also mapping models that keep track of the successive transformations of an element from one model to another. For instance, in the transformation from the task model to the Abstract UI model we generated a *Mapping-Task2AUI* model that specifies what tasks are transformed into what Abstract UI elements. The same principle was applied to obtain a *Mapping-AUI2CUI* model. This allow us to go through the transformation chain and, for instance, retrieve the source task from which a button has been generated.

Fig. 1. Screenshot of the prototype

Figure 1 shows an excerpt of the UI of the prototype. The UI is divided into two main parts. A big area in the middle and a thin area at the bottom. The central area of the UI has two different roles. On one hand it serves as a visual feedback for the user when he/she selects a car model or changes the color of the vehicle (figure 1). On the other hand, it can show dialogs containing all the possible options that the user can select to configure the car with. The thin area at the bottom allows users to navigate through several categories of options for accessing different features of the car such as the electronic equipment or the exterior color of the vehicle.

The prototype was build according to the approach described in [4], so the different types of questions previously discussed could be generated and fully integrated into the UI of the prototype. The infrastructure (figure 2) consists in two model-based UIs, the self-explanatory facility for providing the help and the application. For a discussion on how to mix both set of models see [4]. The functional core of the help UI is composed of 4 modules for generating the list of questions (QG), interpreting (I) a user's request, i.e., inferring the type of question and its parameters, the processor (P) that computes the answer based on such parameters, and the answer generator (AG) that presents the answer back to the user. Each of these four modules of the functional core of the self-explanatory facility has full access to the models of the underlying application at runtime.

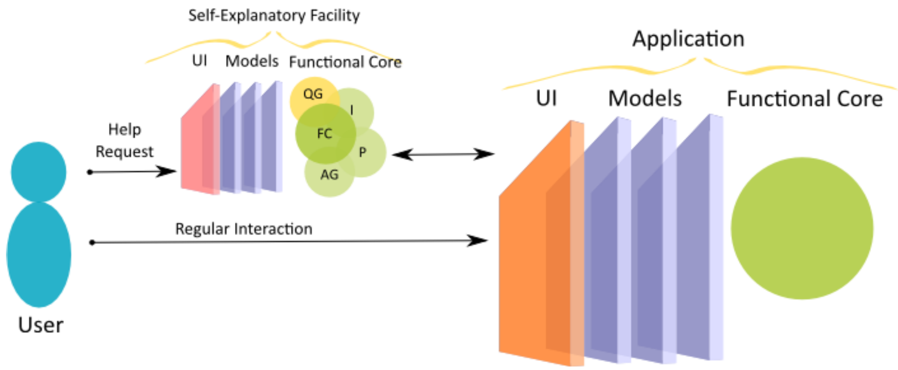


Fig. 2. Infrastructure for self-explanatory UIs. The possible questions are generated by the Question Generator (QG) from the Functional Core of the help facility (FC). For each user's request, the Interpreter (I) determines its type and parameters, used by the Processor (P) to compute the answer, which is presented in some form (textual in this prototype) to the user thanks to the Answer Generator (AG). These four modules use the application models at runtime.

4.2 Self-explanation Infrastructure and Questions / Answers Computation

To check the added-value of model-driven UIs in terms of explanation purposes, we supported six different types of questions built all of them upon the main models of the Cameleon Reference Framework. Figure 3 summarizes the models that have been used to generate the questions and their respective answers. For each of the types, we explain the algorithm we followed to generate the questions, the algorithm for the answers, and the involved models used in each case (summarized in figure 3).

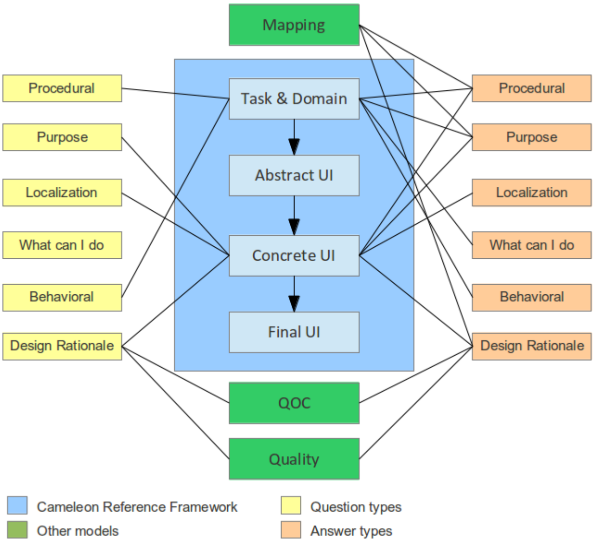


Fig. 3. Models used for generating questions (left) and answers (right)

4.3 Procedural Questions - How

To generate How questions, we explore the task model recursively from the root task to the leaves. For each node representing a task, we create a question in a textual form according to the following grammar:

How to + *Task.name* + ? [1]

were tasks are named starting with a verb following a standardized convention. An example of a How question is:

How to choose Packs?

The computation of the answer is done as follows. First, we locate the task inside the task model. Second, we inspect the mapping model that maps tasks to AUI elements from the AUI model, so we can retrieve the abstract UI element that resulted from transforming such task. Once the AUI element has been found, we repeat the procedure to locate the CUI element derived from this AUI element. This is done by inspecting the mapping model that keeps track of the transformations from AUI elements to CUI elements. Once the CUI element has been retrieved, we compose the answer with following grammar:

Use the + *CUI-elem.name* + *CUI-elem.type*

[1] Original questions were asked in french. The shown grammar as well as the examples and their related answers are adaptations from the original questions presented in this research.

An example of a computed answer using this approach is:

Use the Packs button

Note that the answer can be completed with the information about the localization of the widget, which is computed later in the Where questions. In this way, a more elaborated answer for CUI elements that were not directly visible from the user's were composed as follows:

Use the + *CUI-element.name* + *CUI-element.type* +
in the + *CUI-element.parent* + *CUI-element.parent.type*

where an example is:

Use the 'Pack Connected Drive' checkbox button in the 'Optional Equipment' panel.

4.4 Purpose/Functional Questions - What Is It for

The purpose questions generated in the prototype were of the form:

'What is the + *CUI-element.name* + *CUI-element.type* + for?'

An example of a purpose question is:

What is the 'Optional Equipment' button for?

To compute these questions, we iterate through the CUI model of the UI, adding a question for each new element. We added questions for all the CUI elements except for layouts, as they are the only CUI elements that are not directly visible by the user.

Answers where computed as follows. First we inspect the mapping model between the AUI and the CUI models to retrieve the AUI element from which the CUI element has been generated. Once we have the AUI element, we retrieve the task originating this AUI element, i.e., the source of the transformation chain. Once the task has been retrieved, we directly provide the name of the task, answer is computed using the name of the task in the following grammar:

To + task.*name*

As in the example:

To 'Select the optional equipment'

Even if this question is mostly useful for images or icons that have an unclear meaning, we also generated the questions and answers for the rest of the CUI elements, even if they presented textual information that made clear the purpose of the object.

4.5 Localization Questions - Where

The generated Where questions are of the form

'Where is the + *CUI-element.name* + ?'

As in the example:

'Where is the Tuner DAB?'

The process of generating these types of question is quite similar to the previous purpose questions. We only considered CUI elements having textual information, i.e., labels, any kind of textual buttons such as normal buttons, checkboxes or radio buttons, menus, menu options, and window titles. The reason for avoiding other types of widgets like images is that we didn't want the user to describe such widgets and thus, asking open questions that the system couldn't understand.

Answers were computed by finding the direct parent or container of the CUI element. This is, we first locate the CUI element in the CUI model and then we retrieve its parent, avoiding again layouts that are not visible for the user. For the Where question given in the previous example, the Tuner DAB refers to a checkbox button located on the 'Optional Equipments' panel. Thus, the grammar generating the answer is:

'The + *CUI-element.name* + *is on the* +
CUI-element.parent + *CUI-element.type*'

So the answer given by the system is:

The Tuner DAB is on the Optional Equipment Panel

4.6 What Can I Do Questions

The "What can I do now?" question provides information about what tasks are currently available to the user regarding its current situation in the UI, i.e., depending on the current task that the user is currently performing at the moment of asking the question. As not all the tasks are always available at the time, answers for the same question can vary in time. The presented question is then always of the form:

What can I do now?

The computation of the answer relies on the task model. We first retrieve the current task in the task model. We used the CTT notation to describe the task model so we find the available tasks as follows. From the current task in the task tree, we compute which sister tasks are available regarding the LOTOS operators used by CTT. We add the name of each available task to answer. We then recursively iterate from the current task to the root task of the tree, adding all the available tasks. We finally add the available sub-tasks. The final answer is then a list of tasks shown according to the next grammar:

You can + *task-1.name* + … + *task-N.name*

For example, when the user accesses to the *Optional Equipment* panel, the answer to What can I do now? is:

You can select the external equipment, select the internal equipment, select the internal decorations, select the functional equipment, select the on-board electronics, select the wheel rims, select the maintenance contract.

4.7 Behavioral Questions – Why

Behavioral questions were generated under the form:

Why I can't + *task-N.name* + ?

Where the task *task-N* is unreachable from the current task. For instance:

Why I can't Visualize the car?

To compute questions we proceed as for What can I do now?, locating the current task in the task tree first. We then locate all the unreachable tasks in a similar process, i.e., locating unreachable sister tasks (due to the CTT LOTOS operators) and traveling the task tree to the root and to the leaves. For instance, a task B enabled with information from a task A (A []» B) is unreachable until the information is received. Questions are added for all the unreachable tasks following the previous grammar.

Answers are computed by finding the path that enables the given task. If a task is not reachable it means that some task or tasks need to be done. We find these tasks by traveling the sister and mother tasks (up to the root), locating the LOTOS operators that enable the desired task. For instance, for the task 'Visualize the car', the task was reachable by selecting the model of the vehicle first, so the provided answer is:

You need to Select the model

which conforms to the grammar we used:

You need to + *task-1.name* [+, task-N.name]*

4.8 Design Rational Questions

We also included a QOC and a quality model to compute questions and answers about the design rationale of the UI. A detailed description of the models and the procedure can be found in [6]. The proposed questions were directly retrieved from the QOC model. This model is a simplified version of the one in [6]. Answers were supported by one ergonomic criterion. The answers were computed according to the quality criteria supported by each option of the QOC model as shown in the following example:

Why the engines are ordered by price?

The provided answer, which follows the grammar "Because the ergonomic criterion is + *criterion.description*", is:

Because the ergonomic criterion 'Items of any select list must be displayed either in alphabetical order or in any meaningful order for the user in the context of the task'.

4.9 Self-explanatory Dialogue

The questions were presented in a textual form inside a dialog box (figure 4). Textual answers shown up after clicking on the desired question. In the experiment, questions were presented one by one and only at the end all the questions were shown together. We didn't filter out any question in this dialog, i.e., all the possible questions that the system was able to answer were proposed to the user. The reason for this was to show the user all the questions, so he/she can better realize if the self-explanatory system could cover his/her expectations for the given type of question. For instance, if the user realizes that his/her question is missing in the list.

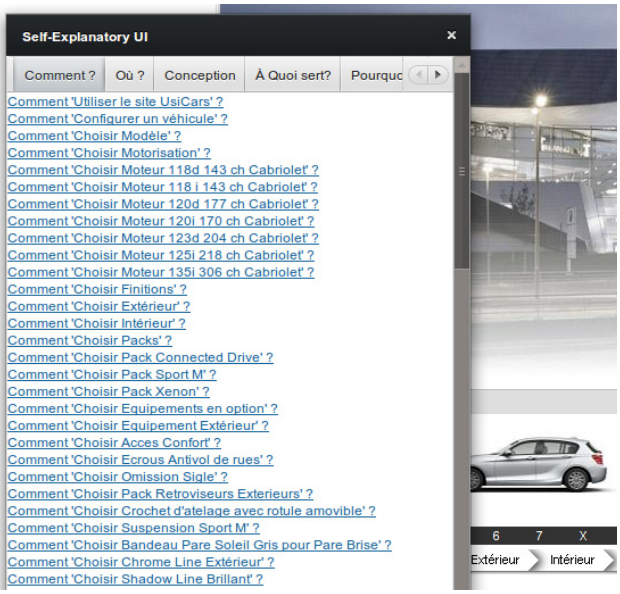


Fig. 4. Self-Explanatory Dialog box showing the full list of available questions by type

5 User Study

We conducted an experiment to evaluate the added value of model-based self-explanations. This section starts describing the participants involved in the experiment. Then it describes each of the different phases that integrate the evaluation protocol. Last section details the prototype that we have developed to carry out the study.

5.1 Participants

We selected 20 participants, all between 23 and 39 with an average age of 27.4. From the 20 participants, 12 were male and 8 female. We recruited individuals regardless their experience with interactive systems because the possible added value of model-based explanations can vary regarding the experience of each profile.

5.2 Method

To carry out the study, we broken-down the evaluation protocol into three different phases.

In the first phase, we asked the participants to answer a questionnaire. This questionnaire allowed us to better know the background of participants, to understand their habits regarding how they use new technologies in general, what are their common uses, the kind of applications they use with a relevant frequency, the problems they use to find with these or other applications, as well as their habits for solving these problems. The questionnaire also included questions regarding how participants used the help provided by the applications they use, and how they used to proceed in case they have a problem with the application. A software recorder was used to record all the answers of all the participants.

In the second phase of the experiment, we asked the participants to use the prototype that we had developed to this aim. We asked them to complete 10 different tasks in an established order. All the participants received the identical 10 tasks. We randomized the order of the tasks for each participant to avoid side effects such as the influence between different tasks or memory effects that can help users to accomplish the tasks better in a certain order. This part of the experiment was conducted on a laptop and the audio was recorded. We asked the participants to verbalize their thoughts, specially the questions they would like to ask to the system and the problems that they find when accomplishing the tasks.

The third part of the experiment presented the prototype including a self-explanatory dialog that contained one type of question at a time. The six questions discussed previously were presented one after another again in a randomized order. For each type of question, the dialog box showed all the possible questions that the participants could ask. Every time we showed a new type of question, we asked the participants their opinion about it, including the possible advantages and disadvantages of asking that question to the UI. We asked as well if the given type of question could be useful in the previous phase of the experiment. At the end of the third phase, all the types of questions were shown together into the same self-explanatory dialog, and we asked some more general questions that are discussed in section 6.

5.3 Tasks

The motivation for the second phase of the experiment was to confront the users with different kinds of problems that are frequently found in UIs. To this end we designed 10 different tasks. The tasks were selected according to their complexity, ranging from easy tasks to more complex ones. We didn't force any specific problem in the tasks that could be easily solved by one of the previous questions. Instead, we tried to reproduce a realistic use case with a varied set of tasks so the answers of the participants in phase 3 were not influenced by the second phase.

The 10 different tasks that we asked the participants to complete are shown in table 1. The accomplishment ratio indicates whether the participants were able to complete the tasks at all. A few users that got stuck and required hints were counted as unsuccessful. The accomplishment ratio gives an idea of how difficult each task was, regardless the expertise of the user.

Some of the tasks involved selection with searches through small lists (1, 2, 4, 8) while others involved selection through lists having multiple options and categories (5, 7, 9) in different locations. Tasks 1, 2 and 4 involved selections through images while the rest of the selection tasks were through options in textual form. Other tasks involved verification (6, 7, 10), comparison (7), or manipulate cars related terminology that was more or less easy to understand (1, 4, 6, 8).

Table 1. List of tasks and their accomplishment ratios. The tasks were randomized to avoid side effects such as the influence between tasks or memory related effects. The accomplishment ratio give an idea of the difficulty of the task.

Task description	Accomplishment ratio
1. Select a "Cabriolet" model	20/20
2. Select a diesel engine for less than 35.000 €	17/20
3. Choose a sport finishing touch	15/20
4. Change the exterior color to *Le Mans Blau*	15/20
5. Ensure that the model has a navigation system. If not, add one.	12/20
6. Ensure that the model has a *Terra* leather upholstery. If not, choose a blue leather instead.	12/20
7. Make sure that you can listen music in the car. If not, choose the best audio system available.	12/20
8. Select the Connected Drive pack	18/20
9. Select a Maintenance Contract of your choice	10/20
10. Visualize the result and check that everything is OK. If not, try to solve the problem.	12/20

We used the accomplishment ratio in the last part of the experiment, specially when we asked the participants if they believed that the model-based explanations could help them to complete one of the problematic tasks, or doing it in a more efficient way. The next section discusses the results of the qualitative analysis that we carried out with all the collected data.

6 Qualitative Analysis

A large amount of qualitative data was collected from the experiment. We extracted around three hundred comments from the records made during the second phase of the experiment, the one in which participants were asked to complete the list of tasks. The selected method of analysis was the thematic type [11]. This method is focused on the answers and comments recorded during the experiment, and classified into categories later on. The aim of the thematic analysis is to group together answers or parts of answers that have the same meaning. The thematic groups were then analyzed to identify the different categories of opinion. The objective is to gather and list all the themes covered by the answers to reflect the widest possible range of opinions, distinguishing the positive ones from the negative ones.

From the extracted comments, we identified around 250 verbatims that referenced types of questions either in an explicit or implicit way. Only those verbatims that clearly related a question type where considered. For instance, verbatims like *"I don't know where the contracts are"* were classified as an implicit question of type *Where*. The table 2 shows the results of this classification, as well as some illustrative verbatims. It is significant that most of the verbatims addressed navigational problems (Where + How) mainly due to usability issues and to the nature of the tasks (table 1). The high number of 'Other types' is mainly due to questions about semantic information relating concepts specific to the domain. These questions are described in section 6.2, while next section presents the findings for both positive and negative opinions, as well as some revealed limitations of the approach.

Table 2. Relationship between question types and ocurrences extracted from the records during the second phase of the experiment. An exemple of verbatims illustrates each type.

Question type	Example Verbatims	Occurrences
How	*I don't see how to do it*	13
Why	*Why do I need to register?*	21
Where	*And where do I find the maintenance contracts*	119
What is it for	*I'm browsing the tabs to see what they do*	7
What can I do now	*I must find my way (inspecting all the UI with the mouse)*	2
Design rationale	*Why they are not ordered by type?*	1
Other types	*What does Cabriolet stand for? (Definition)*	81
	What are the differences between the packs? (Differences)	
	Is it included in the price I guess? (Confirmation)	
	What happens if I click here? (What if)	

6.1 Findings

In the first phase of the study we collected the data described in the previous Participants section, and we also found that 16/20 liked new technologies, 17/20 use new technology everyday, and 20/20 have found problems in their use. To face these problems, 11/20 inspect the UI to try to solve them by themselves, 8/20 ask other people about the problem, 15/20 search for solutions in the Internet, and 7/20 use the help provided by the system.

The last phase of the study revealed that questions of types *How* and *Where* were identified by most of the users (15/20) as useful and helpful with statements such as "*it can be very useful in certain situations*" or "*It could be very helpful for locating all the options of the vehicle in a faster way*". This last statement refers also to a gain of time, which was also identified as a positive value by a total of 10/20 users with statements such as "*It is a gain of time*" or "*it makes me go faster without losing my time*". The good acceptance of *How* questions contrasts however with the low number of verbatims. This suggests that users find the information useful but they are not thinking of asking it. The help UI could encourage/propose questions in these situations.

The *What is it for* and *Why* questions were also identified as useful by an important number of participants, but less useful than the previous ones. This was mainly due to the fact that subjects didn't find useful to ask for the purpose of some elements of the UI, such as check-boxes or labels, that already contain clear information about what they are currently doing. In the case of *Why* questions, the results didn't show a good acceptation by the participants as in the results found by [9, 10]. This was due to the fact that the questions proposed by our algorithms didn't cover all the possible range of questions that the participants asked. For instance, as our algorithms entirely rely on the task model, our system couldn't answer why questions concerning the functional core of the application such as *Why there is no diesel engines?* (for specific kinds of Cabriolet cars). To overcome this limitation in our implementation, we propose to enrich the Cameleon models with other models more suitable for this purpose, as for instance those used in [10, 20]. Another possible solution is to enrich the mapping model to include the ECA rules used in [20], so we can connect the methods of the widgets triggering the action directly to the functional core and vice versa.

Finally, the *What can I do now* and design rationale related questions were found to be useless by most of the participants (16/20), according to statements such as "*I don't see where I would like to ask this question*" (for *what can I do now?)* or "*I am not interested in this information, all I want is to buy my car*" (for design rationale questions).

At the end of the third phase, when we presented the help UI with all the types of questions together, the study revealed that in general, model-based self-explanatory facilities were identified as "useful" and "helpful" by most of the participants (16/20). The study also revealed question types that were not supported by our current implementation. The analysis of the collected data suggests that our model-based self-explanatory UI, with minor design enhancements for major usability improvements, could have the potential to easily help the users. Next section discusses the possible model-based implications for the types of questions that were not supported. Then, we discuss the usability suggestions extracted from the data for our particular implementation of self-explanatory UIs.

6.2 Unsupported Types of Questions

We identified other types of questions not explicitly supported by our system. A minor number of them referred to *What if* questions. Even if most of these verbatims come from users that showed a trial and error approach to understand the consequences of their actions in the UI (i.e., they don't know the consequences of an action but they perform such action on the UI anyway to see what happens), 2 users out of 20 didn't use options from the UI because they didn't know their possible side-effects. For instance, subject 9 didn't perform one of the tasks because *"I have fear of losing all the options"*. Supporting *What if* questions can help this minority of users to feel more comfortable with the UI. These kinds of questions can probably be answered by analyzing the operators of the task model and how they are transformed to CUI elements, (what elements of the CUI model become active/inactive as we enable/disable new tasks. These answers will probably require some improvements for side effects related to the functional core of the application (external to the UI).

We also identified a high number of verbatims requesting confirmation and validation from the UI. For instance, "does the car already have a navigation system?", "are the options included in the price?", were recurrent expressions used by the participants. This observation suggests that the feedback provided by the site was not enough for the users. Supporting questions about confirming and validating the user actions can help to overcome this usability issue. This may require new models for handling user actions, specially those that have effects beyond the UI.

A third group of questions not supported by the self-explanatory dialog concerns definitions. Most of these questions were about specific car-related terminology and concepts such *"What is the Tuner DAB?"* or "*What does Cabriolet stand for?*". To support these questions, the proposed model-based approach needs to be extended with semantic information, either by adding new models or by connecting the UI with sources of semantic information (internet).

Semantic information may be also necessary for answering questions about differences that we identified in a minor number of verbatims, for instance, *What is the difference between the packs?* (or eventually similarities).

6.3 Usability Suggestions and Improvements

We were also interested in usability observations. During the third phase of the experiment, where participants were confronted to the self-explanatory dialog, 14 out of 20 suggested that they would like to type the whole question directly instead of clicking on a predefined answer inside a list. 13 out of 20 would like to access questions by typing keywords in a text area, and 4 proposed to use a vocal interface instead. These observations sustain some of the design principles for help systems of the literature, in particular, “Help should be accurate, complete and consistent” ([1,16]), and “Help should not display irrelevant information” ([7]).

6 participants suggested to classify questions not only by question types but following the categories of the underlying site, for instance, grouping them by equipment or car models.

Regarding the answers, some participants argued that they don't like to read explanations, specially those that have a significant length. With the models used in this approach, the information given in the answers can be represented in non textual forms. For instance, as the CUI model can store the screen coordinates of the widget, *Where* questions can be answered by highlighting the region of interest (as currently done in mac systems), and procedural questions can be explained by means of animations of the cursor over the widget coordinates.

Finally, some participants proposed that it would be preferable to use the questions not as a means to know how to find a specific option, but to “*get there*”. This suggests that self-explanatory UIs could be used as software agents to overcome the usability issues of a UI not only by explaining to the user how to solve the issue, but solving it directly if possible. For instance, navigating to the desired website instead of explaining what website the user should navigate to. This observation opens new research questions: can self-explanatory UIs benefit from agents? If so, what other models are needed and how this can be done?

7 Conclusions and Perspectives

Model-Driven Engineering of UIs has been extensively used for the automatic generation and adaptation of UIs. This paper studies a side effect of keeping these models at runtime. They can be used for supporting the users through explanations based on such models. These explanations have interesting advantages as the dynamic generation of answers and the automatic evolution along with the models, so no manual updates of the support are required when the program specification (the models) changes. But do users think that this model-based support is relevant and useful enough? The experiment that we conducted shows that most of the users identifies model-based explanations as potentially useful.

Our future work includes to test how scalable model-based explanations are, either with a huge number of models or with a huge number of users requesting answers. We will study how to support the new types of questions that we have identified, and better support the current questions that we are able to compute.

We also plan to investigate the use of design rationale questions to support the learning of HCI design methods.

Acknowledgments. This work has been funded by the UsiXML. We warmly thank all the participants that collaborated in this experiment. Note that all the content used in the prototype, extracted from the original website during the reverse engineering process, belongs to this website and has been used for research purposes only.

References

1. Dix, A., Finlay, J.: Human Computer Interaction. Prentice Hall, London (1993)
2. 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)
3. Eisenstein, J., Rich, C.: Agents and GUIs from Task Models. Information Science (2002)
4. García Frey, A., Calvary, G., Dupuy-Chessa, S.: Users need your models! Exploiting Design Models for Explanations. In: Proceedings of the 26th BCS HCI Group Conference, Birmingham, UK, September 12-14 (2012)
5. García Frey, A., Calvary, G., Dupuy-Chessa, S.: Xplain: an editor for building self-explanatory user interfaces by model-driven engineering. In: Proceedings of EICS 2010, pp. 41–46. ACM, New York (2010)
6. García Frey, A., Ceret, E., Dupuy-Chessa, S., Calvary, G.: Quimera: a quality metamodel to improve design rationale. In: Proceedings of EICS 2011, pp. 265–270. ACM, New York (2011)
7. Horton, W.: Designing and Writing On-line Documentation, 2nd edn. John Wiley & Sons, New York (1994)
8. Lim, B.Y., Dey, A.K.: Assessing demand for intelligibility in context-aware applications. In: Proceedings of Ubicomp 2009, pp. 195–204. ACM (2009)
9. Lim, B.Y., Dey, A.K., Avrahami, D.: Why and why not explanations improve the intelligibility of context-aware intelligent systems. In: Proceedings of CHI 2009, pp. 2119–2128. ACM (2009)
10. Myers, B.A., Weitzman, D.A., Ko, A.J., Chau, D.H.: Answering why and why not questions in user interfaces. In: Proceedings of CHI 2006, pp. 397–406. ACM (2006)
11. Paillé, P., Mucchielli, A.: L'analyse qualitative en sciences humaines et sociales. Armand Colin, Paris (2003)
12. Palanque, P., Bastide, R., Dourte, L.: Contextual help for free with formal dialog design. In: Fifth International Conference on Human-Computer Interaction. Elsevier Science Publisher (1993)
13. Pangoli, S., Paterno, F.: Automatic generation of task-oriented help. In: Proceedings of UIST 1995, pp. 181–187. ACM, New York (1995)
14. Purchase, H.C., Worrill, J.: An empirical study of on-line help design: features and principles. International Journal of Human Computer Studies 56(5), 539–567 (2002)
15. Sellen, A., Nicol, A.: Building user-centred on-line help. In: Baecker, R., Grudin, J., Buxton, W., Greenburg, S. (eds.) Readings in Human Computer Interaction, 2nd edn., Morgan Kautmann, San Francisco (1995)
16. Shneiderman, B.: Designing the User Interface, 2nd edn. Addison-Wesley (1992)
17. Spieker, P.: Natürlichsprachliche Erklärungen in technischen Expertensystemen. Dissertation, University of Kaiserslautern (1991)
18. Sukaviriya, P., Foley, J.D.: Coupling A UI framework with automatic generation of context-sensitive animated help. In: Proceedings of UIST 1990, pp. 152–166. ACM, New York (1990)
19. Vanderdonckt, J.: Model-driven engineering of user interfaces: Promises, successes, failures, and challenges. In: Proceedings of ROCHI 2008, pp. 1–10. Iasi, Bucarest (2008)
20. Vermeulen, J., Vanderhulst, G., Luyten, K., Coninx, K.: PervasiveCrystal: Asking and answering why and why not questions about pervasive computing applications. In: Proceedings of IE 2010, pp. 271–276. IEEE Computer Society, Washington, DC (2010)

Comparing Input Modalities for Peripheral Interaction: A Case Study on Peripheral Music Control

Doris Hausen, Hendrik Richter, Adalie Hemme, and Andreas Butz

University of Munich (LMU), Human-Computer-Interaction Group
Amalienstr. 17, 80333 Munich, Germany
{doris.hausen,hendrik.richter,andreas.butz}@ifi.lmu.de,
hemmea@cip.ifi.lmu.de

Abstract. In graphical user interfaces, every application usually asks for the user's full attention during interaction with it. Even marginal side activities often force the user to switch windows, which results in attention shifts and increased cognitive load. Peripheral interaction addresses this problem by providing input facilities in the periphery of the user's attention by relying on divided attention and human capabilities such as proprioception and spatial memory. Recent work shows promising results by shifting tasks to the periphery for parallel task execution. Up to now, most of these interfaces rely on tag-based objects, tokens or wearable devices, which need to be grasped and manipulated, e.g., by turning, moving or pressing the device.

To explore this design space further, we implemented three modalities for peripheral interaction with a desktop audio player application – graspable interaction, touch and freehand gestures. In an eight-week in-situ deployment, we compared the three modalities to each other and to media keys (as the state-of-the-art approach). We found that all modalities can be successfully used in the (visual and attentional) periphery and reduce the amount of cognitive load when interacting with an audio player.

With this work we intend to (1) illustrate the variety of possible modalities beyond graspable interfaces, (2) give insights on manual peripheral interaction in general and the respective modalities in particular and (3) elaborate on paper based prototypes for the evaluation of peripheral interaction.

Keywords: Peripheral Interaction, Input Modalities, Audio Control.

1 Introduction

While writing a research paper, a user might repeatedly switch between a text processing program, an online dictionary and related articles. In addition, she might switch to an instant messenger to change her availability status or open her audio player to skip a song. While the first three activities are directly related to her current primary task – writing the paper – the other two are marginal side activities she either has to do or wants to do, but which are in no direct relation to the primary task. Although these activities are simple and fast in themselves, current graphical user interfaces make the user switch windows or click small icons and thereby impose context

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 162–179, 2013.

and focus switches. These switches cause a so-called resumption lag, which describes the time needed to resume work on the primary task [2]. Users also feel more stressed and cannot focus well if they are frequently interrupted by secondary tasks [4,24].

Peripheral interaction can reduce these interruptions of the users' primary focus by making use of peripheral attention. Current projects in this domain primarily use devices, which are grasped and then manipulated (e.g., Fireflies [8], PolyTags [26]). This is a logical choice as many peripheral interactions in our (non-digital) daily life are carried out with the hands (e.g., tying shoe laces while talking) [6]. However, other types of manual input, such as touch or freehand gestures, are hardly explored up to now for peripheral interaction. Opening the design space for other means of peripheral interaction makes room for additional application areas benefitting from less disruptive secondary tasks.

In order to explore this design space, we compared three input modalities (graspable interaction, touch and freehand gestures) (see Figure 1) in an eight-week in-situ deployment with eight participants. As an example use case, we chose the control of an audio player (play/pause, previous/next song and volume), since it is a common but short secondary task when working on a computer. For comparison, we included the commonly adopted media keys built into many keyboards.

We found that participants embraced all peripheral modalities (graspable, touch, freehand) and were able to use them in the periphery of their attention, which broadens the design space for manual peripheral interaction. Further we discuss the use of paper prototypes as an extension of the design process and in addition to the typical in-situ deployments of peripheral interaction. Finally, we highlight implications for manual peripheral interaction in general (e.g., the dominant hand is preferred even for very simple interactions) as well as some distinct characteristics of each modality.

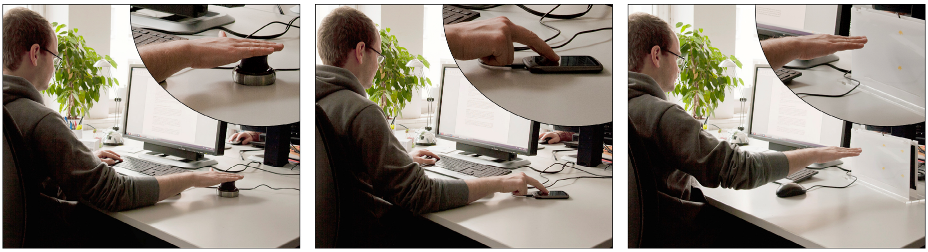


Fig. 1. User interacting with the peripheral modalities (graspable, touch and freehand) to control his audio player. The actual study was executed in the participants' homes, but in order to preserve privacy, these photos were taken in an office.

2 Related Work

Our work builds on existing work on peripheral interaction and previously explored input mappings for audio player control.

2.1 Peripheral Interaction

Peripheral interaction describes the interplay between several tasks, similar to multitasking. However, multitasking research mostly focuses on interruption management, i.e., finding the best possible moment for interrupting with a secondary task [25]. In contrast, peripheral interaction tries to minimize cognitive load and hence the effect of interruptions. The concept is similar to ambient information [28], but uses the periphery not only for perceiving information, but also for active interaction. Bakker et al. give an overview of attention theories relevant for peripheral interaction design [5].

Projects in the scope of peripheral interaction so far mainly use physical artifacts that can be manipulated by grasping them. Edge and Blackwell [13] propose a token-based task management system intended for "imprecise interaction with independently meaningful, digitally-augmented physical tokens". Olivera et al. [26] built tag-based tangibles to allow changing the status in a social network or controling the environment with commands such as *power on* and *off*. Their prototypes are intended for brief interactions while the main interaction focus is somewhere else. NoteLet [7], a bracelet for primary school teachers, facilitates taking notes of observations in class. To do so, teachers can take a photo by squeezing the device on their wrist. FireFlies [8] are also designed to aid teachers for example by giving feedback on the fly. The wearable device can be attached to the clothes. By squeezing a bead (each pupil has one corresponding bead) signals are sent to a light object located at each desk. StaTube [17] is a cylindrically shaped tangible, which can be pushed down and rotated to update one's Skype status. All these graspable prototypes have been evaluated and show that graspable devices are one possibility to support peripheral interaction

The Appointment Projection [15] is the only prototype in the scope of peripheral interaction, which relies on freehand gestures to acquire information about the next upcoming event in a calendar or to silence an ambient reminder animation. Similarly, related fields such as eyes-free interaction – which is in line with the concept of peripheral interaction, but does not address its full scope (e.g., interaction in the attentional periphery) – made use of gestural input (e.g., [20]). In addition to related work on freehand gestures, there exists work on touch-based input in parallel to the classic desktop setup (e.g., [10,21]), which does not explicitly address the characteristics of peripheral interaction. However, findings from these works encourage us to apply touch and freehand interaction to peripheral interaction.

2.2 User Interfaces for Audio Player Control

Playing music with instruments or digital systems or devices (such as, for example the reacTable* [19]) is usually a focused task. In contrast, monitoring and controlling music is an everyday task that is often performed as a secondary task. As we are particularly interested in the latter we focus on audio player control in this section. A large number of prototypical audio player interfaces have been proposed and many of them rely on the input modalities we want to evaluate for peripheral interaction:

Graspable: Graspable interfaces for audio player control can mostly be considered cube- or knob-based designs: The Gesture Cube [22] detects the movements of the user's hand and controls parameters such as volume or play/pause. The Music Cube [1] further presents a button that can be pressed and rotated. Ferscha et al. [14] present a 'key knob', which allows both continuous and discrete control of a music application. Andersen [3] also compares a knob to media keys and found that the knob was preferred over the keys. Butz et al. [11] present additional tangible audio player UIs, some of which are knob- or cube-like or meant to blend into the periphery.

Touch: On mobile touch-based devices, sweeping and tapping the screen can let the user control a music player without looking at the UI [27]. Döring et al. [12] adapted this interaction style for cars on a touch-sensitive steering wheel and reduced visual demand during interaction. Further, several commercial products now use gestural input on touch devices to remotely control a music application on a PC[1].

*Freehand***:** Strachan et al. [29] control a music player by placing a device near different body parts. This gestural input method utilizes inertial sensing and pattern recognition. Other solutions use sensor-equipped gloves and interpret predefined hand movements [9]. Additionally, the accelerometer in mobile devices is used to track geometric forms drawn in midair, which are mapped to commands [23].

3 Designing the Modalities for a Peripheral Music Controller

We selected the specific interaction sets in a two-step process: We analyzed existing systems using graspable interfaces, touch and freehand gestures, extracted interaction mappings and used them in a paper prototype study to identify the most suitable ones.

3.1 Interaction Possibilities

There are well-established symbols (▶ ⏸ ⏮ ⏭) representing the most common functions of audio equipment and media software. However, they cannot be used here, because peripheral interaction is supposed to work with minimal visual attention and gestural UIs provide no obvious place for graphical symbols. In the absence of such graphical hints, interaction sets have to be easy to use and to remember. Table 1 proposes possible gestures based on the previously mentioned related work or direct metaphors such as the volume knob. Similar to volume knobs and sliders, we want volume control to be continuous (i.e., the volume gesture continuously changes the volume to the desired level). The other commands (pause/play and next/previous) require a discrete gesture to have an effect (e.g., stop the music).

3.2 Study with Paper-Based Prototypes

As a next step in the design process, we presented all possible input mappings, listed in Table 1, to participants in a study based on paper prototypes to identify the most accepted ones and thereby ensure easy and natural interaction for all peripheral devices in the long-term deployment.

[1] `www.iospirit.com/products/remotebuddy/ajaxremote/#gesturedemovideo` (accessed: 06.12.12)

Table 1. Possible input mappings, and the finally selected one (bold and shown in Figure 3). Some gestures (e.g., one/two finger gestures) were presented individually but are summarized here.

	Graspable	Touch	Freehand
= ▲	· **click** · double click · long click	· **one/two finger tap** · one/two finger double tap · one/two finger long tap · draw square/arrow	· **hold vertical hand in mid-air**/thumbs up · move horizontal hand up/down · draw square/arrow
⏫ ⏬	· **tilt knob left/right** · turn knob left/right · push down & turn knob left/right	· **one/two finger swipe left/right** · tap left/right area on the surface	· **flick left/right** · thumb left/right
▼	· **turn knob left/right** · push down & turn knob left/right · tilt knob up/down	· **one/two finger swipe up/down** · circle left/right (cf. iPod click wheel) · two finger circle left/right (cf. rotation on toch displays) · tap top/bottom area repetitive/long	· **move horizontal hand up/down** · grasp gesture up/down or left/right (cf. slider) · circle with hand/finger left/right · thumbs up/down · pinch gesture

Paper Prototypes. To mimic the different modalities, we built paper prototypes out of white cardboard [32] (see Figure 2) resembling existing artifacts (e.g. touch pads). To symbolize a 3D gesture-tracking device we used a box (HxWxD: 6x8x5cm). For the touch sensitive surface, we used a tilted paper surface (HxWxD: 0-2x13x13cm). Looking at existing graspable systems to control an audio player, two shapes are well established in related work: the cube and the knob. We chose a knob mounted on a pedestal for our design (cf. knobs for volume control on stereos). The graspable paper interface was mounted on a bottom panel (20x20cm for stability) with the help of a brass fastener. The knob (Height x Diameter: 5.5x4cm) could be turned, tilted and pressed down.

Fig. 2. Paper prototypes mimicking a knob, a gesture tracking device and a touch sensitive surface.

Participants. We recruited 18 participants (six female) ranging in age from 21 to 31 (avg. 25 years). 14 participants had a background in computer science or math, while four participants had a background in economics or languages.

Procedure. We used a repeated measures design. The independent variables were modality (graspable, touch, freehand), command (pause/play, next/previous, volume) and gestures. Modality was counterbalanced. We presented 35 gestures to each user. Within one modality, the gestures were presented in a randomized order for each command. Participants were free to use their preferred hand for the interaction, but the paper prototype was located in the periphery (i.e., not directly in front of them). Further, participants were instructed to minimize gazing at the prototype and imagine working on another primary task simultaneously. Every gesture was first presented and carried out by the instructor and afterwards carried out by the participants themselves. Participants rated each gesture as “bad”, “ok”, or “good”. Additionally, they had to choose their preferred gesture for each modality and command. Afterwards, they were asked to answer a short questionnaire. Each run took about 30 minutes.

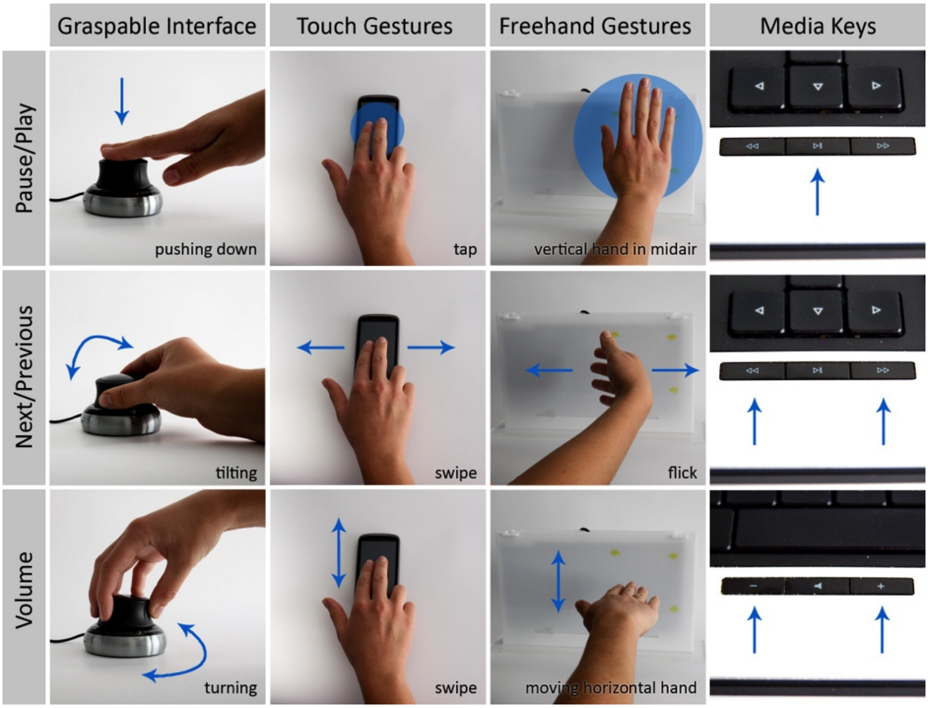


Fig. 3. Implemented interaction sets for each modality to control an audio player (pause/play, next/previous, volume). The media keys are located on the keyboard below the arrow keys (pause/play; next/previous) and the spacebar (volume).

Results. The most preferred and hence implemented gestures are highlighted in bold in Table 1 and shown in Figure 3.

Graspable: To start or stop a song, participants preferred a click (89%). Most participants wanted to switch to the next/previous song by tilting the knob left or right (83%). To change the volume participants wanted to turn the graspable knob left and right similar to a volume knob on a stereo (67%).

Touch: Using touch to start/stop the player, participants were undecided whether a one-finger tap (44%) or a two-finger tap (11%) felt more suitable. Some also mentioned that they liked both equally (22%). Generally participants were in favor of a tap (77%). We observed similar results for next/previous. A one- as well as a two-finger swipe left/right was preferred (94%). Similarly participants preferred a one- and two-finger swipe up/down (as on a mixer console) to change the volume (67%). We therefore decided to support both one- and two-finger interactions for all three commands as Wobbrock et al. also reported similar results [33].

Freehand: Participants preferred the vertically oriented hand to pause the music (67%). We had originally intended to use a thumbs-up gesture as the counterpart for play, but this gesture was not well received. Participants suggested using the vertical hand also for play, which we implemented as a tap of the vertical hand (cf. Figure 3). To change the song (previous/next), participants favored a flick gesture (78%). Half of the participants preferred an up/down movement of the horizontally oriented hand to change the volume (50%).

We further asked our participants if they could imagine using a separate device for audio control at home. On a 5-point Likert scale (1 = totally disagree; 5 = totally agree), users rated their willingness with a median of 4. They also expected it to make interaction with the audio player easier (median=4) and that they would enjoy it (median=4). Eight participants (44%) preferred freehand gestures, seven (39%) preferred touch and three users (17%) the graspable interface.

4 Building the Peripheral Music Controller

The prototype consists of the hardware for the respective input modalities and the software connection to an audio player.

4.1 Hardware

Depending on the modality, different hardware was chosen (cf. Figure 3). We decided to use commercial products wherever possible, in order to ensure a reliable function during the in-situ evaluation.

Graspable: To implement the graspable interface, we used the Connexion Space-Navigator, which is originally designed as a 3D mouse. (We had also considered the Griffin Technology PowerMate, but ruled it out because it could not implement all the gestures selected in the pre-study.) The hardware offers six degrees of freedom and an SDK. Pushing down the knob controls the play/pause function, next/previous corresponds to a lateral push or tilt, and volume to a rotation of the knob. Only volume uses a continuous manipulation, meaning that every 60ms the volume is decreased or increased as long as the user rotates the device. All three commands included a threshold to avoid accidental executions.

Touch: For the implementation of the touch modality, several hardware options were tested. The Apple Magic Trackpad, which we originally had in mind, did not offer raw data access under Windows and no generic driver, which rendered it unusable for our purpose. The same is true for the Logitech Wireless Touchpad and the Bamboo Pen&Touch. Instead, we used the touch-sensitive surface of an Android Nexus One mobile phone. The screen was set to black without any GUI elements and the device never went to sleep. Participants were asked to keep the mobile phone connected to their computer via USB for power supply. The phone communicated with the computer through UDP. A threshold was used for swiping up/down (volume) and left/right (previous/next) to avoid interference with the tap gesture for pause/play.

Freehand: Freehand tracking is based on a preproduction prototype[2] of a capacitive sensing device. We opted for this prototype because it can easily be put on a desk and detects gestures within a range of 10cm, which is perfectly suitable for our use case because this limits the interaction space and reduces the danger of unintended gestures. Using a Kinect, we would not have achieved this easy setup (i.e., small distance to tracking device). The prototype consists of five capacitive sensors located in the four corners and in the middle. X, Y and Z coordinates are provided for the location of the intruding hand's center in the electrical field. The API already recognized flick gestures (next/previous) as well as tap gestures (pause/play). Volume control was implemented as a movement along the Y-axis (upwards axis). The box (HxWxD: 19x31x3.5cm) with the tracking device offered a tracking area of 7.5x10cm (HxW).

Media Keys: We used the Cherry EASYHUB MultiMedia Keyboard, which offers grouped media keys below the regular keys. The volume control is located on the left, below the space bar. On the right, below the arrow keys, are pause/play and next/previous. Media keys are activated without additional keys (e.g. function key).

4.2 Software

We chose Windows as the prototype platform. The software, implemented in C#, connects the different hardware prototypes with the audio player. The GUI, used only for closing the application, offered a status icon in the taskbar, thereby providing feedback that the application is currently running. We selected iTunes as the audio player for implementation reasons (e.g., logging mouse interactions is possible).

5 In-Situ Deployment

Peripheral interaction is intended to blend into the periphery of the user, similarly to ambient information systems. Consequently, a long-term in-situ deployment is the preferred way to explore such a system [18]. All reported medians are based on 5-point Likert scales ranging from 1 = I totally disagree to 5 = I totally agree.

5.1 Participants and Study Environment

We carried out our exploration with eight participants (two female) ranging in age from 21 to 32 (avg. 25). All participants used Windows and iTunes before this study. Six of them were students of computer science and two working (computer scientist and engineer). Three participants had a computer setup, which provided media keys, but two of them had to additionally press the function-key to activate the media keys.

[2] The manufacturer prefers to stay anonymous for strategic reasons.

Another two participants had previously had media keys and missed them on their current keyboard. Additionally, four participants had keys for volume control only. All participants were right-handed.

During the study all devices were used at home. Consequently, their primary tasks were very diverse ranging from leisure usage of the computer (e.g., browsing the web, writing emails) to working (e.g., for university). Participants were free to use the computer in any way. Based on their backgrounds, all participants spent much time at their PC, which was helpful to get comprehensive insights of usage with the peripheral modalities alongside a standard desktop computer during the duration of the deployment. Leaving the PC while music was playing of course also was possible.

Participants' Music Listening Habits. Seven participants stated that they listen to music on a daily basis for one to four hours. The other participant listens to music three to four days a week. Our participants stated that they listen to music for half of the time (51%) while at their computer. They tend to select a playlist or an album and then listen to it but skip songs they dislike (median=4.5). They rated their interaction frequency with the audio player as medium (median=3). The most used command was pause/play (median=3.5), followed by volume control (median=3) and next/previous (median=2.5). They wish for a faster interaction with the iTunes player (median=4). Further, they are bothered by the focus switch to iTunes (median=4) and consider the distraction of it high (median=4). Apart from the focus switch, interaction with the mouse and the GUI is not considered to be mentally demanding (median=1).

5.2 Procedure

The eight-week in-situ deployment contained five semi-structured interviews with each participant. Every participant tested each modality for two weeks (repeated measures design). The distribution of devices was counterbalanced, i.e., we used a Latin square design to minimize learning effects.

1st Meeting. During the 1st meeting, we collected data about the participants' listening habits, usual interaction with their audio player and demographics. We installed the application, set it to auto start and asked them to leave it open whenever their computer was turned on. They were handed their first device and received an instruction about the gestures. At this point, participants did not know which other devices they would be testing in the upcoming weeks.

2nd to 4th Meeting. During the 2nd to 4th meeting, which took place in two-week intervals, participants were asked about their experiences with the device they had at home. We asked about the general usage as well as their perceived mental load and the ability to carry out interaction in the periphery. They handed back the old device and received the next one together with the respective instructions.

5th Meeting. During the last meeting, participants were interviewed about the last device. Additionally, we carried out a comparative interview about all devices. In the end, the application was uninstalled and the last device was given back to us.

Logging. For all modalities, we logged: (1) when and which command (pause/play, next/previous, volume) was executed, (2) which input modality (peripheral device, mouse, media keys) was used, and (3) whether iTunes was in focus while carrying out a command. Furthermore, we logged (4) the duration that iTunes was opened and music was played.

Conditions and Variables. All participants tested all four conditions: Peripheral devices (graspable, touch and freehand) and media keys. Based on our logging data, our dependent variables are number of commands, type of command, input modality used, focus of iTunes, duration of iTunes opened and duration of music listening.

6 Results of the In-Situ Deployment

During the deployment, we collected quantitative and qualitative data. A listening session is defined as listening to music without a break longer than 30 minutes [34].

6.1 Duration of Each Modality

When conducting a long-term evaluation over eight weeks, unexpected events invariably occur. One participant went on a spontaneous holiday. We therefore excluded the corresponding data from the quantitative evaluation. Another participant unexpectedly changed the workplace and only spent the weekends at home for the last four weeks of the study. Furthermore, we cannot control how often participants actually listen to music. Each modality was tested on average for the following number of days by each participant: graspable: 14.4, touch: 13.1, freehand: 12.8, and media keys: 15.1.

We logged 391 listening sessions overall. These added up to 12 days, 5 hours, 32 minutes and 2 seconds. An average session lasted 45:03 minutes. The longest listening session we logged lasted 6:48:40 hours. Our participants listened to 5652 songs (average 14.5 songs per session) and they executed 6119 commands (peripheral devices, media keys and mouse).

6.2 Frequency of Use

Looking at the probability of a command being issued during one minute for each respective modality, the highest probability is observed for freehand gestures (m=93%) (probably some of these logged interactions are due to tracking issues of freehand interaction, which is still not as reliable as other input modalities), followed by touch gestures (m=72%) and graspable interaction (m=51%) (see Figure 4). Media keys have the lowest probability (m=20%). Overall we did not observe any significant change between each of the two weeks the participants

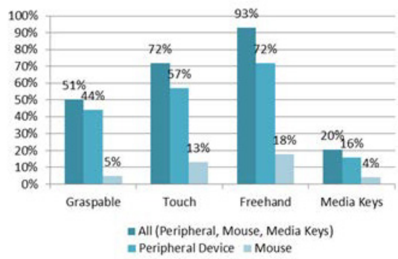


Fig. 4. Probability of issuing a command for all available input modalities (peripheral device, mouse and if available media keys) during one minute

used one device. In other words, usages slightly differed between the two weeks for each participant but sometimes they used the peripheral device more, sometimes they used it less during the second week. We assume this is due to external circumstances occurring during an in-situ deployment but not related to the peripheral devices.

6.3 Modalities vs. Mouse

Participants were given a peripheral device but of course they were free to use the traditional GUI with a mouse or media keys, if their standard keyboard allowed them to. Still, participants used the additional devices (with the exception of the media keys) significantly more than the traditional GUI (t-tests: graspable vs. mouse $p = 0.003$; touch vs. mouse $p = 0.003$; freehand vs. mouse $p = 0.006$).

6.4 Usage in the Periphery

All devices were intended to be used in the (visual and attentional) periphery. Participants quickly familiarized themselves with the modalities (graspable: median=5; touch: median=5; freehand: median=4.5; media keys: median=4.5). However, they mentioned that the location of the media keys as well as the freehand tracking was not immediately clear. Interaction sets were all rated as being intuitive (graspable: median=5; touch: median=5; freehand: median=4.5), but media keys were given a medium rating (median=3), because they required a targeted key press, even though the keyboard layout already emphasized the media keys by their position. Interaction only caused minimal mental load for all peripheral devices (graspable/freehand/touch: median=5), but participants experience more mental load for media keys (median=2.5). When asked whether disruption from the primary task, while interacting with the devices, was low, the graspable device was rated best (median=4.5), while the other modalities were rated nearly indifferent (touch: median=3.5; freehand: median=3.5; media keys: median=3)

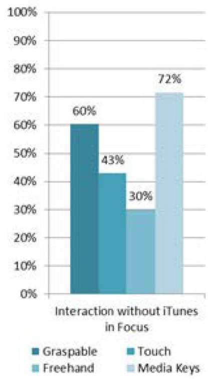


Fig. 5. Percentage of all commands carried out without iTunes in focus (not possible with the mouse)

6.5 Focus

Looking at the logged data depicted in Figure 5, we can observe that participants were able to carry out the interaction without iTunes in focus with all modalities. Using the graspable interface or media keys, participants even carried out more commands without focus than with focus. Participants stated that they felt that they were able to carry out the commands without looking at iTunes when using the graspable (median=5) and touch (median=4). In contrast, they stated to prefer looking at iTunes when using the media keys (median=2.5, which contradicts the logged data) and freehand gestures (median=2.5). Concerning the latter, they stated that they miss the haptic feedback, but some also said that it also felt “magical” and that they enjoyed observing the change in iTunes (e.g., the movement of the volume slider). This was not intended by us, especially since we wanted to shift the device to the periphery of the attention, but the novelty of the interaction seemed to have pushed the participants to look at the visual feedback.

6.6 Feedback

Due to the nature of our use case, functional feedback [31] was given. Users could hear whether a song stopped, started to play, was skipped or how the volume changed. Thus, participants stated that they did not miss further feedback for most devices (graspable: median=5; media keys: median=5; touch: median=4.5; freehand: median=3.5). Most participants (88%) named freehand gestures as the modality that offered the least feedback because of a lack of physical contact. Asked about the device with the most feedback, six participants (75%) named the graspable interface, because they could feel its physicality.

6.7 Location of Devices

Participants (all right-handed) were free to position the peripheral device on their desk wherever they felt comfortable. Only the media keys had their fixed locations on the keyboard. We asked them where they put each device while they had it at home.

Graspable: Five participants located the graspable interface on the right side of the keyboard. Two participants placed it on the left side and one participant changed the location several times. All participants interacted with the corresponding hand.

Touch: Three participants positioned the touch-sensitive device on the right. Three positioned it on the left. The respective hand was used for interaction. Two participants never used it on the desk, but carried it around, because the cable was only necessary for power supply and not for transmission. They interacted with the right hand. Other two participants also stated that they enjoyed the ability to take the device with them, even though we did not intend this use case.

Freehand: Seven participants positioned the freehand gesture tracking on the right side. With the exception of one, who used the left hand, they also interacted with the corresponding right hand. One participant changed the location frequently and thereby also the interacting hand.

Media Keys: Media keys were located on the bottom of the keyboard. Participants mostly used their right hand for interaction, but occasionally also used the left hand.

6.8 General Remarks

Generally, our participants considered the extension to be positive (median=4), especially because they were able to control iTunes independently of the current system status (median=5). They considered the peripheral input fast.

Participants pointed out at least some weakness for every modality. The preproduction prototype for the freehand gestures was considered to be too big. The touch device had to be correctly oriented. In future implementations, this could be overcome by including the finger orientation [30], or by tilting the device a bit, as is already common with separated track pads. Participants considered grasping the knob to be more difficult with the non-dominant hand than touch or freehand interaction. Media keys ask for a very precise interaction. On the other hand, two participants mentioned that they liked media keys because they were integrated and no extra device was necessary. Others enjoyed the possibility of using touch remotely. Furthermore, they liked freehand interaction because it was considered futuristic and playful. The graspable knob was described as ergonomic and similar to the known mouse. Overall, our participants could imagine using a peripheral device for further commands like opening the player, activating shuffle or even other applications.

7 Interpretation and Discussion

In the following we will present interpretations of the results and propose considerations for future projects incorporating peripheral interactions and their evaluation.

7.1 Alternative Peripheral Interaction

Table 2 summarizes the results for the three peripheral devices (one 'plus' being a medium score here). This table confirms previous research based on graspable interfaces, but also shows that touch and freehand gestures are an alternative. This is the case especially when a graspable device is unsuitable, for instance, because there is no space to place it. Some current findings, like the missing haptic feedback for freehand interaction, might be less problematic once users are more accustomed to freehand interfaces and trust them more.

Table 2. Characterization of all peripheral modalities. One plus here is a medium score.

		Graspable	Touch	Freehand
Usage:	**Preference over mouse & GUI**	+++	+++	+++
	Little mental load	+++	+++	+++
Periphery:	**Interaction without focus on GUI**	+++	++	+
	Inherent haptic feedback	+++	++	+
Learnability:	**Easy familiarization**	+++	+++	++
Location:	**Preference for dominant hand interaction**	++	++	++

7.2 Usage of Peripheral Devices

Participants issued fewer commands (including commands carried out with the mouse) during the media keys condition. We see two possible explanations for this observation. First, the peripheral devices (graspable, touch, freehand) could have been more innovative for our participants and therefore triggered their play instinct, second, subjective data shows that media keys imposed more mental load on the participants, which might have kept them from interacting with the player although they would have changed the song if interaction was more comfortable. This is also in line with subjective findings where participants told us that they would like to have a faster interaction for iTunes and consider the distraction imposed through window switches as high. Consequently, we attribute the increased interaction with the peripheral devices to an easier and smoother interaction and think that it indicates that peripheral interaction can lower the barrier to start interaction. While a low barrier for interaction might not be extremely relevant for audio player control it might help for tasks that participants should care about but do not particularly enjoy (e.g., keeping their availability status up to date for colleagues in an instant messaging client).

For all peripheral modalities, we can observe significantly more interaction with the peripheral device than with the mouse. While we acknowledge that this might be partly due to the novelty effect and the fact that participants knew their interaction was logged and tried to (maybe unconsciously) please us, we still consider this a tendency towards a preference for peripheral interaction in contrast to traditional mouse interaction in this use case. This is supported by general research in peripheral interaction [17] and the fact that this effect arose in all modalities (graspable, touch, freehand) also indicates that the design space for peripheral interaction can be extended beyond graspable interfaces.

7.3 Shift to the Periphery

Measuring how much a device or interaction style moves into the periphery is difficult and peripheral interaction researchers still struggle with it. Although in-situ deployments are feasible because they offer the necessary time for the users to get accustomed to the device, they also limit researchers in their observations. For example, it would not be viable to equip participants with an eye tracker in their home for several weeks to evaluate the visual periphery. As a weaker indicator of visual attention, we therefore used the input focus of the GUI.

We observed that participants carried out commands without the focus on iTunes with every modality, which strengthens our belief that the design space for peripheral interaction can be extended and indeed interaction in parallel to other tasks is possible. Furthermore, subjective results show that the graspable and touch modality were especially easy to be carried out in parallel. Physical feedback therefore seems to be helpful even (or particularly) when interacting with minimal attention in the periphery. In our use case functional feedback provided by the task itself (e.g., music started or stopped to play) also supported peripheral interaction.

7.4 Learnability

Usually when designing systems for peripheral interaction the interaction itself is very easy and straightforward (e.g. swiping up and down the touch surface or carrying out a flick gesture) and does not really have to be learned. Our participants also familiarized themselves with each modality very quickly. On the other hand, when evaluating peripheral interaction, all prototypes always require time to get used to and to shift to the periphery [16]. Interestingly, our study suggests that it is not necessarily only the device itself that the user needs to get accustomed to (especially when they are generally used to an interaction style such as touch input), but also the general usage of the periphery alongside traditional input. One participant mentioned that he had more trust in peripheral interaction without looking at the GUI after he had been testing several devices. Another participant told us that he first had to break loose of habits like using alt+tab to switch to iTunes.

7.5 Location of Devices

Up to now, previous work on peripheral interaction did not consider the question which hand was used to carry out the peripheral interaction. Generally, we observed a tendency towards positioning and interacting on the right side, which was the dominant hand for all of our participants. This is an interesting observation because apart from a possible preference for the dominant hand, one can argue for either direction when considering a traditional desk setting. On the one hand, most users are used to moving the right hand away from the keyboard to interact with the mouse, and therefore might also feel more comfortable interacting with the peripheral device with their right hand. On the other hand, the mouse often occupies the right hand while the left hand would be free for other tasks. However, the latter did affect the decision where to put the device less than the preference for the dominant hand. For future systems it is feasible to offer peripheral interaction for the dominant hand if the primary tasks allows.

7.6 Paper Prototyping vs. Field Study

The paper prototyping study was very successful in terms of finding suitable interaction sets for all devices. Participants considered all gestures to be very intuitive and natural. At the end of the paper prototyping study, we asked the participants which modality they preferred. The most mentioned modality were freehand gestures, followed by touch gestures and graspable interaction. The participants of the field evaluation answered the same question after the eight weeks and the most mentioned device was the graspable interface followed by touch and freehand gestures. This clearly hints that while paper was great to test the general gestures, the preferences differed very much in the end. Possible reasons are that (1) technical issues (e.g. freehand gestures being the least reliable input modality) were not present with the paper prototypes, (2) haptic issues, which were mentioned several times, did not play such a big role when touching just paper and (3) the missing interactivity when using paper. Especially since most devices incorporating peripheral interaction are tested in the field, paper prototypes here have proven to be a good tool to test interaction mappings intended for the periphery even though during a paper study the periphery cannot be mimicked properly.

8 Limitations

There are some limitations inherent in the nature of our study. The biggest problem turned out to be the fact that the three peripheral devices used hardware in different development stages. The graspable knob is a commercial product with good haptic and ergonomic characteristics. The touch device was based on a mobile phone, which evokes different ideas than a track pad. The freehand tracking was based on a preproduction prototype, which had worse recognition rates than the other two devices for some commands (e.g., flicks (next/previous) were already very well recognized). Consequently, freehand tracking also resulted in more incorrectly interpreted gestures. This might also explain the higher frequency of freehand gestures compared to graspable interaction. We assume that this affected subjective ratings and favored the graspable interface. In other words, although touch and freehand gestures already seem suitable for peripheral interaction, in this exploration they might even be more suitable if implementation quality increases.

Further we only tested this in one use case (audio player control) with very distinct features (e.g. only input was evaluated, output and feedback was inherent to the task, i.e., music started to play). Furthermore we used one example implementation for each modality – thoroughly chosen based on a paper prototyping study – but of course these example do not completely cover the capabilities of each modality.

Additionally, unexpected events like holidays and job changes affected the individual times spent with each modality. Two participants also mentioned that they would have liked to have more time with each device to fully embrace it.

9 Conclusion and Future Work

We presented an eight-week in-situ deployment evaluating different modalities (graspable, touch, freehand) for peripheral interaction. We contribute a first exploration in the usage of touch and freehand gestures as a means for peripheral interaction. Our research confirms previous work relying on graspable devices but shows that similar results – smooth interaction intensively used without focus on the audio player itself – could be achieved with touch-based interaction as well as freehand gestures. This broadens the design space for peripheral interaction making room for other use cases, where graspable interaction might not be fit (e.g. in settings with high hygienic demands).

We further found that participants had a preference for the dominant hand, although the interaction itself (e.g. a flick gestures) could easily be performed with the non-dominant hand. This has not been addressed yet in peripheral interaction research but can act as guidance for future manual peripheral interaction.

Last but not least, for evaluation of peripheral interaction one can observe a tendency towards field evaluation. We managed to find suitable interactions by means of a paper prototyping study even though we were not able to mimic interaction in the periphery properly. By utilizing paper prototypes or similar techniques in the early design process, only actual promising ideas need to undergo a field trial reducing the resources needed overall.

Related work already shows diverse applications areas for peripheral interaction. Further research in other, especially non-desk-based, scenarios is necessary. We therefore aim at further case studies to substantiate the promising results of this study. Additionally, as participants stated that they do not want an extra device for each peripheral task, we imagine adapting peripheral touch input for touch pads, as well as freehand interaction based on camera tracking with the help of built-in webcams. Thus, the reduction of additional hardware and the trend towards miniaturization of tracking devices (e.g. Leap Motion) will help to further spread the use of beneficial peripheral interaction.

Acknowledgments. This work is funded by the German state of Bavaria. We thank Sebastian Boring, Sebastian Löhmann, Sarah Tausch, and Michael Winking for their valuable input during this project. Further we thank everyone participating in our paper prototyping study as well as the field deployment.

References

1. Alonso, M.B., Keyson, D.V.: MusicCube: making digital music tangible. In: CHI EA, pp. 1176–1179 (2005)
2. Altmann, E.M., Trafton, G.J.: Task Interruption: Resumption Lag and the Role of Cues. In: CogSci (2004)
3. Andersen, T.H.: Searching for music: How feedback and input-control change the way we search. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 144–157. Springer, Heidelberg (2005)
4. Bailey, B.P., Konstan, J.A.: On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. Computers in Human Behavior 22(4), 685–708 (2006)
5. Bakker, S., van den Hoven, E., Eggen, B.: Design for the Periphery. In: Eurohaptics, pp. 71–80 (2010)
6. Bakker, S., Hoven, E.: Acting by hand: Informing interaction design for the periphery of people's attention. Interacting with Computers, 119–130 (2012)
7. Bakker, S., Hoven, E., van den Eggen, B., Overbeeke, K.: Exploring peripheral interaction design for primary school teachers. In: TEI, pp. 245–252 (2012)
8. Bakker, S., van den Hoven, E., Eggen, B.: FireFlies: Supporting Primary School Teachers through Open-Ended Interaction Design. In: OzCHI, pp. 26–29 (2012)
9. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. Communications of the ACM 36(7), 28–35 (1993)
10. Bi, X., Grossman, T., Matejka, J., Fitzmaurice, G.: Magic Desk: bringing multi-touch surfaces into desktop work. In: CHI, pp. 2511–2520 (2011)
11. Butz, A., Schmitz, M., Krüger, A., Hullmann, H.: Tangible UIs for Media Control - Probes Into the Design Space. In: CHI Extended Abstracts (Design Expo), pp. 957–971 (2005)
12. Döring, T., Kern, D., Marshall, P., Pfeiffer, M., Schöning, J., Gruhn, V., Schmidt, A.: Gestural Interaction on the Steering Wheel – Reducing the Visual Demand. In: CHI, pp. 483–492 (2011)

13. Edge, D., Blackwell, A.F.: Peripheral tangible interaction by analytic design. In: TEI, pp. 69–76 (2009)
14. Ferscha, A., Holzmann, C., Resmerita, S.: The Key Knob. In: ICDCS Workshops (2006)
15. Hausen, D., Boring, S., Polleti, J., Butz, A.: Exploring Design and Combination of Ambient Information and Peripheral Interaction. DIS Work in Progress (2012)
16. Hausen, D.: Reducing Cognitive Load by Using the Priphery of our Attention. Workshop Designing for Cognitive Limitations. In: Workshop Designing for Cognitive Limitations. In Conjunction with DIS (2012)
17. Hausen, D., Boring, S., Lueling, C., Rodestock, S., Butz, A.: StaTube: Facilitating State Management in Instant Messaging Systems. In: TEI, pp. 283–290 (2012)
18. Hazlewood, W.R., Stolterman, E., Connelly, K.: Issues in Evaluating Ambient Displays in the Wild: Two Case Studies. In: CHI, pp. 877–886 (2011)
19. Jordà, S., Kaltenbrunner, M., Geiger, G., Bencina, R.: The reactable*. In: ICMC (2005)
20. Kajastila, R.A., Lokki, T.: A Gesture-based and eyes-free control method for mobile devices. In: CHI EA, pp. 3559–3564 (2009)
21. Kane, S.K., Avrahami, D., Wobbrock, J.O., Harrison, B., Rea, A.D., Philipose, M., LaMarca, A.: Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In: UIST, pp. 129–138 (2009)
22. Kranz, M., Freund, S., Holleis, P., Schmidt, A., Arndt, H.: Developing Gestural Input. In: ICDCSW (2006)
23. Mäntyjärvi, J., Kela, J., Korpipää, P., Kallio, S.: Enabling fast and effortless customisation in accelerometer based gesture interaction. In: MUM, pp. 25–31 (2004)
24. Mark, G., Gudith, D., Klocke, U.: The Cost of Interrupted Work: More Speed and Stress. In: CHI, pp. 107–110 (2008)
25. McFarlane, D.: Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. Human-Computer Interaction 17, 63–139 (2002)
26. Olivera, F., García-Herranz, M., Haya, P.A., Llinás, P.: Do Not Disturb: Physical Interfaces for Parallel Peripheral Interactions. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 479–486. Springer, Heidelberg (2011)
27. Pirhonen, A., Brewster, S., Holguin, C.: Gestural and audio metaphors as a means of control for mobile devices. In: CHI, pp. 291–298 (2002)
28. Pousman, Z., Stasko, J.: A Taxonomy of Ambient Information Systems: Four Patterns of Design. In: AVI, pp. 67–74 (2006)
29. Strachan, S., Murray-Smith, R., O'Modhrain, S.: BodySpace: inferring body pose for natural control of a music player. In: CHI EA, pp. 2001–2006 (2007)
30. Wang, F., Cao, X., Ren, X., Irani, P.: Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In: UIST, pp. 23–32 (2009)
31. Wensveen, S., Djajadiningrat, J.: Interaction frogger: a design framework to couple action and function through feedback and feedforward. In: DIS, pp. 177–184 (2004)
32. Wiethoff, A., Schneider, H., Rohs, M., Butz, A., Greenberg, S.: Sketch-a-TUI: Low Cost Prototyping of Tangible Interactions Using Cardboard and Conductive Ink. In: TEI, pp. 309–312 (2012)
33. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: CHI, pp. 1083–1092 (2009)
34. Zheleva, E., Guiver, J., Mendes Rodrigues, E., Milić-Frayling, N.: Statistical models of music-listening sessions in social media. In: WWW, pp. 1019–1028 (2010)

Linetic: Technical, Usability and Aesthetic Implications of a Ferrofluid-Based Organic User Interface

Jeffrey Tzu Kwan Valino Koh, Kasun Karunanayaka, and Ryohei Nakatsu

Keio NUS CUTE Center, Interactive and Digital Media Institute,
National University of Singapore, Singapore
jtkvkoh@gmail.com, kasun@mixedrealitylab.org, elenr@nus.edu.sg

Abstract. We present an OUI that combines Hall Effect sensing and actuation through electromagnetically-manipulated ferrofluid. The movement of magnets worn on the fingertips, over a surface embedded with a Hall Effect sensor array and electromagnets, gives the user the ability to interact with ferrofluid. This system provides a three-dimensional, physically animated response, as well as three-dimensional, spatial-sensing inputs. The vibration of the magnets worn on the fingertips, produced by the repulsing polarity of the electromagnets, provides the user with haptic feedback. Linetic is a multimodal interface with a visual, audio and haptic experience. In this manuscript we explain the overall system from a technical, usability and aesthetic viewpoint by outlining significant experiments conducted that contribute to the development of the system. Furthermore we discuss the philosophical and aesthetical implications of the Linetic system, as well as characterize Linetic's disposition to Analogness or Digitalness.

Keywords: Electromagnetic, haptic, TUI, OUI, morphable, shape changing.

1 Introduction

Recent progress in human-computer interaction has paved the way for a new body of research known as Organic User Interfaces (OUIs) [2] [10]. This field focuses on the need to further explore possibilities for interactive user interfaces using the advances made in electronics and material sciences [9]. OUIs are defined by three factors: the input interface and output display should be one and the same; the form of the object should change continuously and correlate directly with the function it embodies, and finally the function performed by the object depends on how the physical shape of said object is changed [22].

OUIs open a path to morph the shape and form of the actual computer interface itself. The interface can now be a piece of fabric, a plastic card, liquid, sand, clay or any other material. OUIs enable users to interact with the interface by manipulating the natural qualities of these materials such as bending, stretching, pulling, stroking, etc. This new paradigm moves away from the traditional approach of metaphors and physical objects as defined by Tangible User Interfaces (TUIs), and explores next generation interfaces focused on the analog, continuous and transitional nature of physical reality and human experiences, as Schwesig points out [20].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 180–195, 2013.

With advances in technology pushing the boundaries in regards to the materials used to create ubiquitous interactive systems [24], it is now possible to expand the computer into the everyday environment through softer, and flexible formats [1]. Using these materials and technologies, this paper presents the implementation of an innovative OUI system based on liquid. This interface explores the potential of liquids as an interface and display device, where the manipulation of liquid becomes both the input and output. Linetic (Link to youtube video - http://www.youtube.com/watch?v=V5EMyRsLLeQ) can provide the user with a natural and fluid experience where three-dimensional, tangible interaction takes place.

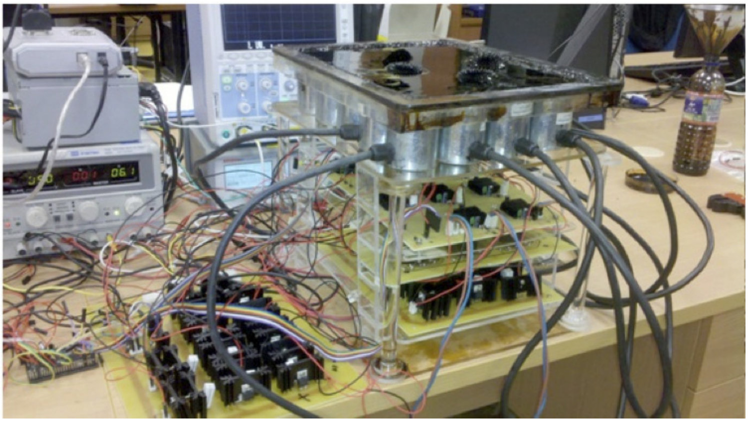


Fig. 1. Version one of the Linetic system

Building on the idea of a ferrofluid display created by Sachiko Kodama [13], Linetic provides an input/output solution based on ferrofluid. Ferrofluid is essentially a liquid that reacts to magnetic fields. The system is composed of a pool of ferromagnetic liquid combined with a sensing and actuation mechanism. The sensing uses an array of Hall Effect sensors that measure the density of a magnetic field, while actuation is produced by an array of electromagnets. Sound generation using a MAX/MSP patch running on a connected server augments the output experience.

Wearing a set of magnetic rings, the user can interact with the ferrofluid. The magnetic ring position is detected by the array of Hall Effect sensors, which in turn actuates the electromagnets and the sound server. The magnetic field of the active electromagnets produces the morphing of the ferrofluid to create transitional physical buttons in conjunction with the gesture, which then generates a sound. At the same time the pulse of the matching polarity electromagnets produces a force feedback vibration on the rings, giving the user haptic feedback.

Through natural movements of the hand, the interface is able to morph from a two-dimensional surface to a three-dimensional form fluidly and dynamically. Using the shape changing quality of ferrofluids, we were able to study how liquids could become a novel form of OUI. In sum Linetic provide a tangible, multi-touch interface with haptic feedback that produces a real 3D morphing surface.

2 Background

Linetic builds on research by both scientists and artists in creating a malleable, three-dimensional, multi-touch interface. Hiroshi Ishii's seminal research in TUIs offers insight for natural user interfaces in which this project builds upon [9]. Ishii later went on to categorize generations of TUIs, leading to the development of guidelines for the next iteration of user interfaces, known as Organic User Interfaces [8] [10] [23].

Research into the field of OUIs is still in its infancy. In 2008 ACM published a groundbreaking collection of essays that collected the works of OUI's pioneering researchers in a special issue of Communications of the ACM. In this special issue, several key authors outlined various aspects and characteristics that make an OUI:

Input Equals Output: In the graphical user interface (GUI) there is a clear division of input and output. The mouse and keyboard input actions from the user. Based on those actions, output is generated graphically on the screen. A key feature of OUIs is that a piece of Organic light-emitting diode paper, or any potentially non-planar object for that matter, is meant to input actions from the user and also outputs them onto the same object.

Function Equals Form: The form of an object clearly determines its ability to be used as an input. The statement Function Equals Form emphasizes this dependency on one another. Holman and Vertegaal [8] argue that these two are in fact inseparable and that it is a mistake to try to deny this in any way.

Form Follows Flow: This principle states that it is of utmost necessity for OUIs to negotiate user actions based on context, e.g. the ubiquitous 'clamshell' phone, where incoming calls alter the phone's function when opening the phone during an incoming call.

In regards to the material we choose to use for our input and output, the ferromagnetic art installations by Sachiko Kodama provide an aesthetic viewpoint of how the power of fluid and transient shapes can capture the imagination of viewers [13]. In her projects ferromagnetic fluids are used to create organic shapes that change structure dynamically. Adjusting the power of the electromagnets actuates the ferromagnetic fluid in varying ways. The magnetic field produced by the electromagnets controls the movement of the ferrofluid, producing a visual output. Linetic builds on this work, providing the user with a means to directly interact with the ferromagnetic fluid and the formation of shape changing buttons, while inheriting the aesthetic and kinetic qualities that Kodama's pioneering works offer.

Other ferrofluid interfaces include SnOil, which is a controllable display that allows the user to interact with ferromagnetic fluids by tilting the display [6]. The display uses a grid of electromagnets to actuate ferromagnetic fluid in selected areas, and a built-in tilt sensor is used to activate the electromagnets. Actuation of the ferrofluid is binary, off or on, depending on the respective electromagnet state. Linetic expand upon SnOil by offering a novel means of interaction with ferrofluid by directly interacting with the material and provides an analog-like and linearly variable state of the actuated fluid, beyond simply a binary on/off state.

MudTub by Tom Gerhardt [7] is an intriguing, quasi-OUI that exploits the material characteristics of mud as an input material, but fails to follow the contemporary definition of OUIs as outlined by Holman and Vertegaal [8], in the sense that the interface medium and output display are not entirely the same. MudTub uses projection to augment the output and provide input. While offering a morphable computer interaction material in the form of mud is novel, our implementation offers more fluid-like properties, a higher degree of controllability, adheres to the tenants of OUIs as outlined by Holman and Vertegaal, and is not nearly as messy.

MudPad [11] is a haptic multi-touch input device that is malleable and can be controlled for localized haptic feedback. It acts similar to Linetic but the materials and methodology of actuation is different. The interaction material is also contained within a bladder and users do not directly interact with the material. Linetic offers the user direct interaction with the display and input material, making it a more compelling example of an OUI.

Programmable Blobs [24] is a compelling example of a user interface that uses malleable, programmable matter. It differs from Linetic in that it still uses a quasi-solid material compared to the much more fluid attributes of ferrofluid in the Linetic system. Latency during transformation of the Programmable Blobs is also much slower when compared to the speed of Linetic.

All the examples discussed above provide an invaluable and important basis for the research conducted for Linetic. Linetic attempts to expand the field of TUIs beyond solid and static-shaped object manipulation, bridging the gap between physical user actuation and animated representation into a unified input/output device.

3 System Description

A system diagram is shown in Figure 2 and the first version of the Linetic system is shown in Figure 1. By moving fingers over the ferrofluid buttons, users can manipulate shapes in the liquid in order to interact with the system. The system identifies a magnetic flux change when a finger moves perpendicular and vertically in relation to the surface. The flux change is detected by the Hall Effect sensors embedded beneath the surface of the ferrofluid, which sense the magnetic field produced by the magnets worn on the fingers. In turn, an electromagnet also contained beneath the surface of the ferrofluid is activated and produces a field in which the ferrofluid reacts to, producing a ferrofluid button. The distance between the finger and the ferrofluid button is mapped to the pressure and intensity of the click. Haptic feedback is felt by the user through the natural vibrations of the magnet worn on the fingertips. In the following section we will describe both the hardware (sensing and actuation) and software design of the system in detail. In essence, the Linetic system consists of a set of controller circuits to control the flow of the current, an array of Hall Effect sensors to detect user interaction, an array of electromagnets to actuate the ferrofluid, a ferrofluid pool, a MAX/MSP sub-system for sound generation and a software application programming interface (API).

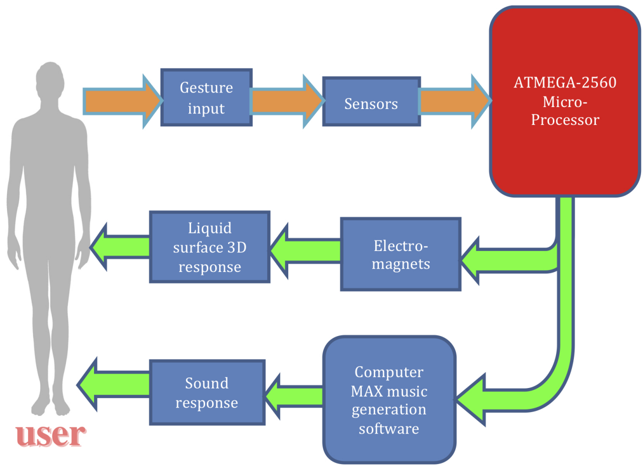


Fig. 2. A System diagram of the Linetic system

3.1 Hardware Design

In the following section, we will outline the various hardware sub-systems of the Linetic system. Hardware sub-systems include the sensing method, the actuation accessory, the electromagnetic array and circuit.

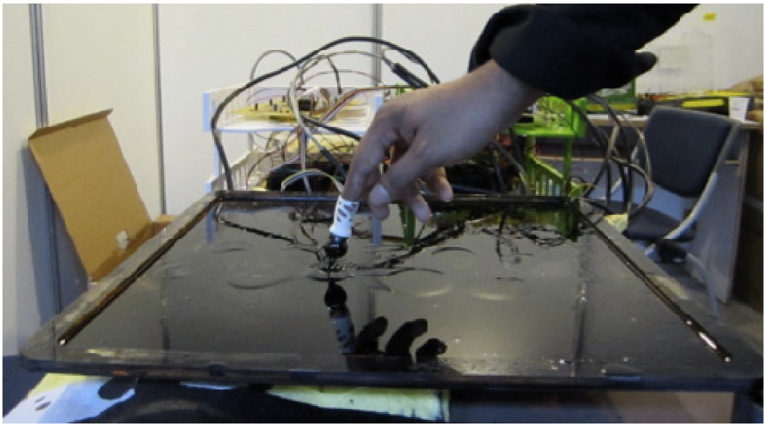


Fig. 3. Interacting with version two of the Linetic system using the wearable accessory

Sensing. The hardware for sensing is performed by magnetic Hall Effect sensors manufactured by Hamlin Electronics. Hall Effect sensors measure the density of a magnetic field. The sensor produces an analog output when it is under the influence of any magnetic field. The sensor output is connected to the analog-to-digital converter of the micro controller used. The sensor is placed directly on top of each electromagnet while the user's finger carries a powerful neodymium magnet with like poles of each magnet facing one another. When the electromagnet is turned on, the sensor output becomes fully saturated. If a neodymium magnet of the same pole is brought close to the surface, the sensor output drops and this is detected as the presence of the user's finger. Voltage drop of the sensor is dependent on the strength of the external magnetic field. Therefore, a sensor voltage drop is much higher when the distance between the finger and the ferrofluid display is reduced. This measurement is used as an input to control the height of the spikes, which are a product of the system.

Actuation Accessory. The actuation of the system, as seen in Figure 3, takes place when the user moves their fingers across the ferrofluid surface while wearing a magnetic finger accessory. To design this accessory, we looked at various accessories musicians wear on their fingers, such as guitar picks, to give us ideas on how to best design an accessory that acted as a natural extension of the finger. Achieving this kind of interaction was important for us so we decided to design a ring that allows the users to wear and play with the system. We also developed a wand that can be waved over the surface in order to provide an alternative method of interaction with the system but in this manuscript we will focus on the wearable configuration.

Fig. 4. Design iterations for the wearable component. The final design is still being developed.

We designed a ring that could be worn on the index finger of each hand. Figure 4 shows some of the design iterations. The rings contain a magnet that allows users to actuate the ferrofluid through natural gestures and finger movements. The ring has a cylindrical base made of plastic with a metal, coned shaped tip containing a magnet. When the users play with the system, they can sense a subtle haptic feedback through magnetic repulsion. This enables users to judge the distance from the liquid surface without touching it. This is important since when the user gets too close to the fluid, the liquid will become attracted to the metal cone. To avoid this, the haptic feedback enables the user to use the fluid interface without wetting or staining the fingers. Furthermore, electromagnet polarity is configured in such a way as to repel the magnet on the user's finger, once again subtly directing the user to not place her hands too close to the liquid.

Actuation. The screen consists of ferrofluid placed in a shallow acrylic container. This container is placed on top of an array of electromagnets. These electromagnets, when turned on, actuate the fluid, causing it to form bumps and/or spiky buttons depending on the strength of the magnetic field produced. The system is powered using an array of ATX power supplies. These power supplies are then fed to a current limiting circuit. The output current from the cur- rent limiter circuits is distributed to the magnet driver board consisting of full-bridge drivers. By feeding the microcontroller's pulse-width modulation (PWM) signals to the magnet drivers, the electromagnetic field produced by the electromagnet array is controlled. This electromagnetic field creates spikes in various shapes and sizes on the surface of the ferrofluid display.

3.2 Software Design

In the next section, we will outline the various software sub-systems of the Linetic system. Software sub-systems include the Linetic OUI Framework, sound generation and the circuit firmware. Software sub-systems include the Linetic OUI Framework, sound generation and the circuit firmware.

Linetic OUI Framework. One of the objectives of the research is to develop a software framework for the project. It consists of a system background service and a Java API extension. The framework hides the complexity of the overall hardware and low-level software code from the developer and provides a convenient way to plug different sub-systems into the Linetic system. Therefore the Linetic system could be configured as a tangible input device, a tangible display, or perform both operations to the external system simultaneously. Various kinds of external systems could communicate with the Linetic system through the Linetic OUI framework API.

Sound Generation. In this application we attached Linetic to an instrumental music generation program developed in the MAX/MSP environment. The Max/MSP patch that is used detects incoming signals from the framework and activates appropriate midi piano keys to create sound.

Circuit Firmware. The circuit firmware is written in C and programmed into the microcontroller within the circuit. It communicates with the PC through USB interface and receives information regarding ferrofluid actuation patterns. According to the given information, it dynamically actuates the ferrofluid and creates dynamic patterns in the ferrofluid display by changing the current of the electromagnet grid underneath the ferrofluid container.

4 Experiments and Results

The previous work of authors includes a series of experiments to measure spike height versus current, distance of two adjacent spikes, transient state of the system and the static linearity of the system [1]. These experiments demonstrated the linearity of the control mechanisms, which are used in the Linetic system. Since magnetic systems are highly nonlinear, these experiments represent a significant step in controlling ferrofluids. The following experiments explore the control mechanisms of the system.

4.1 Experiment 1: Hall Effect Sensor Reading versus Vertical Distance

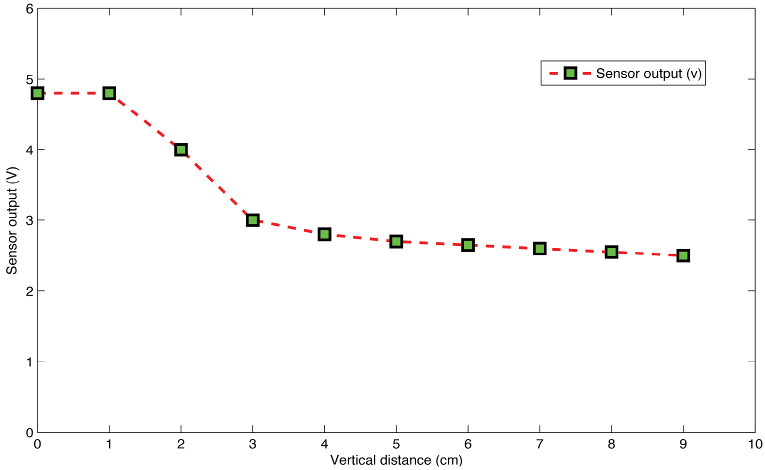


Fig. 5. Sensor output versus vertical distance

The point of this experiment is to understand the Hall Effect sensor readings in relation to the vertical distance of the actuator (in this case, the finger-worn actuation ring) from the surface. This experiment has been conducted using a Hall Effect sensor and an electromagnet, which generates an average flux density on the surface from 450 to 1950 Gauss for the range of 6V to 24V with 1.9A to 7.5A of electrical current. In the experiment we kept the power of the electromagnet at a constant voltage of 10V and a driven current of 2.44A, with the sensor on the vertical axis on top of the electromagnet. The sensor reading is measured versus the distance to the electromagnet. The value of the sensor output voltage taken is the mean value in one second.

This plot shows (Fig.5) that the sensor is most sensitive with respect to the vertical distance from 0cm to 3cm. When the distance is greater than 3cm, the change in output is much lower. At larger distances, for example the values of 6cm and 7cm, the difference in voltage is only 0.011 volts. However such a small voltage difference is not detected by the microcontroller used for this iteration of the system. Therefore, the sensor cannot detect motion from 6cm to 7cm, but it can detect longer distance motion, e.g. moving from 6cm to 8cm.

4.2 Experiment 2: Hall Effect Sensor Reading versus Horizontal Distance

This experiment is similar to this first experiment, but looks at horizontal distance as opposed to vertical distance. Once more keeping the power of the electromagnet constant, the sensor is placed on the vertical axis of the electromagnet at 2cm, since at this distance the sensor is most sensitive, registering the largest change in values with respect to the distance moved. An oscilloscope was used to measure the voltage output of the sensor.

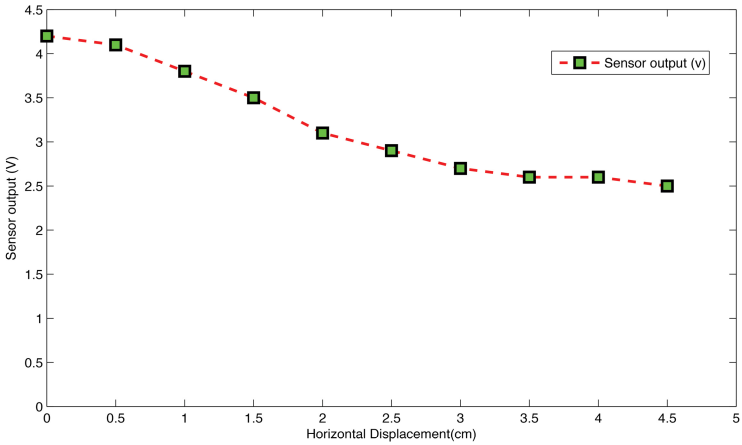


Fig. 6. Hall Effect sensor output versus horizontal displacement

Experiment 1 shows that the resolution of the system cannot distinguish any smaller changes. Within 0.02 volts, the magnetic field at 3.5cm and beyond is too small to cause a change in the microprocessor. This experiment shows that the magnetic field from the horizontal area of the magnet is too small to be detected at the optimal vertical distance. This is advantageous for the system since the tracking system gives an almost perfect horizontal reading without external noise from the neighboring magnets.

4.3 Experiment 3: Characterization of Magnetic Hall Effect Sensor Readings under the Influence of Multiple Magnetic Fields

In this experiment the readings of the Hall Effect sensor are measured to determine the influence of the magnetic fields generated by the electromagnets and neodymium magnets. These readings will provide the appropriate parameters for the microcontroller firmware.

The electromagnet is connected to a power supply, supplying 10V and up to 2A of current. Its output magnitude is controlled by a Pulse-Width Modulated (PWM) input to the driver circuit. It is positioned such that the field of the North Pole is directed vertically upwards and the South Pole vertically downwards. The sensor is supplied the rated of 5V and is positioned such that it is in level with the top of the electromagnet and directly next to it. The reading of the steady-state output voltage of the sensor is recorded using the oscilloscope, while varying the height and direction of the neodymium magnet and the PWM input to the electromagnet. First the default sensor value is taken. Without the neodymium magnet or electromagnet influence, the reading is 2.50V. Next, with the neodymium magnet pole at South Pole (facing down), the PWM values and distances are measured.

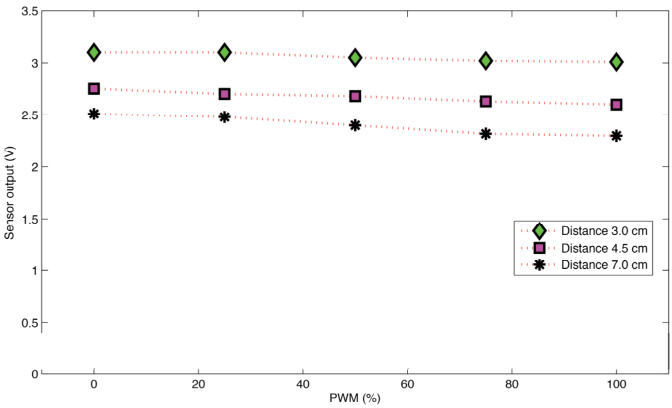


Fig. 7. Sensor reading values obtained for different distances versus PWM

Here (Fig.7) the strength of the electromagnet's field serves to decrease the reading of the sensor, whereas the position of the neodymium magnet field serves to increase the reading of the sensor. This results in a case where the value of the sensor is unable to detect the presence of the neodymium magnet due to the electromagnet's field. From the data we gathered, this occurs in the case when the distance of the neodymium magnet is 7.0cm. If the electromagnet is off, the reading is 2.53V, but if the electromagnet is turned on, the reading falls below the 2.50V neutral value. To circumvent this problem, we could use like poles instead of unlike poles, i.e. the North side of the neodymium magnet facing the North side of the electromagnet. This causes both magnets to boost each other's readings, which can be compensated for by changing the sensitivity in the software. This approach has the peripheral advantages of preventing the two magnets from attracting each other and preventing the neodymium magnet from picking up the ferrofluid contained on the surface of the system. It also provides a means for haptic feedback while using the system.

5 User Interaction Methodology

In regards to the user interaction methodology, we attempted to simulate some interactions that reflect both the aesthetics and playfulness of interacting with fluid. Although the dramatic effects of water as a tactile surface has been previously explored [19], no precedent in regards to the interaction with ferrofluid, which has distinct fluid characteristics, has been recorded. Due to this, an adaptation of previous methods as well as trial and error was employed.

Fig. 8. Version Three of the Linetic system without the splash cover

Due to the messy nature of ferrofluid as well as the nature of sensing inherent to the system, we decided to employ a finger accessory that allowed users to interact with the fluid without actually touching it. This reduced the methods of interaction to that if simple mid-air gestures, some of which are employed in [1]. These gestures included waving and tapping. Like-polarity between the surface and fingertip magnets allowed a force resistance that made tapping a particularly compelling interaction.

Later on, we decided to remove the haptic subsystem in order to concentrate on the visual and audio effects of the system. This led us to the third iteration of the Linetic system.

6 Aesthetic Interactions with Linetic

Although efficiency and speed are two important characteristics for most technical devices, research suggests that these attributes are not sufficient to fulfill all of a user's needs [4] [16]. Aside from the technical characteristics of the device, a user needs a series of other characteristics that enrich the user experience of an artifact, device or technology. People search for these characteristics that allow exploration and a holistic experience of an artifact instead of just basic features. Through Linetic, we hope to tap into these deeper desires and human emotions in order to be able to produce a more enriching, aesthetic experience of the system. The aesthetic experience refers to the pragmatic aesthetics that Dewey [3] and Shustermann [21] describe, as well as the guidelines that Petersen and others have developed for Aesthetic Interactions [18] from the pragmatist aesthetics viewpoint.

Fels [5] states that for an individual to have an aesthetic experience there needs to be an intimacy that is built between the person and the object. Fels believes that when people are able to manipulate an object skillfully or intimately, this produces an aesthetic experience. He describes four types of relationships that one is able to have with an object and these are as follows: the person communicates with the object; the person embodies the object; object communicates with person; object embodies the person. Through Linetic we are able to achieve the first three of these experiences: person communicates with object; object communicates with person; the person embodies the object.

First, the user is able to communicate with the system in a direct and natural way through touch. Interacting with the liquid, the user can start a dialogue in which the system responds to the user. The still, dark pool of ferrofluid, with its reflective surface like a black mirror, has the ability to draw people to it in the same way artist Richard Wilson [26] was able to capture the imagination of his audience with his installation 20:50 - a room filled with thick sump oil. According to Khaslavsky et al. [12], the first step to seduce your audience is to entice them in order to grab their attention. It is the very liquid quality of ferrofluid that makes it a material that probes curiosity and encourages interaction. This curiosity naturally leads to one wanting to touch the liquid and explore it.

The Linetic system possesses the playful characteristics needed for learning by playing, which is a learning mechanism for most people. The system lends itself to playful exploration and thus a relationship with the system can begin. Unlike a solid tangible interface, liquid has the ability to take the shape of its container and allows for the free maneuvering of itself. As people move their fingers above the liquid they begin to learn quickly how to sculpt and manipulate the liquid, creating three dimensional, liquid sculptures combined with sound that respond to the user's tactile exploration. This seemingly magical quality of being able to sculpt ferrofluid has been the main spark of inspiration that motivates us to exploit this interaction.

Aesthetic interaction suggests that the use of the system gives it meaning and creates an aesthetic experience [19], the appropriation of Linetic by the user and the freedom for improvisation makes this a platform for creative expression. Here the user is able to intelligently construct music through the simple and ordinary actions of playing with liquid, making the natural into a magical experience, thus capturing their imagination.

The third way in which a person can relate to an object is when the object communicates to the person; there is no interaction and it is much like how we enjoy a painting. Even at this level Linetic has the ability to engage the audience. The liquid, which seems to have a life of its own, can mesmerize even non-interacting audiences and this quality was evident in the works of Sachiko Kodama. With Linetic, the audience can reach out to the liquid and with the movement of their fingers, can engage with this liquid as a piano player would with piano keys. This allows the user to engage in a more intimate way with the interface as Fels defines intimacy, and allows the user to use this interface as an extension of them.

One of the questions we wanted to address with this project was how we can keep audiences engaged in the experience, and help them take away a meaningful experience. Music or art that has stood the test of time are examples of works where the creator is able to captivate the audience even after decades from when the work was produced [17]. Many interactive systems today have not yet been able to capture users in the same way. People easily adapt to new technologies and once they understand how to use it, it can or frequently looses its novelty. Liquid is a material that people from different disciplines continue to study to understand its powerful effect on our body, mind and beyond. Liquid as an interface for computing opens the possibility of a whole new way of how we understand and engage with the computing world. The exact ways in which we might use liquids in future interactive systems is still to be explored and Linetic is a beginning of this exploration.

7 Characterizing Linetic as Disposed to Analogness or Digitalness

Using the method offered by Koh et al. [15]. we attempted to situate Linetic in the Analogness-Digitalness Continuum (ADC), by comparing the system with the characteristics for interactive systems offered in the research. Table X shows a breakdown of how Linetic scored in regards to its Analogness & Digitalness.

Table 1. A table showing how Linetic scored in terms of its disposition to Analogness or Digitalness

Characteristics of Analogness & Digitalness	Affirmative
The Interface is Intuitive	
Interaction is Continuous	✓
Content and Media are Singular in Embodiment	✓
Content and Media are Dichotomistic in Embodiment	
Interaction is Discrete	✓
The Interface is Mediated	
Hybrid Candidate?	✓

In Analysis of how Linetic scored, we found that our current version of the interaction required some training before being able to use the system and thus did not meet the requirements of an initiative interface.

However Linetic did meet the requirements for a continuous interface where media and content are a singular embodiment, as the morphable ferrofluid used in our interface is both a visual component of information feedback to the user (in addition to sound and formally, haptics), as well as the interface where the user interacts with the system.

Curiously depending on the configuration of Linetic, the interaction characteristics could become discreet. This is mainly due to the possibility to use Linetic as an interface device to an external system.

In conclusion, we find Linetic to lean towards the Analogness side of the continuum, with hybrid-like qualities depending on the configuration of the system.

Future Work

Future work for this project involves efforts in several directions: exploration of new materials, new array configurations, and improvements in the methodology of gestural interactions and most importantly, the development of relevant applications in order to explore the full potential of human-computer interaction using liquids as a medium.

The search for new materials for different effects and increased controllability involves experimentation with materials such as electrorheological (ER) and magnetorheological (MR) fluids. These materials will provide another avenue for OUIs and the Linetic project.

Refinement of the electromagnet array is also important so as to create a more fine-grained and higher-resolution display. The use of other magnet arrangements will also enable us to control electromagnetic fields better, so as to create more complex and intricate shapes, which will enable us to explore more real-world applications.

An effort to create a concise library of gestures that are more intuitive for use with the system is also desirable in order to disseminate a methodology in which other researchers can base new research on. Once a concise gestural library is created the project will focus on developing and executing more user studies in order to fine-tune the usability of the system. Finally an integrated development environment (IDE) for direct programming of Linetic is desired so that the technology can be shared internationally, which can foster application development in the academic and commercial markets.

8 Conclusion

In this manuscript we describe new iterations of the Linetic system that uses the Hall Effect as a novel form of interaction and actuation. By using Hall Effect sensing in combination with a new accessory worn on the fingertips, users avoid messiness and staining, which was a previous downfall of the original system.

We also outline in this paper, three new experiments that enabled us to develop the new iterations of the Linetic system. These include findings regarding the relationship between perpendicular distances of the Hall Effect sensor and the magnetic field generated by the electromagnetic array embedded in the surface of the system, the characterization of the magnetic Hall Effect sensor readings under the influence of multiple magnetic fields, and the relationship of distance from the sensor versus PWM.

A preliminary user study was also performed in order to understand the usability of the system, as well as to inform the authors on the possible challenges of evaluating such a system for use as an interface and display device. We also discuss the addition of haptic feedback facilitated by the vibration of the magnets placed on the fingertips, providing an additional modality of feedback, along with the previously developed modalities of sound and visual output. This new actuation accessory provides a means for users to interact with the Linetic system without the need to actually touch the ferrofluid, and still provides instantaneous tactile feedback.

Finally we also discuss the aesthetic implications in regards to the user's experience when using the Linetic system. The aesthetic implications of using liquid as a means for human-computer interaction are an important topic in regards to our research. We felt a need to outline these implications in order to accurately prepare for future liquid-based interactive systems.

Acknowledgement. This research is supported by the Singapore National Research Foundation under its International Research Center Keio-NUS CUTE Center @ Singapore Funding Initiative and administered by the IDM Program Office.

References

1. Bailly, G., Vo, D.B., Lecolinet, E., Guiard, Y.: Gesture-aware remote controls: guidelines and interaction technique. In: Proceedings of the 13th International Conference on Multimodal Interfaces, ICMI 2011, pp. 263–270. ACM, New York (2011)
2. Coelho, M., Maes, P.: Sprout I / O: A Texturally Rich Interface, pp. 221–222. ACM (2008)
3. Dewey, J.: Art as experience, vol. 12. Perigee (1934)
4. Dunne, A.: Hertzian Tales. Electronic Products, Aesthetic Experience, and Critical Design, vol. 2. RCA Computer Related Design Research (1999)
5. Fels, S.: Intimacy and Embodiment: Implications for art and technology. In: Computer Engineering, pp. 13–16 (2000)
6. Frey, M.: Snoil. a physical display based on ferrofluid (2006), `http://www.freymartin.de/en/projects/snoil/` (accessed August 18, 2011)
7. Gerhardt, T.: Mud tub (2009), `http://tomgerhardt.com/mudtub/` (accessed August 18, 2011)
8. Holman, D., Vertegaal, R.: Organic user interfaces: designing computers in any way, shape, or form. Communications of the ACM 51, 4855 (2008)
9. Ishii, H.: Tangible bits: designing the seamless interface between people, bits, and atoms, p. 3. ACM

10. Ishii, H.: The tangible user interface and its evolution. Communications of the ACM 51, 32 (2008)
11. Jansen, Y.: Mudpad: Fluid haptics for multitouch surfaces. In: Human Factors, pp. 4351–4356 (2010)
12. Khaslavsky, S.N.: Understanding the seductive experience. Communications of the ACM 42, 45–49 (1999)
13. Kodama, S.: Dynamic ferrofluid sculpture. Communications of the ACM 51, 79 (2008)
14. Valino Koh, J., Karunanayaka, K., Sepulveda, J., Tharakan, M., Krishnan, M., et al.: Liquid interface: a malleable, transient, direct-touch interface. In: Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology, ACE 2010 (2010)
15. Koh, J.T.K.V., Peiris, R.L., Zhu, K., Polydorou, D., Nakatsu, R.: Uncovering analogness and digitalness in interactive media. In: Proceedings of the 30th ACM International Conference on Design of Communication, SIGDOC 2012, pp. 233–242. ACM, New York (2012)
16. Korsmeyer: Picard, R.W.: affective computing, Minds and Machines, 9 (1999)
17. Locher, P., Overbeeke, K., Wensveen, S.: A framework for aesthetic experience. In: Advances, pp. 9–12 (2009)
18. Petersen, M.G., Iversen, O.S., Krogh, P.G., Ludvigsen, M.: Aesthetic interaction. ACM Press (2004)
19. Pier, M.D., Goldberg, I.R.: Using water as interface media in vr applications. In: Proceedings of the 2005 Latin American Conference on Human-Computer Interaction, CLIHC 2005, pp. 162–169. ACM, New York (2005)
20. Schwesig, C.: What makes an interface feel organic? Communications of the ACM 51, 67 (2008)
21. Shusterman, R.: Pragmatist aesthetics; living beauty, rethinking art, vol. 50. Blackwell (1992)
22. Vertegaal, R., Poupyrev, I.: Introduction (special section on organic user interfaces). Communications of the ACM 51, 26–30 (2008)
23. Wakita, A., Nakano, A., Kobayashi, N.: Programmable blobs. ACM Press (2011)
24. Weiser, M.: The computer for the 21st century, Humancomputer interaction: toward the year 2000, pp. 933–940 (1995)

When Paper Meets Multi-touch: A Study of Multi-modal Interactions in Air Traffic Control

Cheryl Savery[1,3], Christophe Hurter[3,4], Rémi Lesbordes[2], Maxime Cordeil[2,3,4], and T.C. Nicholas Graham[1]

[1] School of Computing, Queen's University, Kingston, Canada, K7L3N6
{cheryl.savery,nicholas.graham}@queensu.ca
[2] DGAC DSNA DTI R&D 7, Avenue Edouard Belin 31055, Toulouse, France
[3] ENAC 7, Avenue Edouard Belin 31055, Toulouse, France
[4] IRIT, Université de Toulouse, 118 Route de Narbonne, Toulouse cedex, France
{christophe.hurter,maxime.cordeil}@enac.fr,
remi.lesbordes@aviation-civile.gouv.fr

Abstract. For expert interfaces, it is not obvious whether providing multiple modes of interaction, each tuned to different sub-tasks, leads to a better user experience than providing a more limited set. In this paper, we investigate this question in the context of air traffic control. We present and analyze an augmented flight strip board offering several forms of interaction, including touch, digital pen and physical paper objects. We explore the technical challenges of adding finger detection to such a flight strip board and evaluate how expert air traffic controllers interact with the resulting system. We find that users are able to quickly adapt to the wide range of offered modalities. Users were not overburden by the choice of different modalities, and did not find it difficult to determine the appropriate modality to use for each interaction.

Keywords: Paper computing, augmented paper, digital pen, interactive paper, tangible interfaces, air traffic control.

1 Introduction

In modern air traffic control centres, paper flight strips continue to be used by air traffic controllers because they provide a tangible interface that aids in the visualization of the aircraft under their control. Flight strips allow easy collaboration and sharing of duties and provide an efficient means of recording communication between pilots and air traffic controllers [16]. However, as the flight strips are paper artifacts, the information they contain cannot be digitally processed to help air traffic controllers perform more efficiently and more safely. For example, with a paper-based system, it is impossible to automatically alert an air traffic controller when a plane is given clearance to descend to a level before that level has been vacated, or to warn the air traffic controller if an aircraft overshoots its assigned flight level. Although computer-based systems exist, controllers have been reluctant to adopt them. This reluctance is in part due to the superior interaction qualities offered by paper.

New technologies including multi-touch surfaces, Anoto pens [2] and augmented reality (AR) have led to new interaction techniques that have the potential to bridge the gap between the digital world and paper flight strips, allowing information entered

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 196–213, 2013.

onto flight strips to be recorded digitally, while still allowing air traffic controllers to manipulate the strips in a physical manner. In addition to increasing safety, a digital system can assist the controller, making it easier to locate flight strips and schedule aircraft. For example, when the controller points to an aircraft on the radar screen, the associated flight strip can be highlighted. Similarly, if the controller points to a flight strip, the plane can be highlighted on the radar screen. This has led us to the development of the Strip'TIC augmented flight control system where controllers can fluidly switch between pen and touch interaction on both physical and digital strips. This supports the familiarity and rapid interaction of traditional techniques while providing the safety and collaboration opportunities of digital interaction.

Combining multiple technologies in this way has the potential, however, of burdening users with too much choice. In a given situation, users must determine whether it is better to use a pen on a paper strip, pen on a digital strip, or touch-based manipulation of the paper strips. It is not obvious that this proliferation of input possibilities necessarily leads to a better user experience. We therefore ask the question of whether, in the domain of air traffic control, more options and modalities are helpful, or merely lead to confusion. We are interested in how users deal with the availability of different input techniques: do they take advantage of the numerous options, or do they settle on a smaller subset of the options? And if they use only a subset, do all users choose the same subset, or different ones?

In this paper, we present an interface that combines multi-touch, Anoto pen, tangible objects and augmented reality and investigate the interactions involved in several tasks typically performed by air traffic controllers. For each task, we have developed a variety of possible interaction techniques. The interactions were initially developed based on sessions observing the existing processes used by air traffic controllers at the Toulouse Blagnac Airport and at an air traffic control simulation centre in Toulouse. The interactions were then evaluated by air traffic controllers and modifications were made based on their feedback. The air traffic controllers then completed a scenario based on the task of flight stack management during which we observed the techniques they selected to use. Semi-structured interviews were used to gain further insight into the controllers' actions and their reasons for preferring different types of interactions.

This paper makes two main contributions. First, our technical contribution addresses the complex challenges of combining of paper strips, digital pen and finger tracking. Our implementation is simple, flexible and allows finger, pen orientation and tangible object detection. Second, we have investigated the impact of combining multiple input techniques. We find that users are able to quickly adapt to an interface that offers such a wide range of modalities. The availability of different modalities did not overburden the users and they did not find it difficult to determine the appropriate modality for each interaction.

The paper is organized as follows. We first present related work on technologies and interactions for combining multiple types of input, specifically focusing on those that involve pen and/or paper. Next, we describe air traffic control centres, followed by a description our system that combines multi-touch and pen technology with tangible paper objects and augmented reality. We then present a user study in which air traffic controllers tested and evaluated various interaction techniques, and discuss the results from these sessions. Next, we discuss how combining different modalities impacts the user experience. Finally, we conclude by describing how this work will be used to guide the development of the next iteration of our augmented strip board.

2 Related Work

Related work comes from two areas. First we look at research into technologies that have been used to create new user interfaces that combine other technologies with pen and paper computing. Following that, we look at research involving the types of interactions that are used with these interfaces.

2.1 Paper and Pen Computing

Technologies for interacting with pen and paper can be divided into four broad categories [22]:

- Scanning and interpreting the content of paper documents;
- Identifying and tracking the location of paper artifacts;
- Capturing input; and
- Outputting information onto paper

Many pen and paper interfaces have been developed that focus solely on capturing input from a digital pen. These applications cover a diverse range of areas such as documenting scientific research [17], filling out paper forms, composing music [25], mapping [6] and managing medical records [6]. These applications either record information digitally for later transfer to computer or they may display the data as it is entered, allowing verification and correction in real-time.

Other applications focus on tracking paper artifacts and using gestures to change the information displayed. Paper Windows [11] uses markers to track the locations of several sheets of paper and supports gestures for manipulating the information displayed. Mouse-light [21] uses Anoto technology to track the position of a small projector on engineering drawing and display augmented information. Similarly, a hand-held display such as a smart phone can be used to augment information displayed on paper maps [19]. Markers such as ARTags [3] allow the position and orientation of objects, including paper, to be tracked in three-dimensional space. Do-Lenh et al. [7] use a combination of ARTags and touch detection to create multi-finger interactions with paper.

Recent research has investigated techniques for combining pen input with multi-touch surfaces by applying transparent Anoto film to an LCD display [10] or a back-projected FTIR touch display [15]. Aitenbichler and Schnelle-Walka [1] describe an architecture for combining Anoto, touch and AR based markers. However in order to detect the markers through the back-projection foil, the authors found that they needed to be constructed using reflective foil for the white portions of the markers.

Despite the wide range of applications and technologies that exist for interacting with pens and/or paper, to the best of our knowledge, no applications exist that combine the identification and tracking of paper artifacts with the capturing of pen input and outputting information onto the paper artifacts. Our Strip'TIC application for managing paper flight strips tracks the position of paper strips using ARTags, allows users to input information by writing directly on the strips, and provides feedback by projecting augmented information directly onto the paper strip. In addition, the paper strips are positioned on a back-projected surface supporting both Anoto and multi-touch input.

We next look at the types of interactions that are possible with direct input techniques such as multi-touch, pen input and tangible objects.

2.2 Interaction Techniques

Multi-touch surfaces have lead to a wide range of gestures and interactions. For selection and dragging tasks on a table top surface, Forlines et al. [8] show that for a uni-manual task, users performed better using indirect mouse input compared to direct touch. However, for symmetrical bi-manual tasks, direct-touch was superior. Ringel et al. describe gestures for rotating, panning and resizing documents on a multi-user interactive surface [20]. Wu et al. present a series of gestures for multi-touch surfaces and discuss guiding principles for designing such gestures [28]. Wigdor expands on this by defining a classification system for gestures based on the number of fingers involved, the number of shapes and the type of movement [27]. Morris et al. explore cooperative gestures, where the system interprets the input of multiple users as contributing to a single command [18].

Combining pen technology with multi-touch tabletop displays opens up new interactions that combine the expressiveness of touch gestures and interactions with the precision of pen technology. Most research on interaction techniques that combine touch and pen inputs focuses on extending touch surfaces to also support pen input. Yee [29] investigates single touch and pen interactions, suggesting that the finger be used for panning and the pen for drawing. Brandl et al. compare bimanual interactions using two pens, two-handed touch and a combination of pen and touch [4]. They found that the pen-touch combination was superior (faster and fewer errors) to either touch or pen interaction alone. They discuss general design principles for combining pen and touch input and in particular the difference between the roles of the dominant and non-dominant hand. Hinckley et al. [9] identify nine design considerations for combining pen and touch interactions and discuss how people's behaviour when interacting with physical pen and paper relates to touch+pen interactions. They propose a clear distinction between the functionalities of pen and touch: the pen writes and touch manipulates, although they encounter some situations where this rule needs to be ignored; particularly, when interacting with menus, users expected pen and touch to be interchangeable. For information visualization tasks, Walny et al. [26] show similar results, finding that users have clear expectations about what type of interactions should be touch or pen based.

Tangible user interfaces have many advantages over traditional graphical user interfaces [13]. They provide direct haptic feedback as the user is able to manipulate real physical objects and does not need to wait for indirect visual feedback on screen. Tangible objects provide persistence; even if computers fail they still work. They provide a seamless representation of information across physical and digital domains and they tend to encourage two-handed and multi-user interaction.

Pen and paper interfaces form a subset of tangible user interfaces. Users can manipulate paper objects by moving, rotating, stacking and folding them [11] as well as writing and drawing on them and pointing to them with a pen [30].

Holman [11] describes basic gestures such as holding, collating, flipping, rubbing, stapling, pointing and two-handed pointing for interacting with PaperWindows which can be moved about in three dimensions. Do-Lenh et al. [7] suggest a set of finger gestures suitable for interacting with paper on a touch surface. The finger gestures may either interact with the touch surface or with the paper. As well they define gestures for transforming the paper: moving, rotating and covering.

Yeh [30] divides pen and paper interactions into two categories: drawing and commands (gestures). Within these categories, he suggests that there are four types of interactions: selection, writing, drawing, gestures.

In this paper, we extend this work by investigating how multi-touch, Anoto and augmented paper can be combined in the real world application of air traffic control at an approach control centre. We show how the combination of technologies allows us to leverage the strengths of each technology and point to new interactions that utilize multiple technologies.

The following section describes the different types of air traffic control centres followed by a description of the process of stack management, one of the tasks performed in an approach control centre.

3 Air Traffic Control Centres

Based on the portion of the flight for which they are responsible, air traffic control centres are divided into three categories: tower, approach, and en-route. The tower control centre is responsible for the plane from the moment it begins take-off until the plane is in the air and appears on the radar screen. At this point, responsibility is passed to the approach centre. The controllers at the approach centre are responsible for the aircraft until it has left the airspace around the airport. At this time, control is passed to an en-route centre. A series of en-route centres will pass control from sector to sector and possibly country to country until the aircraft approaches its destination airport. Control is then passed to the approach centre for the destination airport. The approach centre accepts responsibility for aircraft arriving from multiple directions and must schedule their arrival at the runway to ensure optimal spacing between flights. When the plane has been scheduled for landing and is on final approach to the runway, responsibility is passed to the tower control centre.

Our work focuses on the tasks performed at approach control centres. An approach centre will typically have two or three controllers depending on the volume of traffic.

- The planner controller communicates with the en-route centres and accepts responsibility for arriving aircraft. The planner is responsible for scheduling the order in which the aircraft will land.
- The radar controller communicates with the pilots and executes the plan developed by the planner controller.
- The intermediate approach controller handles the aircraft from the time they pass the final beacon until control is passed to the tower control centre. At smaller airports, this position is not used and the radar controller retains responsibility for the aircraft up until control is passed to the tower.

When the volume of air traffic is high, the planner controller, must often organize the arriving aircraft in holding stacks prior to scheduling them for landing. The process of managing these stacks is described in the following section.

3.1 Stack Management

Holding stacks are used to delay aircraft when they cannot land, typically due to congestion or weather conditions. Planes fly in horizontal loops as shown in Figure 1. A large airport typically has multiple holding stacks and the planner controller must coordinate the planes in each stack and time their departure from the stack to provide optimum spacing between the aircraft, ensuring both safety and efficiency.

Planes enter the stack at the highest level and are successively given clearance to descend to the next level. When a plane exits the stack, a known and fixed amount of time elapses before the plane lands on the runway. When two stacks are involved, these times are usually different. For example, at the Orly airport when landing eastbound on the QFU 06 runway, the time required when leaving the ODILO stack is 9 minutes and when leaving the MOLBA stack the time is 17 minutes. The air traffic controller must mentally calculate when each plane should leave the stack, interleaving planes from each stack and maintaining a 90 second gap between landings.

As shown in Figure 2, air traffic controllers organize their flight strips on the strip board to help provide a mental image of the aircraft in the holding stacks. The strips at

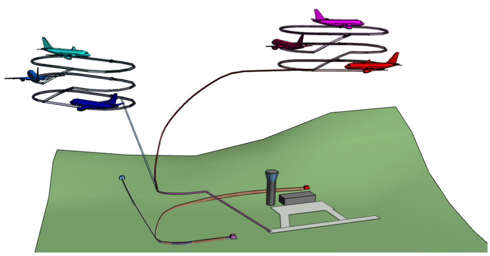


Fig. 1. The departure of aircraft from multiple holding stacks must be coordinated to ensure optimum spacing between the aircraft

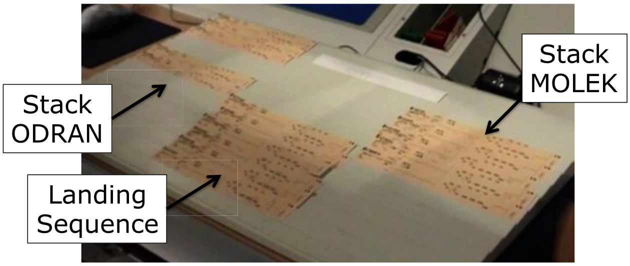


Fig. 2. The flight strips on the strip board are organized into three groups. The upper left and right groups contain the strips for aircraft in each of the two stacks. The centre group holds the strips for the aircraft that have left the stack and are approaching the runway beacon.

the upper left and right represent the aircraft in the two stacks, and the strips in the centre represent the flights that have been cleared to exit the stack. The time of departure from the stack and the expected landing time are written on the strip.

Our augmented strip board, shown in Figure 3, maintains the same arrangement of flight strips, but provides the air traffic controller with augmented information. The flight levels within each stack are clearly shown along the left and right sides of the board and lines connect each flight strip with its level within the stack. When planes are scheduled to depart the stack, the departure time and the expected arrival time are automatically calculated and projected onto the flight strip.

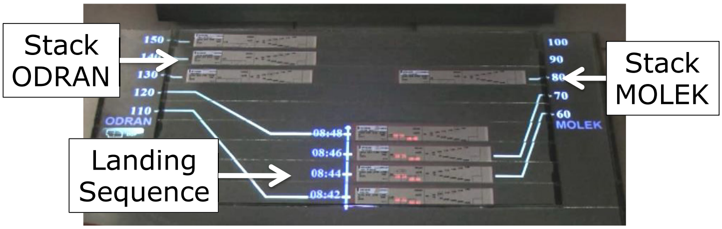


Fig. 3. Augmented information clearly indicates the level of the aircraft in each stack. The times for departing the stack are automatically calculated and displayed.

We next describe our augmented flight strip board and the technical challenges of providing multi-touch capabilities.

4 Augmented Strip Board

Strip'TIC [12,14] was developed to bridge the gap between paper flight strips and fully electronic systems. The challenge was to provide all the efficiency and safety of an electronic system while continuing to allow air traffic controllers to interact with and write on paper flight strips. Strip'TIC was initially designed as a pen-driven interface. This allowed controllers to write on the paper strips and have the information recorded electronically. To provide feedback, we needed to track the position of the strips. This was accomplished using AR tags on the bottom of the strips and a camera located beneath the strip board. Top projection was added to display the feedback directly on the paper strips. Bottom projection was used to display additional information on the strip board such as a virtual representation of a flight strip when the paper strip was removed from the board, and buttons for selecting different modes of operation and for inputting data. We also wanted the users to be able to interact with both the strip board and the radar screen and thus these surfaces were covered with Anoto film.

During initial user testing, it was found that air traffic controllers wanted to be able to use their fingers to interact with either the paper or virtual strips. Thus, we began investigating how to add multi-touch interactions. The next section describes some of the technical challenges in adding touch capabilities to the strip board.

4.1 Multi-touch Technical Challenges with Strip'TIC

The physical shape of the strip board makes multi-touch interaction challenging and makes many of the standard techniques for touch interaction such as Frustrated Total Internal Reflection (FTIR), Diffuse Illumination (DI) and LED frames infeasible. Instead of a smooth surface as found on typical multi-touch tables, the strip board contains a series of ridges or steps that allow the controllers to easily align the paper strips horizontally and keep them in place. To provide a light source for detecting touches, we shine a layer of infrared light above the surface of the strip board [24]. When the surface of the board is touched, this light is reflected down through the strip board and is detected by a camera beneath the board (see Figure 4). The Community Core Vision (CCV) library processes the camera image and outputs finger tracking data. To create the layer of light, a row of 48 infrared LED's was attached slightly above the strip board along the left and right sides of the board as shown in Figure 5. Because the paper strips are flat and below the level of the infrared light, they are not detected and do not interfere with touch detection. However, the surface of the strip board itself does cause multiple false touches along the edges of the ridges. Fortunately, the bands where these false touches occur are narrow and are found in areas where the user does not need to touch. Thus, in software, we are able to ignore any touches that originated within one of these bands.

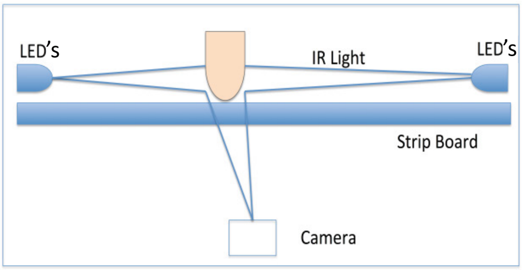


Fig. 4. Light from the infrared LED's is reflected through the strip board when it is touched

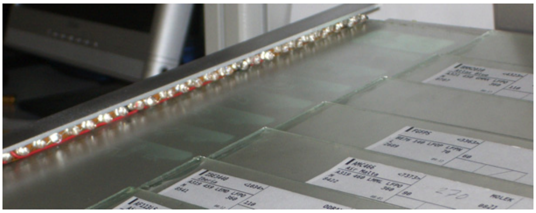


Fig. 5. Infrared lights along one side of the strip board

4.2 Pen Orientation Detection

Because the surface of the strip board is covered with transparent Anoto film, users can interact interchangeably with either the Anoto pen or with touch gestures. The Anoto pen uses an infrared LED to illuminate the Anoto pattern on the strip board. Some of the light from the Anoto pen LED is transmitted through the strip board, allowing the camera used for touch detection to interpret it as a touch. Thus, we have two positional inputs from for the pen, one based on the position of the pen relative to the Anoto pattern, and the other from the touch detection. When the pen is in use, we ignore "touches" generated by the pen and use only the Anoto input. One interesting side effect of obtaining both touch and pen positions when using the Anoto pen is that it allows us to determine the orientation of the pen, and from that infer whether the pen is being held in the left or right hand of the user. As shown in Figure 6, when the pen is held in the right hand (and oriented up and to the right) the light from the pen is detected to the right of the pen tip. When the pen is held in the left hand the light from the pen is detected to the left of the pen tip.

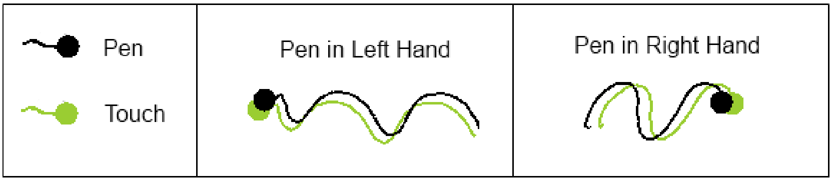


Fig. 6. When the pen in held is the right hand the light from the pen is detected to the right of the pen tip. When the pen is held in the left hand the light is detected to the left of the pen tip.

a b

Fig. 7. The strip board allows for the investigation of other input devices such as (a) a static raised area on the board for entering flight headings or (b) a device with three LED's that can be placed anywhere on the board

Other research has used techniques such as hand occlusion [5] or pens with Wacom tablets [23] to determine the pen orientation. The light detection from our touch surface provides an additional technique.

4.3 Tangible Input Devices

The multi-touch surface of the strip board creates the potential for other input devices. We investigated two potential devices for entering flight headings, represented as a value from 0 to 355 rounded to the nearest five degrees. The first device (Figure 7a) consisted of a raised ring of plastic mounted on the strip board. The position of the user's finger on the ring was used to calculate the heading angle. The second device (Figure 7b) consisted of three LED's forming an isosceles triangle powered by a nine-volt battery. The device could potentially be placed next to a flight strip to enter the heading for that flight.

Thus we see that in addition to providing finger detection, our touch detection system also allows for the detection of both pen orientation and of other tangible objects.

5 Interactions with Paper Flight Strips

Our observations at Toulouse Blagnac Airport approach control centre and at the air traffic control simulation centre at École Nationale de l'Aviation Civile in Toulouse confirmed that air traffic controllers at approach centres interact with paper flight strips in a manner similar to air traffic controllers at en-route centres [16]. Air traffic controllers arrange strips using one hand or two hands, and point at strips to indicate issues with or potential conflicts between aircraft. The controllers switch rapidly and frequently between holding a pen to write and using the same hand to manipulate or point to flight strips. The pen is also used to point to locations on the strip board and on the radar screen. In the approach centre, there is a segregation of duties between the planning controller who coordinates with the en-route control centres and passes control of aircraft to and from these sectors and the radar controller who communicates with the pilots to execute the flight plan determined by the planning controller. Frequent collaboration occurs between the two controllers, with the planning controller inserting new strips on the board of the radar controller. The paper flight strip plays a key role in this collaboration. As the strip is passed from one controller's strip board to another, responsibility for that aircraft is implicitly transferred between controllers.

In developing interactions with augmented flight strips, we wanted to preserve as much as possible the existing interactions with paper strips while providing opportunities for increased safety and efficiency that computer based systems have the potential to provide.

As an aircraft descends through multiple levels within the holding stack, the flight strip can become covered with circles and lines. We preserve the same interaction, allowing the controller to write on the flight strip using the Anoto pen. The system recognizes the circles and lines and highlights the last level selected using top projection. In addition, when a flight level clearance is given, a thick white line appears that connects the flight strip to the appropriate level in the stack. When the flight strip is moved on the strip board, the top projected images and the white line follow the strip.

We also created several alternate interactions for entering flight level clearance information into the system, such as buttons projected along the sides of the strip board.

6 User Study

Our augmented strip board has the advantage of combining traditional paper-based and AR-based digital interaction with flight strips. We wished to address the question of whether users were capable of usefully working with all of these modalities, or whether they led to confusion.

To investigate this question, we held design sessions with 10 air traffic controllers with 5 to 35 years experience (average 16 years) from eight different approach centres in France including Paris, Orly, Brest, Marseille and Blagnac. Seven of the participants were male and three were female; eight were right handed and two were left-handed. Eight of the controllers had previously seen a demonstration of the Strip'TIC system; however, none had used the system. Nine had experience using a touch surface such as an iPad or smart phone and one participant had previously used a digital pen. Each session lasted approximately 45 minutes.

During the sessions, the system was demonstrated to the controllers who were given time to experiment with different interactions and become familiar with the system. The controller was then given a series of tasks to complete using different interaction techniques. Semi-structured interviews were used to obtain the controllers' impressions of the interaction techniques. All of the sessions were videotaped. The tasks performed by the controllers were as follows:

1. Writing Commands
   - On the paper flight strip, enter flight level 120
   - On the virtual flight strip, enter flight level 120
2. Freehand Writing
   - On the paper flight strip, indicate that there is a radio problem
   - On the virtual flight strip, indicate that there is a radio problem
3. Moving Strips
   - Move a paper strip horizontally and vertically
   - Move a virtual strip horizontally and vertically using a finger
   - Move a virtual strip horizontally and vertically using the pen
4. Pen vs Touch with Buttons on Left
   - Assign flight level 130 using touch only
   - Assign flight level 120 using pen and touch
5. Pen vs Touch with Buttons on Right
   - Assign flight level 80 using touch only
   - Assign flight level 70 using pen and touch

In addition, six of the controllers completed a scenario that involved several tasks.

1. Using any technique you wish assign the following flight levels:
   - Flight: AF113LS – Level: 110
   - Flight: BMM2010 – Level: 120
   - Flight: BZ716WH – Level: 140
   - Flight: AMC466 – Level: 60
   - Flight: AF015TM – Level: 70
2. Switching Flight Levels:
   - Change AF113LS to flight level 120 and BMM2010 to flight level 110
   - Change AF015TM to flight level 60 and AMC466 to flight level 70
3. Writing: Make a note on one of the strips that the transponder has failed
4. Schedule the next three flights for leaving the stack
5. Add two more flights to the stack and assign them a suitable flight level

7 Observations

We now describe our observations of each of the tasks performed in the user study.

7.1 Writing Commands on Virtual or Paper Strips

All of the participants found it much easier to circle the flight level on the paper strip than on the virtual strip. With the virtual strip there was always a slight offset between where the participant saw the tip of the pen and where the projected ink marks were displayed. The offset was due to the thickness of the glass and the angle at which the participant viewed the display. The participants were consistently more accurate when entering a flight level directly on the paper.

7.2 Freehand Writing on Virtual or Paper Strips

Again, all of the participants preferred writing on the paper strip compared to the virtual strip. Although the position of the pen was not as critical, the lag in the feedback when writing on the virtual strip, made writing more difficult.

7.3 Moving Flight Strips

Moving the physical paper strips about the board was found to be the quickest way to reposition strips on the board. However, eight of the ten participants, all of whom had previous experience using a touch surface, found using touch to select and drag the virtual strips about the board nearly as simple and efficient provided that two of the strips did not become overlaid. When this occurred, it was difficult to select the strip that they wanted to move. With the paper strips, it was easy to separate strips that were on top of each other. None of the participants felt that the pen was useful for moving the virtual strips.

7.4 Selection Tasks - Interaction Directed

We created an interaction in which the participants could assign a flight level by selecting a bottom-projected button near the edge of the strip board and then selecting one of the flight strips. The interaction was designed so that they could select the button and the strip in either order or simultaneously. They could also use either the pen or their finger for one or both of the touches.

In general the controllers, found that touch and pen worked well for pointing tasks such as selecting flight levels. As shown in Figure 8a, four of the ten controllers preferred using the pen, while the remaining six preferred using touch.

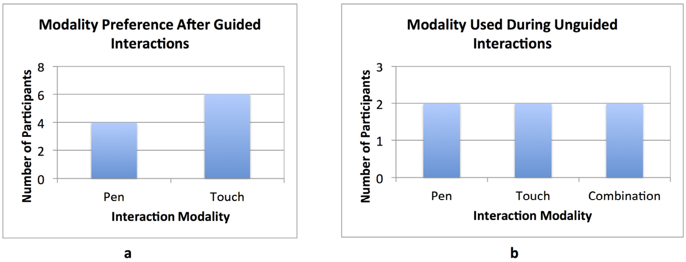


Fig. 8. Users indicated a slight preference for touch over pen interactions. However during the unguided scenario, the users were equally divided between using touch, using the pen and using a combination of both modalities.

7.5 Selection Tasks - Task Directed

Six air traffic controllers completed the final unguided scenario in which they performed selection tasks to assign initial flight levels and to change flight levels. Of these six controllers, two used touch for all the interactions, two used the pen for all the interactions and two used a combination of pen and touch (Figure 8b).

7.6 Scheduling Leaving the Stack

All six of the controllers who completed this task moved the paper strips into the correct position to complete this task. No one attempted to remove the paper strips and move the virtual strips.

The feedback from the controllers about the stack management time calculations was uniformly very positive. All six controllers liked how simple it was to move the flight strips and have the times calculated and displayed automatically. They felt that having the times calculated automatically would be quicker and in dense air traffic could be very helpful for the planner.

7.7 Devices for Flight Headings

The two devices for entering flight headings (Figure 7) were demonstrated to six of the air traffic controllers. They all had strong opinions about the devices, although there was no consensus. Three of the controllers liked the device with the LED's. They liked the possibility of using the device anywhere on the strip board and that it could be used in either hand. Right-handed controllers did not like the idea of using their left hand to enter the headings with the ring device. One controller, Participant 10, preferred the ring that was mounted on the strip board, but did not like its current location. She felt that adding the LED device was just one too many things. It was ok to have paper and pen and touch, but she did not want another device to pick up. Two controllers did not like either option, with Participant 7 suggesting a form of keypad might be more suitable for entering the headings.

7.8 General Comments

The feedback on the system from the air traffic controllers was all very positive. They liked the visual representation of the flight levels at the sides of the strip board and thought it would be good for organizing stacks. The controllers liked the idea of retaining paper strips, especially for passing the strips to another user. They also thought that the idea of a virtual strip was useful commenting "Paper can get lost so a virtual strip underneath could help with that".

8 Discussion

Strip'TIC was initially designed as a pen-driven interface. However, early user testing indicated that air traffic controllers would also like to be able to use their fingers to interact with either the paper or virtual strips. Thus, we added multi-touch capabilities. Through our user study, we attempted to assess whether this added modality was confusing rather than helpful, and to extract design guidelines for our prototype.

8.1 Combining Multiple Modalities

We used our study to address three potential concerns about combining multiple modalities:

Ease of Training: Did having multiple modes for completing a task make it difficult for the air traffic controllers to learn how to use the system?

In general, the air traffic controllers had little difficulty in learning to use the system. With less than 20 minutes of training time, they were able to use all the modalities and found it straightforward to do so. Air traffic controllers are expert users of their systems, and so 20 minutes of training is a negligible cost.

Cognitive Overhead: Did having multiple modes of completing a task require the air traffic controllers to think about which mode to choose before completing the task?

Having multiple ways of completing the same task can cause users to spend more time thinking about which technique to use to complete the task rather than about the task itself. In general, we found that users quickly selected one mode for completing a task and continued to use that mode for all the tasks. We did see some examples of indecision, for example, when first beginning the task involving assigning flight levels to five flight strips, participant 9 initially reached for the pen. However, he did not pick it up and then used touch interactions to enter all the flight levels.

Confusion: Were the air traffic controllers confused by the limitations of some of the modalities?

Confusion may arise when an interaction does not work in the same manner on all system features. The current implementation of Strip'TIC has several potential issues of this type:

- The pressure of writing with a pen on a paper strip can cause the strip to move, requiring the user to hold the strip in place with the non-dominant hand when writing. However the virtual strip does not move when writing in the centre area of the strip. In contrast, placing a hand on a paper strip will prevent it from moving while writing, but placing a hand on the virtual strip will make the strip move, which is not expected.
- Users can touch a virtual strip to select it; however, because the camera used for touch detection is located under the strip board, the camera could not "see" touches on the paper strips. Thus it was not possible for users to select a flight strip by touching the paper strip.
- Depending on the location of the buttons for selecting flight levels, there was a difference in the ease of using the pen for assigning flight levels. For a right-handed user, it was easy to select a flight strip using the pen and then use their left hand to select a flight level to the left of the strip. However, if a person used the same approach for selecting a flight level to the right of the strip, he would become cross-handed and the interaction would be awkward.

We saw instances of confusion when the participants first used Strip'TIC. For example, participant 10 tried placing her non-dominant hand on the virtual strip to hold it in place when writing on it and participant 1 became cross-handed when using a combination of pen and touch to enter flight levels on the right hand side of the strip board. However, these users quickly adapted and did not display any signs of confusion upon subsequent interactions. We found that all six of the air traffic controllers who completed the final unguided scenario adopted essentially the same workflow and used all of the modalities: pen for writing on strips, touch for selecting strips and flight levels and paper for arranging the flight strips on the strips board. The availability of different modalities did not overburden the users and they did not find it difficult to determine the appropriate modality for each interaction.

8.2 Technical Considerations

The Strip'TIC system addresses a range of complex challenges: combining paper flight strips, digital pen and finger tracking. Our implementation of touch interaction is simple, flexible and allows finger detection, pen orientation and tangible object detection.

Our user study highlighted some of the technical limitations of the system primarily related to the use of Anoto pen technology as well as areas for future implementations and investigations. These are discussed in the following sections.

Anoto Pen: Our results show that Anoto pens, while suitable for written input on paper surfaces, should not be used for written input on digital surfaces. There are four technical issues that need to be resolved:

- *Accuracy:* Due to a combination of parallax and alignment of the bottom projected image with the Anoto film, there was a slight difference between where the participant saw the tip of the pen and where the tip was detected on the virtual flight strip that made writing on the virtual strip difficult.
- *Feedback time:* When writing on the paper strips, the controllers received instant feedback seeing the physical ink on the paper. However, on the virtual strip, the delay between moving the pen and seeing the virtual ink trail made writing difficult. Surprisingly, the air traffic controllers found the lag more problematic for freehand writing than for circling flight levels.
- *Pen Angle:* The angle of the pen was also an issue when writing on the virtual strips. On paper, the pen functioned well as long as it was held at 35 degrees or more above horizontal. However on the strip board, the pen needed to be held at an angle greater than 45 degrees. Half of the participants initially held the pen at an angle between 35 and 45 degrees and had to adjust their grip of the pen in order to write on the strip board.
- *Feel:* The contact between the pen and glass was not as pleasing as between the pen and paper.

The Anoto pens were also not suitable for dragging virtual strips on the strip board. The issues here were due to the design of the strip board and of the virtual strip. First, because the pen could be used for writing on the strip as well as for moving the strip, we designated only the 0.5 cm coloured border at the top of the strip as the region that needed to be selected in order to move the strip. The remainder of the strip was reserved for writing on the strip. The users found that this border at the top of the strip was too narrow to select accurately and they often began writing instead of moving. Second, due to the stepped surface of the strip board, the pen often lost contact with the board when trying to move the strips vertically, requiring the users to try repeatedly before the strip would move up or down to the next row.

The Anoto pens did work well for selection tasks on both the paper flight strips and on the strip board.

Touch Detection: We found that multi-touch is a natural feature for moving paper and virtual strips and that users valued having this additional modality of interaction.

However, our current implementation is based on a camera located underneath the strip board and cannot detect when a user touches a paper strip. Thus it is not possible for paper flight strips to be selected by pointing at or touching the strip. We plan to investigate whether other technology such as Microsoft's Kinect or a LED light frame might enable this interaction.

Our current touch detection system, does allow for the detection of both pen orientation and tangible input devices. To date, we have not fully explored the range of possibilities afforded by these inputs.

9 Conclusion

In this paper, we have described our design evolution with the Strip'TIC project. Our technical contribution addresses the complex challenges of combining multi-touch, digital pens and tangible objects. However, it was not obvious whether combining these technologies necessarily leads to a better user experience, and whether more options and modalities would be helpful or merely lead to confusion.

Through our user study we found that users adapted quickly to the interface. We found that all the controllers who participated in an unguided interaction scenario adopted essentially the same workflow and used all of the modalities: pen for writing on strips, touch for selecting strips and flight levels and paper for arranging the flight strips on the strips board. The availability of different modalities did not overburden the users and they did not find it difficult to determine the appropriate modality to use for each interaction.

Acknowledgments. We wish to thank the LEIF Transatlantic Exchange Partnership Project for supporting this student exchange and Stéphane Chatty who made the research possible.

References

1. Aitenbichler, E., Schnelle-Walka, D.: An extensible architecture for multitouch & pen interactive tables. In: Proceedings of EICS (2010)
2. Anoto, digital pen technology, `http://www.anoto.com`
3. ARToolkit, `http://www.hitl.washington.edu/artoolkit`
4. Brandl, P., Forlines, C., Wigdor, D., Haller, M., Shen, C.: Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In: Proc. AVI, pp. 154–161. ACM (2008)
5. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., To, P.: Occlusion aware menu design for digital tabletops. In: Proc. CHI, pp. 3223–3228. ACM (2009)
6. Cohen, P., McGee, D.: Tangible multimodal interfaces for safety-critical applications. Communications of the ACM 47(1), 41–46 (2004)
7. Do-Lenh, S., Kaplan, F., Sharma, A., Dillenbourg, P.: Multi-finger interactions with papers on augmented tabletops. In: Proc.TEI, pp. 267–274. ACM (2009)
8. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-touch vs. mouse input for tabletop displays. In: Proc. CHI, pp. 647–656. ACM (2007)

9. Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H., Buxton, B.: Pen+ touch= new tools. In: Proc. UIST, pp. 27–36. ACM (2010)
10. Hofer, R., Kunz, A.: Digisketch: taming anoto technology on lcds. In: Proc. 2nd ACM SIGCHI Symposium Engineering Interactive Computing Systems, pp. 103–108. ACM (2010)
11. Holman, D., Vertegaal, R., Altosaar, M., Troje, N., Johns, D.: Paper windows: interaction techniques for digital paper. In: Proc. SIGCHI, pp. 591–599. ACM (2005)
12. Hurter, C., Lesbordes, R., Letondal, C., Vinot, J., Conversy, S.: Strip'TIC: exploring augmented paper strips for air traffic controllers. In: Proc. AVI, pp. 225–232. ACM (2012)
13. Ishii, H.: Tangible bits: beyond pixels. In: Proc. TEI, pp. xv–xxv. ACM (2008)
14. Letondal, C., Hurter, C., Vinot, J., Lesbordes, R., Conversy, S.: Strip'TIC: Designing a paper-based tangible interactive space for air traffic controllers. In: Proc. CHI. ACM (2013)
15. Liwicki, M., Rostanin, O., El-Neklawy, S., Dengel, A.: Touch & write-a multi-touch table with pen-input. In: 9th Int. Workshop on Document Analysis Systems, pp. 479–484 (2010)
16. MacKay, W.: Is paper safer? The role of paper flight strips in air traffic control. ACM Transactions on Computer-Human Interaction (TOCHI) 6(4), 311–340 (1999)
17. Mackay, W., Pothier, G., Letondal, C., Bøegh, K., Sørensen, H.: The missing link: augmenting biology laboratory notebooks. In: Proc. UIST, pp. 41–50. ACM (2002)
18. Morris, M., Huang, A., Paepcke, A., Winograd, T.: Cooperative gestures: multi-user gestural interactions for co-located groupware. In: Proc. CHI, pp. 1201–1210. ACM (2006)
19. Paelke, V., Sester, M.: Augmented paper maps: Exploring the design space of a mixed reality system. Journal of Photogrammetry and Remote Sensing 65(3), 256–265 (2010)
20. Ringel, M., Ryall, K., Shen, C., Forlines, C., Vernier, F.: Release, relocate, reorient, resize: fluid techniques for document sharing on multi-user interactive tables. In: Proc. CHI, pp. 1441–1444. ACM (2004)
21. Song, H., Guimbretiere, F., Grossman, T., Fitzmaurice, G.: Mouselight: bi-manual interactions on digital paper using a pen and a spatially-aware mobile projector. In: Proc. CHI, pp. 2451–2460. ACM (2010)
22. Steimle, J.: Survey of pen-and-paper computing. In: Pen-and-Paper User Interfaces, Human Computer Interaction Series, pp. 19–65. Springer, Heidelberg (2012)
23. Taub, D.: The BoPen: a tangible pointer tracked in six degrees of freedom. PhD thesis, Massachusetts Institute of Technology (2009)
24. Teiche, A., Rai, A., Yanc, C., Moore, C., Solms, D., Cetin, G., Riggio, J., Ramseyer, N., Dintino, P., Muller, L.: Multi-touch technologies. NUI Group (2009)
25. Tsandilas, T.: Interpreting strokes on paper with a mobile assistant. In: Proc. UIST (2012)
26. Walny, J., Lee, B., Johns, P., Riche, N., Carpendale, S.: Understanding pen and touch interaction for data exploration on interactive whiteboards. IEEE Transactions on Visualization and Computer Graphics (2012)
27. Wigdor, D.: Architecting next-generation user interfaces. In: Proc. AVI, pp. 16–22. ACM (2010)
28. Wu, M., Shen, C., Ryall, K., Forlines, C., Balakrishnan, R.: Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In: TableTop, pp. 183–190. IEEE (2006)
29. Yee, K.: Two-handed interaction on a tablet display. In: Proc. CHI, pp. 1493–1496. ACM (2004)
30. Yeh, R., Paepcke, A., Klemmer, S.: Iterative design and evaluation of an event architecture for pen-and-paper interfaces. In: Proc. UIST, pp. 111–120. ACM (2008)

3D Visualization and Multimodal Interaction with Temporal Information Using Timelines

Giannis Drossis[1,2], Dimitris Grammenos[1], Ilia Adami[1], and Constantine Stephanidis[1,2]

[1] Foundation for Research and Technology - Hellas (FORTH), Institute of Computer Science
N. Plastira 100, Vassilika Vouton, GR-70013, Heraklion, Crete, Greece
[2] University of Crete, Computer Science Department, Heraklion, Crete, Greece
{drossis,gramenos,iadami,cs}@ics.forth.gr

Abstract. This paper reports on the design, development and evaluation of TimeViewer, a system allowing the storing, visualization and multimodal interaction with temporal and semantic information using timelines in 3D environments. We focus on the appliance of 3D interactive timelines for temporal information visualization, extending their capabilities with the novel concept of a time-tunnel. Our work is intended to investigate beyond the current state of the art interaction with 3D environments using a variety of input modes, such as touch and full-body kinesthetic interaction. Qualitative results were elicited through an evaluation with 16 users in order to assess the users' opinion of the system, the efficiency of the offered interactions and direct future research in the area.

Keywords: Interactive timelines, multimodal interaction, 3D information visualization, full-body interaction, virtual environment, usability evaluation.

1 Introduction

Timelines are a widespread concept for temporal information representation (i.e., events that took place during a time period) and are employed in various environments and contexts, ranging from exhibition spaces and museums to educational textbooks. In general, timelines aim to present information regarding specific temporal periods along with an overview of related occurring events in condensed, yet rich, form. The contents of a timeline can vary from historical events (e.g., [1, 25]) to numerical data representation (e.g., [11, 12]), while the only limitation is that the various items should be temporally related.

In this context, this paper reports on the design, development and evaluation of TimeViewer, a system allowing the modeling, storing, visualization and multimodal interaction with timelines.

TimeViewer extends the typical visualization techniques offered by the majority of the available interactive timelines, which orders the various events in a 2D space to denote their temporal relation by providing an innovative alternate view that aims to highlight the time dimension and facilitate a more immersive exploration of the available information. This view employs a "time tunnel" metaphor, i.e., a long corridor

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 214–231, 2013.

along which events are placed in chronological order, using distance to represent time. On the tunnel sides, next to each event, exists an extendable ‘showcase’ comprising various multimedia objects (e.g., texts, images, videos and 3D models).

TimeViewer supports a variety of multimodal interaction techniques, even in combination, so as to offer natural interaction in a wide range of hardware set-ups. Apart from common desktop-based interaction techniques (i.e., point and click or multi-touch gestures), TimeViewer supports non-instrumented full-body kinesthetic interaction targeted to navigation and manipulation of 3D virtual environments. To assess TimeViewer in terms of usability and user experience and also compare the various supported interaction methods, an evaluation study was conducted with 16 participants.

2 Background and Related Work

2.1 Timeline-Related Concepts

Events are a common concept of timelines and represent any type of incident that took place at some point in time. Information concerning events is represented in chronological order and usually includes at least a title/short description, as well as when it happened. Event occurrences may be (or considered) instant, e.g., the birth of a person, or have a duration, e.g., the construction of a monument.

Periods are time frames during which incidents share some common characteristics or hold a distinct meaning.

Finally, **Categories** define groups of semantically relevant events whose common denominator can be their type, their context or a specific attribute (e.g., the director in the case of a movie). Categories are mainly used to facilitate interaction by minimizing the amount of the displayed information through filtering out certain objects.

2.2 Interactive Timelines

Interactive timelines may be broadly classified into three distinct categories according to the type of information they present, as well as the aspect of information they focus on: (a) Historical Events Representation (e.g., [1, 25 and 20]), (b) Temporal Data Representation (e.g. [11, 12]) and (c) Semantic Timelines (e.g., [8, 21]). The first category includes timelines that represent historical events, aiming to provide an overview around a specific topic. Timelines representing temporal data consist of numerous values of one or more variables that change over time; therefore, the second category can be considered as enriched graph visualization techniques, which primarily focus on the analysis of data distribution over time. Finally, the third category, semantic timelines, represents events with metadata that can be interrelated with others, primarily focusing on displaying the relationships between events rather than providing detailed information about the events per se.

Faceted navigation is a common practice used for temporal data visualization in timelines [2, 16, 23 and 25], implemented through the adoption of taxonomies to classify data in multiple ways and allows the application of filters to information. Faceted display is implemented through hierarchical trees [23, 25], separate quantitative controls [17] and separate toggle / filtering controls [10].

Existing 3D timelines include Beedocs [3] and Kullberg's Dynamic Timelines [9], both of which place events on a two dimensional plane. The major drawback of both is that they do not support a formal way to acquire input from a knowledge source, while Kullberg's approach [9] additionally fails to provide rich interactivity and an overview. Furthermore, another issue of these approaches is the lack of content categorization and time segmentation. In all aforementioned referenced cases user input is provided through a mouse, a keyboard or a touch surface.

2.3 Multimodal / Natural Interaction

Multimodal user interfaces support interaction techniques which may be used sequentially or concurrently, and independently or combined synergistically, aiming to increase the systems' usability by the application of different modalities that each time suit the system's needs.

Gestures can be defined as a form of non-verbal communication in which visible body actions communicate particular messages; gestural interaction is not only widely studied in the literature, but is actually one of the most popular approaches that users prefer due to its intuitiveness [13, 18, 24]. Furthermore, hand gestures can be used to augment systems and allow supplementary interactions [7] when combined with other means of interaction, such as single- or multi-touch. Apart from single-hand gestures, both users' hands may be used in combination to enrich the set of available gestures [5]. Finally, the use of feet gesturing [18, 22] and foot tapping [4, 17] are occasionally conceded as supplementary helpful interaction tools.

In addition to hand gestures, body movement can augment and facilitate natural interaction. Body movement indicates the pose of a user's body as mentioned in [6], which can be tracked and used for selective interaction with the environment. Papadopoulos et al. [15] use defined body poses recognition in order to allow navigation in 3D environments.

3 Motivation and Design Requirements

The overall goal of the presented work was to design, develop and assess an integrated approach for modeling, storing, retrieving, and visualizing temporal information using multimodal, natural, interaction techniques. In brief, according to the envisioned design requirements the system should:

1. allow expressing, storing and retrieving event-related data augmented with semantic information, and thus be able to retrieve knowledge from a formal data model;

2. support the arbitrary grouping (categorization) of events in a non-restrictive way; thus, apart from temporal categorization, the system should offer a content-independent semantic categorization mechanism;
3. encompass natural user interaction techniques, such as body movement, hand and leg gestures, which can be used independently or in combination. Any single interaction technique should allow users to manipulate all parts of the system in a consistent way in all supported alternative views;
4. provide alternative, appropriate and complementary, ways (views) to represent information among which the user should be able to dynamically switch 'on-the-fly';
5. provide suitable display modes for the presentation of event information, both for fundamental data such as titles and for extended details such as descriptive text, images, videos and 3D models.

Brainstorming sessions with HCI experts took place in order to define the appropriate design, regardless of the content displayed and the timespan examined. Interviews were conducted with real users with no technological background, after short experimentation with prototypes, in order to assess the fundamental functionalities they would expect to be offered from such a system. The most significant finding of these interviews involved the segmentation of the process of exploring a timeline in two sections: the users expected the system to offer firstly an overview of the examined time region and secondly a technique to examine details of interest on demand.

From the early design stages it was decided to instantiate the envisioned system through a 3D user interface, thus adding some additional requirements, as interaction complexity in 3D space radically increases due to the additional available degrees of freedom. In this respect:

1. interaction should be precise and rich, while remaining as simple as possible;
2. user disorientation should be prevented;
3. interaction for temporal exploration should offer means to travel both easily through prolonged and precisely through short time frames;
4. the system should be robust and tolerant to user behavior (e.g., gestures, movement) that may be beyond the foreseen 'repertoire' of actions, avoiding unexpected 'reactions' or states;
5. the system should be able to render both in 2D and stereoscopically, allowing to explore and assess in practice the potential benefits and shortcomings of each rendering approach.

4 System Overview

This section provides an overview of the characteristics and design rationale of the TimeViewer system, which was created based on the aforementioned rationale and requirements. In order to cover the need for semantic information to represent knowledge, timelines are described and stored as an ontology using the Web Ontology Language (OWL) [14], a widely used formal data model used for knowledge information representation in many state of the art systems. Data stored in the ontology can be retrieved through the use of SPARQL queries (SPARQL Protocol and Resource Description Framework Query Language) executed either at initialization or at runtime. Information retrieval is performed by a dedicated module which can easily be substituted by another implementation which uses an alternate way to represent knowledge.

4.1 Visualization Approach

A principal characteristic of TimeViewer is the provision of an immersive display of temporal information supplementary to a view that favors the provision of an overview. In this direction, TimeViewer supports two distinct co-existing modes (i.e., views) for visualizing information respectively: the 'Tunnel' (Figures 3 and 4) and 'Classic2D' (Figures 1 and 2). Users can freely and seamlessly switch between the two modes at any time. In this respect, a key design goal was to preserve the consistency between alternative representations of the same information regarding both the 'look' and 'feel' of corresponding interactive items. For example, multimedia elements are displayed and manipulated in the same way in both modes, while, although the actual position and scale of events may occasionally change, their appearance and the way that users interact with them remains identical.

In contrast to the display differentiations required for each view of TimeViewer, the fundamental information such as its title holds the same display in both modes for the sake of consistency. The timeline's title is placed at the upper side of the display in order to inform users about the subject that the TimeViewer presents. Next to the title, resides an interactive label showing the category of events presented. Upon selection, a drop-down list appears containing all the available event categories. As event categories in TimeViewer are mutually exclusive, only one can be active at a time. The placement of the related interactive object next to the timeline's title is intended to aid towards understanding the functional role of categories, i.e. filtering and hiding information. Furthermore, using a drop-down menu allows supporting a large number of choices using minimum display estate. Similarly to simplified faceted navigation [2, 16, 23,25], events not belonging to the selected category are just "dimmed" (Figure 1) and not completely removed from the visualization, as on the one hand, they constitute integral parts of contextual information regarding the rest ones, while, on the other hand, hiding and showing events would result in significant changes to the visualization of the timeline and eventually to confuse the users.

To avoid potential visualization problems (e.g., overcrowding, empty spaces) that arise when some periods contain too many or too few events, in both modes, the visual space taken up by a period can be appropriately scaled (up or down) out of proportion. Furthermore, color coding is also supported to highlight the time period each event belongs to.

Events may be 'spontaneous' or have a duration and can be accompanied by multimedia information, including text, videos, images and 3D models. Each multimedia type affords different interactions: images may be zoomed, videos may be played, paused and stopped, texts may be read aloud, and 3D models may be magnified and rotated. All multimedia types may be accompanied by a descriptive label.

Classic2D Timeline Visualization Mode. The goal of this mode is to provide an overview of the available information in a manner familiar to the user, easy to understand and convenient for navigation (Figure 1). Thus, in accordance to the convention followed by similar timelines, the horizontal axis is used to represent time. Periods are also represented along this axis. Time and period representation is placed in the upper side of the virtual world and acts as an upper boundary, below which all events are displayed. The vertical axis is used for the visualization of events, taking care so that closely occurring events do not overlap. More specifically, in the case where any two events overlap, the one that occurs later is placed below the other, thus providing an additional visual clue for perceiving the temporal order of closely occurring events.

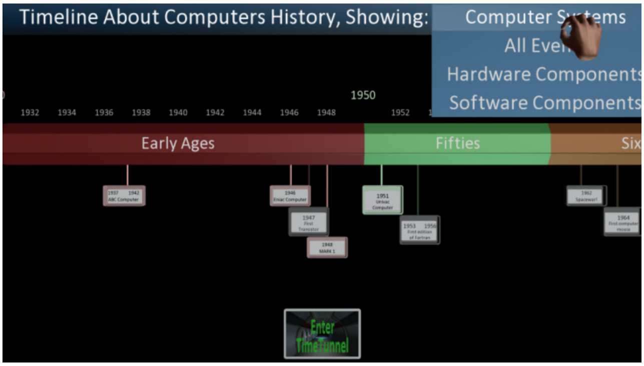


Fig. 1. Enabling the category ‘Computer Systems’ in Classic2D Mode

Time legends appear over time periods, which can adapt to the current level of detail (i.e., zoom level), in order to avoid overcrowding the scene with redundant information; when not viewing the maximum detail level, respectively, the legends are scaled up in order to remain clearly visible. Finally, a button resides at the bottom of the display, which allows the user to switch to the Tunnel mode at any time.

When the timeline is presented at a small scale, events are displayed as rectangular boxes containing a title and a time/date. Events not belonging to the currently selected category are dimmed, indicating that they are not available for interaction. When an events’ size increases, its representative item is added to the container box. Upon selection, the event is brought to the center of the display, scaled up and its container is transformed in order to match the need for hosting additional information: it extends so as to host two additional elements, which are placed at the sides of the box inside an expanded case (Figure 3). The item in the center may be examined in further detail – according to its type –, whereas the next and previous ones are placed on the side. Apart from examining the centered item in detail, the users may proceed to the next or previous items so as to explore all the available information regarding the specific event. In case no further multimedia elements exist, the boxes collapse, indicating that the user has no other items to further explore. At last, events that extend over a span of time are treated as the ‘spontaneous’ ones (placed according to their start time) differentiated only by the display of their start and end time as a time of occurrence.

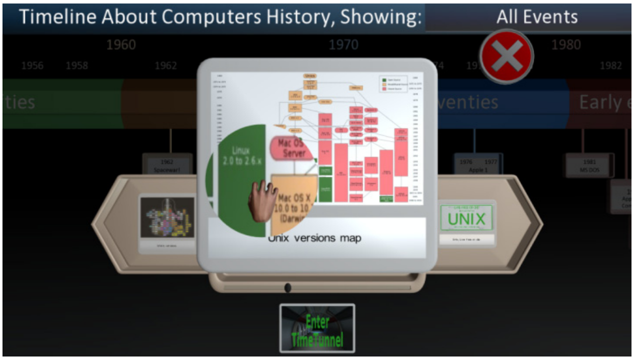


Fig. 2. Magnifying an event's image

Tunnel Visualization Mode. This mode works complementarily to Classic2D and is targeted to supporting the task of sequential, exhaustive, exploration of the presented information, allowing the user to physically experience the time dimension as well as the unambiguous display of each event. The time dimension is visualized as a "tunnel" i.e., a long corridor, along which all information is integrated. Events are placed on the tunnel floor, while periods are placed on the side corners of the tunnel roof. The tunnel walls serve a dual role: they considerably constraint navigation in 3D space, also offering rich orientation cues, and can host contextual information, such as event-related items.

All basic timeline concepts are supported, but are transformed in an equivalent way suitable to fit the needs of detailed event information demonstration. In essence, the virtual tunnel is a sum of individual tunnel parts (slices). Each distinct part may either represent an event or be just a connecting segment used to connect the various events. The distance between events remains constant and linear, as time inside each period is also linear, depicting the time that intervenes between their occurrences. However, in the case where two or more events take place in a very short time, the events' slices would overlap as their extent is fixed. Therefore, TimeViewer adopts the idea of "freezing" time in a region around the event, equal to its slice, in order to allow the straightforward sequential placement of events occurring even in small time intervals. Thus, the proposed approach is capable of supporting any time interval between two events.

Periods are represented as horizontal signs along both sides of the tunnel roof which are color-coded consistently to the Classic2D mode. Additionally, to further emphasize the organization of events in periods, the lightning of tunnel sections belonging to a specific period follows the respective color code.

An event's title is displayed on a board residing on the tunnel floor with an indication of the event's occurrence time on the side. In addition to the event's title, a 'cavity' with a showcase on its front is placed on the one side of the walls of the tunnel. The cavity includes a glass showcase housing the event's representative item, highlighting the visual representation of the event and providing an image to which the users may be familiar to. The showcase is placed inside the wall and protrudes, when the user virtually passes near an event. As indicated by a handle placed at the outer side of the glass, the showcase may be dragged outwards so as to display additional information regarding the event. Upon dragging the multimedia elements that host the event's additional information extend in a row, covering the space that the expanded showcase occupies. Events that span over time are presented using two separate boards, one representing the beginning being fully functional, and the ending one being supplementary and only displaying a link to the event's start. The display of the end of an event with duration is altered, as its title and date is grayed out, unambiguously designating its semantic difference.

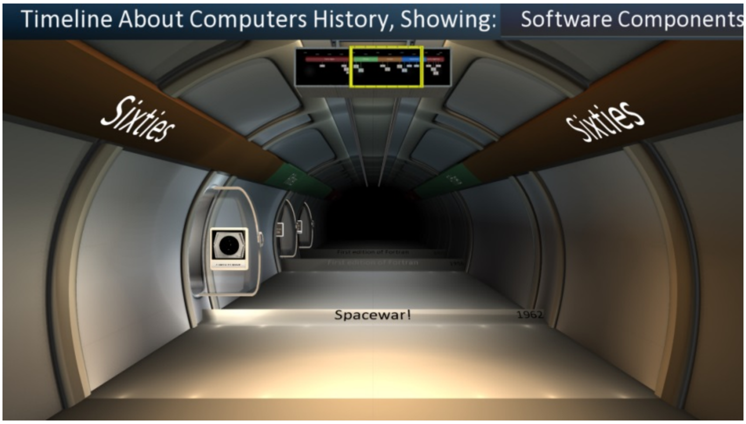


Fig. 3. The nearest event is hidden in the tunnel due to categorization

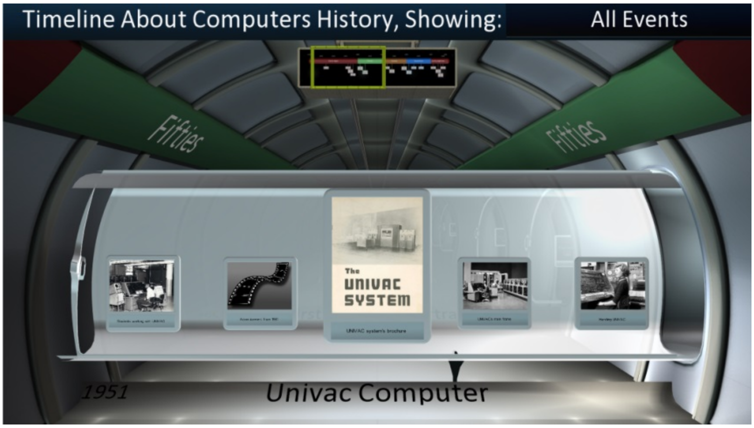


Fig. 4. The view of an event with its showcase expanded

Events that do not belong to the currently selected category collapse on the axis that the virtual tunnel extends. However, their presence is still evident, due to the fact that the tunnel's union with each event's collapsed slice is specially designed so as to be different in comparison to the rest of the tunnel's walls, allowing the user to identify the presence of an event at that spot (Figure 3).

The user's perception of the time currently inspected is enhanced by the presence of a minimap in the form of a railway wagon, which is placed at the tunnel's roof, covering the whole timeline extent. The minimap rendered is the Classic2D representation along with a frame that provides live feedback regarding the equivalent position in the other view of TimeViewer, while also acting as a gateway to Classic2D view.

4.2 Supported Interaction Methods

TimeViewer is designed to support diverse interaction devices and techniques including desktop-based (e.g., touch screen) and full-body kinesthetic interaction.

Touch Screens. The user's touch point on the screen is projected to the virtual world, allowing the selection and dragging of items. Item selection is accomplished by either clicking or dwelling upon (i.e., touching for some time) an item. Item lists (e.g., multimedia elements related to an event) can be scrolled.

Navigation in the virtual world is achieved by dragging. In the case of Classic2D representation, users can slide their finger anywhere on the 2D plane in order to move left, right, up and down. Navigation inside the virtual tunnel is accomplished in a similar way, as the view of the tunnel is changed by dragging the tunnel's floor. The available directions towards which the users may drag the tunnel include forward, backward, left and right along the tunnel's extent.

Kinesthetic Interaction. TimeViewer supports kinesthetic interaction through body movement and gestures which are tracked using Microsoft's Kinect. The general idea regarding TimeViewer's kinesthetic interaction involves the distinction of navigation in the virtual world and interaction with the timeline's elements: users use their feet to move and their hands to reach out virtual elements, corresponding to a person's actions in the real world. As TimeViewer is currently a single-user system, when multiple persons are within Kinect's field of view, only the nearest person's actions are taken into account, as follows.

User Position. The user's position in comparison to the display is used as a tool through which the user is able to manipulate TimeViewer in respect to the virtual environment shown. Firstly, the user may stand off-center, towards either side of the display, in order to scroll the projected view towards the respective direction in Classic2D mode. Furthermore, the system interprets the distance of the user from the display as a zoom mechanism, enlivening the experience of his or her movement in space. Finally, the Tunnel mode of TimeViewer aims to provide an immersive experience of the users' actions; thus, the virtual tunnel augments the real space and the system maps the actual position to the place from which the user looks in the tunnel, creating the illusion of "being there".

Virtual Hand. TimeViewer adopts the concept of using a virtual cursor to allow the selection of items due to the remote control of the system. The cursor, in the form of a three dimensional virtual hand, appears when the user extends his or her hand towards the display, following and mapping the user's real hand. Items' selection is accomplished by keeping the hand cursor over them for a short period of time.

Hand Gestures. The usage of the virtual hand, however, is not the optimal solution in some cases, where gestures may suit better: such examples may be the act of pushing an item away as well as scrolling through elements in a row. Therefore, Time-Viewer offers the interpretation of users' gestures, using either both hands or a single one. Both user hands are tracked and a number of single and dual hand gestures are supported.

The simplest single-hand gesture supported is the movement of one hand left or right in order to scroll multimedia elements lists in both modes and fort navigation purposes in Classic2D mode. Another common task in three dimensional virtual environments is the provision of camera rotation in a straightforward way: the TimeViewer equivalent single hand gesture involves raising the hand left or right while turning the torso in the corresponding direction, which results to the rotation of the projected view around its vertical axis.

Dual-hand gestures include pushing and pulling by moving both hands forward and backward respectively. These gestures can be used with items residing in front of the user, such as the multimedia elements. Additionally, in the Tunnel mode, users may move to the next or previous available event by pulling or pushing respectively. In Classic2D mode, users can push the event back to its original position when finished examining detailed information.

Leg Gestures. Legs are a natural means for supporting navigation-related interactions. Time-Viewer adopts the notion of stepping: while a user steps right, left, forward and backward, navigation is achieved in the corresponding direction. In Classic2D mode, stepping forward or backward results in increasing or decreasing the scale of the time-line respectively, while stepping left and right results in exploring the timeline's display at the respective direction. In Tunnel mode, stepping forward or backward allows navigation along the tunnel, while stepping at the side results in displaying the tunnel as if the user is looking from the equivalent side.

5 Evaluation

5.1 The Evaluation Process

The evaluation session was divided into three separate segments, one for each of the primary interaction modes (touch and kinesthetic interaction) and a complementary one for the assessment of stereoscopic view: 3D user interfaces lately involve the adoption of stereoscopic displays, which can offer a sense of depth in the visualized information, creating richer representations and making the systems more immersive, but, can also be tiresome since the related technologies has not yet matured. Therefore, an assessment of their impact and usefulness should provide insights on their helpfulness in comparison to traditional displays. The first evaluation segment included the interaction using a touch screen, the second one integrated the manipulation of the system remotely, through the use of user tracking, and the third one integrated the usage of NVidia's 3D Vision technology. The aim of the evaluation process was:

1. to evaluate the system design in terms of perceived usability, effectiveness, likeness, and learnability
2. to assess kinesthetic interaction opposed to desktop interaction, as well as provide an insight regarding the factor of fatigue
3. to assess the effects of stereoscopic vision in comparison to two dimensional display

In order to eliminate bias towards either interaction method, users were split into two even groups. The first group started the evaluation process using the touch screen interaction mode first and then the hand gesturing interaction mode, whereas the second group did the opposite. Both groups evaluated the 3D vision at the end.

The evaluation process started with informing the users of the goals and the process of the evaluation. They were then requested to sign their consent to participating in the evaluation and allowing the video taping of the evaluation. After the introduction, the functionality of the TimeViewer system was explained to each user, along with the available gesture sets that TimeViewer supports.

A series of 7 user tasks were assigned to the participants for each one of the evaluation segments. The participants were instructed to express their thoughts aloud (Think Aloud) throughout the completion of each task. In addition, at the end of the evaluation they were requested to fill out a Likert type of questionnaire.

Fig. 5. The set-up of the second segment of the evaluation (projection + Kinect)

5.2 Set-Up and Participants

Two different set-ups were used to evaluate the system. The first comprised a touch screen, while the second one a short-throw projector (supporting stereoscopic vision through the use of appropriate glasses) and a Microsoft Kinect allowing full-body tracking.

A total of 16 volunteers participated in the evaluation, 7 females and 9 males. The age of the participants varied from 20 to 40 years old. Twelve of the users (75%) had intermediate or high computer expertise whereas the other participants had limited expertise. Even though the majority of the users were familiar with computers and touch screen systems, they did not have great familiarity with hand gesturing as a mode of interaction with a system.

5.3 Results

Overall System Design. Overall, the users found that the two different content presentation models, the Classic2D and the Tunnel mode, were complimentary to each other. The majority of the users preferred using the Classic2D mode to obtain a chronological overview of the events and the Tunnel mode to explore specific events. This differentiation in the information display was one of the fundamental goals of this system, as far as design is concerned, and the comments from the evaluation participants indicate that this goal is accomplished. Classic2D mode was regarded by all the users as efficient and intuitive in the way it presented the various periods and their events. The users immediately knew what they were looking at. The application of color coding to the chronological periods and their events made the information easy to follow. The immersive design of the virtual tunnel was met with enthusiasm and excitement by the users who in their majority commented on the "fun" factor of the model. The users understood right away the concept of the TimeViewer and how the information was structured in the three dimensional space.

One of the design issues exposed during the evaluations was the content categories filtration mechanism. The users could view all content on the screen, or select one of the available content categories from a list. However, the graphic used to note the existence of the categories and its placement in the design did not make it obvious to the users who failed to acknowledge its existence altogether. Another design issue had to do with the icons used to note the various multimedia elements which proved to create confusion to a lot of the users, as they were not representative enough of the multimedia element.

Touch Screen Interaction. The first segment of the evaluation process indicated that the touch interaction proved to be an efficient, effective, and intuitive way of interacting with both the 2D and the 3D models. The vast majority of the users instantly understood which touch gesture they had to apply to each content element to produce the desired action without much instruction. For example, they understood that in the 3D Tunnel they had to apply the dragging motion in order to advance through the tunnel and the respective periods. Or, in the instance of the glass showcases, where a grabbing hand icon would appear over the handle, the users understood that they could drag it across the screen in order to open and reveal the showcase's content. Overall, the users were successful in completing the given tasks and there were only very few instances where they needed assistance or gave up completely on a task.

One of the main comments that came up repeatedly through the evaluations was the lack of multi-touch capabilities, as a lot of users suggested adding it as a feature. Other suggestions made by the users included, controlling the zoom level by dragging the slider instead of just pressing the (+) and (-) buttons, iterating through multimedia content by dragging the slider instead of just the content elements, and using the minimap frame of the 3D model to navigate to the corresponding position in the 2D view.

Kinesthetic Interaction. The enthusiasm of the users was apparent during their interaction with TimeViewer with hand gesturing and user tracking, as only one user out of 16 stated his preference to interact with the system using more traditional modes. The users were captivated by the remote handling of the system achieved without the use of any wearable component, especially as passersby and not every-day users. The moves of the hands were regarded as very natural, as indicated by the comments during the evaluation and the completed questionnaires. All the users found the gestures representative of their functions and the only request made by a small percentage of the users (4 out of 16) was that the gestures should become shorter. The rationale of this request was due to the factor of fatigue that could emerge after prolonged interaction with the system. For instance, the gestures for dragging items to the side as well as the gestures of pulling or pushing items using both hands were proposed to be shorter. Finally, gestures were considered easy to learn, as users were generally able to accomplish them on their own after being shown by the evaluator.

Table 1. Qualitative results for each interaction technique, T (Touchscreen), H (Kinect/Hand Gestures) and L (Kinect/Leg Gestures)

Questions	**Strongly Disagree (%)**			**Disagree (%)**			**Neutral (%)**			**Agree (%)**			**Strongly Agree (%)**		
	T	**H**	**L**	**T**	**H**	**L**	**T**	**H**	**L**	**T**	**H**	**L**	**T**	**H**	**L**
I liked interacting with the system.	0	0	0	0	0	0	19	6	31	31	44	38	50	50	31
No special training is needed to handle it.	0	0	0	0	0	6	0	31	13	8	44	31	63	25	50
Corresponded precisely to my actions.	0	0	6	0	6	0	6	44	0	56	44	44	38	6	50
It was awkward to use.	63	38	56	25	38	13	13	25	31	0	0	0	0	0	0
Responded promptly to my actions.	0	0	0	0	0	6	6	31	0	63	44	38	31	25	56
It was tiring to use.	69	38	56	25	25	25	6	31	19	0	6	0	0	0	0
I would prefer another method of interaction	56	44	50	25	31	25	6	19	19	6	6	6	6	0	0

The manipulation of the virtual hand (cursor) came easily to the users. A common remark involved the stability of the cursor, which the users requested to be more precise in order to be able to select items which are small in size. Furthermore, some of the users would prefer the duration of the hand being over items to be shorter so as to select them more easily. The hand gestures were embraced without difficulty and the users largely preferred them to the use of the virtual hand. Moreover, the system reacted as expected as far as hand gesturing is concerned, avoiding false interpretation of the users' movements.

An issue that was raised and that was common among the vast majority of the users was the fact that single-hand gestures were deliberately ignored when the cursor was visible. This design decision was taken so as to avoid unexpected system behavior, as the users could simply move their hand in one direction to select an element. After explaining this limitation to the users, they felt comfortable with the system interaction and managed successfully to apply gestures with respect to their goal.

Leg gesturing was almost unanimously accepted, as only one user had trouble with navigation using his legs (Kinect failed to successfully recognize the exact placement of his legs due to the material of his trousers). The idea of stepping in any direction in order to travel in space came naturally to the users and interacting with the system elements using their hands served the initial goal of discriminating the navigation from the interaction. This observation is more evident in the non-expert users, who supported the leg gestures even more than the expert users, as they felt more comfortable with handling the system naturally but in a strictly defined manner. The participants did not have any problem understanding the conceptual model of moving in the space as they found it to be efficient, tireless and fascinating, which is clearly illustrated in the table containing the answers from the users' questionnaires.

In conclusion, although kinesthetic interaction was considered as fractionally less responsive and more tiring to use (questions 5 and 6 respectively), both hand and leg gesturing are preferred to touch interaction (question 7). Despite the error rates not being as good as touch in total, the users were almost unanimously (15 out of 16) in favor of kinesthetic interaction as the fun factor was considered as more important than mediocre precision.

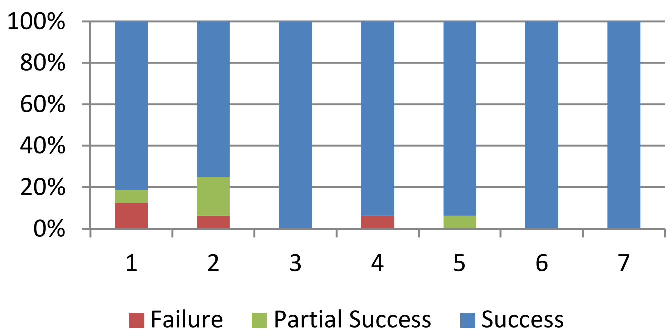


Fig. 6. Touch interaction success rates

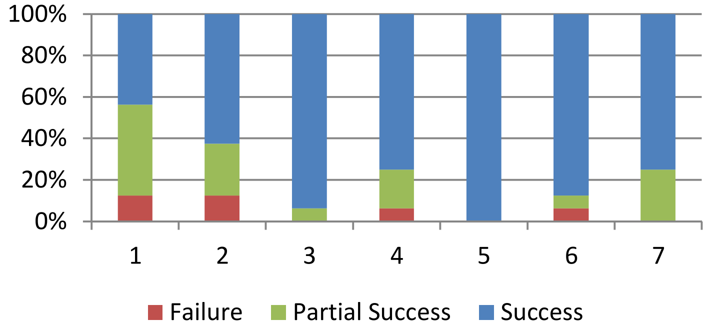


Fig. 7. Kinesthetic interaction success rates

Figures 6 and 7 above display the users' error rates for each given task as percentages. A task is considered to be completed successfully if the user accomplishes it at the first or second try, partially successful if the user needed some help or guidance and is marked as a failure in the case where the user needed too much assistance, gave up on completing the task or could not complete the task with no less than four attempts.

The tasks that users had issues on completing were due to misperception of the system's design rather than inability to understand how to interact with the system. This fact is evident in task 1, where users were asked to find the existing categories of the timeline: the users who partially succeeded or failed to find them did not observe the category filtration mechanism, until the evaluator provided help. Upon assistance on where to focus, the users were able to manipulate the mechanism in both interaction methods. Some users also faced difficulties in task 2, where they were asked to explore a specific event's information and find the types of presented information. In the case of kinesthetic interaction, a larger user percentage needed minor guidance in order to complete a gesture compared to touch interaction (for instance how to pull an item). Finally, task 4 involved finding a video, where the users' failure was due to inability to identify a video (the video element was thought of an image). The presentation of the tasks that were successfully carried out is beyond the scope of this paper, as we focus on the aspects of the evaluation that can be used for the system's improvement or provide interesting outcomes.

3D Vision. After the evaluation process regarding interaction by user tracking with Kinect was finished, the participants were given NVidia's 3D Vision Glasses to assess the effects of stereoscopic information presentation. The users were overall in favor of the stereoscopic view (table 2, last row), as 13 participants state that 3D Vision adds up value to the system's representation, while only two disagree and would prefer the normal view in a total of 16 users. Another interesting outcome of great importance is the fact that none of the participants expressed that the usage of 3D glasses was tiring or awkward, as such an issue would be a major drawback in the future. Users did not find that stereoscopic view had any direct effect, positive or negative, on their performance (this fact is depicted in the question 5 at the table below); however, it improved the overall user experience and entertainment value. Some of the users stated that they felt intrigued to further explore the timeline when using the glasses. In conclusion, the stereoscopic display was regarded as a feature that added up value to the system, providing a more immersive feeling, especially in the Tunnel mode of TimeViewer.

Table 2. Qualitative results of stereoscopic vision

3D Vision Questions	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I liked the interaction in 3D Vision.	0%	0%	13%	44%	44%
It was awkward to use.	75%	25%	0%	0%	0%
It was tiring to use.	56%	19%	25%	0%	0%
I prefer to interact with the system to its normal dimension (2D).	6%	56%	31%	0%	6%
3D Vision improves the system usability	6%	25%	19%	25%	25%
3D Vision display adds up value to user interaction experience.	0%	13%	6%	50%	31%

6 Conclusions

In this paper we have presented the design and characteristics of TimeViewer, a system that allows the storing, 3D visualization and multimodal interaction with temporal information. In addition to the typical 2D representation of timelines supported by the majority of existing systems, this paper introduced the metaphor of a time-tunnel in a virtual 3D environment, that can represent the same information in a different manner also supporting appropriate interaction methods.

TimeViewer was assessed in terms of usability and user experience through an evaluation study with 16 participants. Overall, the results of the evaluation process were very encouraging. As the developed system supports multiple multimodal interaction techniques, even in combination, so as to offer natural interaction in a wide range of set-ups ranging from desktop-based interaction, to non-instrumented full-body kinesthetic interaction in 3D environments, particular emphasis on the evaluation was given on the comparison among all these supported interaction techniques.

To this end, the evaluation participants were excited by the diversity of the supported interaction techniques and were "open" to any new form of interaction beyond the more traditional mouse, keyboard and touch options, embracing innovative and state of the art natural interaction techniques. Therefore, further work should aim to expand the possibilities of this research area, taking into consideration the factor of fatigue after extensive use. Additionally, as the evaluation results indicate, the perception of, and interaction with, the 3D environment was easy and straightforward for all users, thus encouraging the "expansion" of other widely used 2D visualization approaches to 3D worlds, which provide more creative options for information representation and interaction. Finally, stereoscopic rendering proved to have a positive influence to the users' perception of the visualization.

Our future work includes further assessing the system with a larger set of users with highly diverse profiles after refining and implementing issues that came up during the last evaluation. Apart from improvements on existing aspects of TimeViewer, a number of system extensions are planned, such as the capability of nesting of timelines inside other timelines and supporting the dynamic creation of timelines through information retrieval from online sources, such as Wikipedia.

Acknowledgements. The work reported in this paper has been conducted in the context of the AmI Programme of the Institute of Computer Science of the Foundation for Research and Technology-Hellas (FORTH). Special thanks to Anthony Katzourakis for his artistic work and 3D modeling.

References

1. Allen, R.B.: Visualization, Causation, and History. In: Proc. iConference 2011, pp. 538–545. ACM Press (2011)
2. Andre, P., Wilson, M.L., Russell, A., Smith, D., Owens, A., Schraefel, M.C.: Continuum: Designing Timelines for Hierarchies, Relationships and Scale. In: Proc. UIST 2007, pp. 101–110. ACM (2007)
3. Beedocs 3D Timeline, `http://www.beedocs.com/timeline3D/`
4. Crossan, A., Brewster, S., Ng, A.: Foot tapping for mobile interaction. In: Proc. BCS 2010, pp. 418–422. British Computer Society Swinton, UK (2010)
5. Fikkert, W., van der Vet, P., van der Veer, G., Nijholt, A.: Gestures for large display control. In: Kopp, S., Wachsmuth, I. (eds.) GW 2009. LNCS, vol. 5934, pp. 245–256. Springer, Heidelberg (2010)
6. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: A survey. Journal Computer Vision and Image Understanding Archive 108(1-2), 116–134 (2007)
7. Hilliges, O., Izadi, S., Wilson, A., Hodges, S., Mendoza, A.G., Butz, A.: Interactions in the air: adding further depth to interactive tabletop. In: Proc. UIST 2009, pp. 139–148. ACM (2009)
8. Jensen, M.: Visualizing Complex Semantic Timelines. NewsBlib (2003)
9. Kullberg, R.L.: Dynamic Timelines Visualizing Historical Information in Three Dimensions. M.S. Thesis at the MIT Media Laboratory, Boston, MA (September 1995)
10. Kumar, V., Furuta, R., Allen, R.B.: Metadata visualization for digital libraries: interactive timeline editing and review. In: Proc. DL 1998, pp. 126–133. ACM (1996)
11. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C.: TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration. In: Proc. CHI 2011, pp. 227–236. ACM (2011)
12. 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.) SG 2010. LNCS, vol. 6133, pp. 208–219. Springer, Heidelberg (2010)
13. Nickel, K., Stiefelhagen, R.: Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation. In: Proc. ICMI 2003, pp. 140–146. ACM (2003)
14. OWL, Web Ontology Language Overview, `http://www.w3.org/TR/owl-features/` (last accessed on January 22, 2013)

15. Papadopoulos, C., Sugarman, D., Kaufmant, A.: NuNav3D: A Touch-less, Body-driven Interface for 3D Navigation. In: Proc. IEEE VR 2012, pp. 67–68. IEEE (2012)
16. Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B.: LifeLines Visualizing Personal Histories. In: Proc. CHI 1996, pp. 221–227. ACM (1996)
17. Ronkainen, S., Häkkilä, J., Kalev, S., Colley, A., Linjama, J.: Tap input as an embedded interaction method for mobile devices. In: Proc. TEI 2007, pp. 263–270. ACM (2007)
18. Sangsuriyachot, N., Mi, H., Sugimoto, M.: Novel Interaction Techniques by Combining Hand and Foot Gestures on Tabletop Environments. In: Proc. ITS 2011, pp. 268–269. ACM (2011)
19. Serrano, M., Nihay, L., Lawson, J.-Y., Ramsay, A., Murray-Smith, R., Denef, S.: The OpenInterface Framework: A tool for multimodal interaction. In: Proc. CHI 2008, pp. 3501–3506. ACM (2008)
20. Simile Project, `http://www.simile-widgets.org/timeline/` (accessed on January 22, 2013)
21. Stab, C., Nazemi, K., Fellner, D.W.: Sematime - timeline visualization of time-dependent relations and semantics. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Chung, R., Hammound, R., Hussain, M., Kar-Han, T., Crawfis, R., Thalmann, D., Kao, D., Avila, L. (eds.) ISVC 2010, Part III. LNCS, vol. 6455, pp. 514–523. Springer, Heidelberg (2010)
22. Valkov, D., Steinicke, F., Bruder, B., Hinrichs, K.: Traveling in 3D Virtual Environments with foot gestures and a multitouch enabled World in Miniature. In: Proc. VRIC 2012, pp. 171–180. IEEE (2010)
23. Wongsuphasawat, K., Gomez, J.A.G., Plaisant, C., Wang, T.D., Shneiderman, B., Maimom, M.T.: LifeFlow: Visualizing an Overview of Event Sequences. In: CHI 2011, pp. 1747–1756. ACM Press (2011)
24. Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Partk, D., Kim, C.: 3D User Interface Combining Gaze and Hand Gestures for Large-Scale Display. In: Proc. CHI 2010, pp. 3709–3714. ACM (2010)
25. Zhao, J., Drucker, S.M., Fisher, D., Brinkman, D.: TimeSlice: Interactive Faceted Browsing of Timeline Data. In: Proc. AVI 2012, pp. 433–436. ACM (2012)

Activity Theory as a Tool for Identifying Design Patterns in Cross-Modal Collaborative Interaction

Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin

School of Electornic Engineering & Computer Science, Queen Mary University of London
{oussama,nickbk,tonys,fiore}@eecs.qmul.ac.uk

Abstract. This paper examines the question of how to uncover patterns from the process of designing cross-modal collaborative systems. We describe how we use activity patterns as an approach to guide this process and discuss its potential as a practical method for developing design patterns.

1 Introduction

A key challenge in the design of interactive systems is how to leverage existing design knowledge. Design patterns play a role in addressing this challenge by capturing and communicating reusable design knowledge on how to map requirements to design solutions. In the field of human-computer interaction (HCI), design patterns have received increasing attention and sets of patterns have been published for a variety of application domains (e.g. [4, 14]). Similarly, a number of studies within the area of computer-supported cooperative work (CSCW) have addressed the question of how to use design patterns to document successful design knowledge, some of which have generated patterns that can be readily employed to develop practical solutions [2].

We are interested in the design of interactive systems that support collaboration between individuals who use different sets of modalities to interact with each other. Such differences can arise because of permanent sensory impairments or due to the capabilities of the technology used by each collaborator and the context of their interactions with it; we refer to group interactions in these settings as *cross-modal collaboration*. In this paper, we focus on exploring the process by which design patterns for cross-modal collaboration can be identified. We describe how we used an approach based on activity patterns [10] to uncover design patterns from two phases of a typical design process. First, we exemplify the application of activity patterns by reflecting on the evaluation phase of a cross-modal tool that supports collaboration between visually impaired and sighted coworkers. Second, we outline how activity patterns can be used to capture insights from a participatory design workshop that we conducted with visually impaired audio producers and musicians in order to understand their work process and gather requirements about how to design technology that could contribute towards increasing their inclusion in their workplace.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 232–240, 2013.

2 Cross-Modal Collaborative Interaction

Despite significant progress in the use of audio and haptics in multimodal interaction design, research into cross-modal collaboration has so far remained sparse. Initial investigations have nonetheless identified a number of issues that impact the efficiency of collaboration in cross-modal settings. For example, an examination of collaboration between sighted and visually impaired individuals on an interactive puzzle game highlighted the importance of providing visually impaired collaborators with a continuous display of the status of the shared game [15]. Providing collaborators with independent views of the shared space, rather than shared cursor control, was also found to improve orientation, engagement and coordination in shared tasks. In another study [12], a multimodal system combining haptic devices with speech and non-speech auditory output was used to examine collaboration between pairs of visually impaired users on graph reading tasks. Results showed that the use of haptic mechanisms for monitoring activities and shared audio output improves communication and promotes collaboration. Although scarce, the literature on cross-modal collaboration has therefore started to generate insights into the knowledge that is needed to come up with effective designs to support interactions involving individuals with differing perceptual abilities across various domains. We propose to use design patterns as a means to capture such knowledge so that it can be effectively leveraged.

3 Design Patterns

The concept of design patterns originated from the field of urban architecture as a systematic approach to capturing solutions to architectural design problems [1]. Each pattern would characterise a given problem in terms of its context, a set of conflicting forces that are at play, a solution to the problem and the consequences of its application. Particular attention is given to the presentation of patterns as a clear and generic set of instructions so that they can be applied in new situations. A critical aspect of design patterns is that they are cross-referenced. Each pattern refers to higher level patterns describing further context and lower level patterns that could be used to refine the solution. This hierarchical organization structures a collection of patterns into a meaningful network and results in what is called a *pattern language* [7]. The notion of pattern language has also been used to emphasise that patterns should support a "*lingua franca*" amongst all the parties involved in a design process [9]. Patterns in this case can be used as a common vocabulary to improve communication of design expertise between stakeholders [4]. Indeed, design patterns have been shown to have a considerable potential as a method to support participatory design activities [6].

4 Uncovering Design Patterns

In both software engineering and HCI, design patterns are often derived from a reference to existing design solutions and hence are the product of observation, trial and error, and experience [5, 7]. In CSCW, finding patterns has also been described as a process that relies on extensive observation of how people use tools in context. For instance, [11] followed a retrospective approach to examine data from field studies of cooperative interaction in order to uncover patterns that could be generalised across a number of work environments.

The body of work on patterns in CSCW does not always make it clear how patterns come into existence in the first place, however. So the question often remains whether the Alexandrian way of identifying and organising patterns should be used to guide this process, or whether a different approach should be adopted. On a closer look, the hierarchical nature of the scales of patterns described in [1] does not match the hierarchy of problems that are typically encountered in CSCW and an appropriation is therefore required to understand the problem-hierarchy that is relevant to CSCW domains. The domain of cross-modal collaborative systems design presents a set of challenges often characterised by the conflicts associated with reconciling interaction that occurs across more than one modality [13]. In this case, we are interested in uncovering patterns both from direct observation of use of cross-modal collaborative systems as well as from design activities that seek to improve such systems.

4.1 Activity Patterns

We consider activity patterns [10] as a potential guiding framework for addressing the issues highlighted above. According to this framework, Alexander's patterns could be appropriated to embody the principals of activity theory (AT) and hence could be used to analyse activity in terms of understanding tool-mediated work in its context [3]. AT views human activity in terms of a system of tool-mediated actions carried out by a subject (i.e. and individual or a group of individuals) in order to achieve a desired outcome. Actions are characterised in terms of how they are organised within a community context, and how they are regulated by internal rules and mediated by a division of labour. This unit of analysis is conventionally represented by a triangular model to show how its elements interact with each other [8] (see Figure 1).

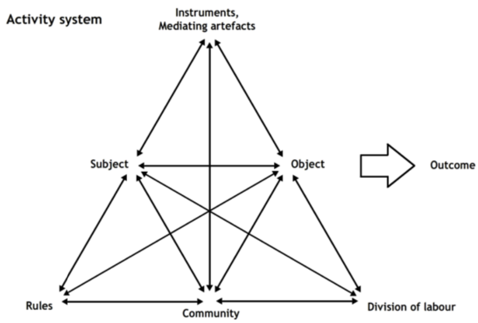


Fig. 1. An activity system as conceptualized by activity theory

According to [10], there are parallels between the design patterns principles introduced by Alexander and that of human activity as conceptualised by the unit of analysis in AT. These include:

- The definition of a pattern in terms of three related components expressing the relationship between a given context, a problem and its solution, which is consistent with the method of AT.
- The characterisation of a problem in a given context as being caused by a system of conflicting forces that arise in that context, which could be captured through AT's conceptual tool of *contradictions*.

Additionally, the hierarchical levels of activity in AT can also be used in a similar way to the Alexandrian concept of scales to help structure the scope covered by activity patterns. [1] used scales to organise their patterns at different levels of architectural design, e.g. city, road, building, room, window, etc. The concept of activity levels in AT are hierarchically structured into three levels, where a given *activity* is realised through a set of concrete *actions*, which are in turn accomplished through a series of *operations*. This gives patterns a sense of scale from high-level activities down to low-level operations. According to [10], patterns could be written to reflect each element in a given activity system – the design of *mediating artefacts*; the work of a *subject*; the *rules* and procedures; and the roles within the *division of labour* or *community* of the work group – as well as organised into a coherent pattern language that preserves the unity of these elements within each of the three levels of activity.

4.2 Activity Patterns in the Evaluation Phase

In order to assess the feasibility of this organisational framework in supporting the process of uncovering design patterns, we applied it to data gathered from two phases of a typical process of designing for cross-modal collaboration; evaluation and requirements capture. In the first instance, we used the activity patterns approach to reflect on the design of an existing cross-modal tool, which we evaluated with visually impaired and sighted users.

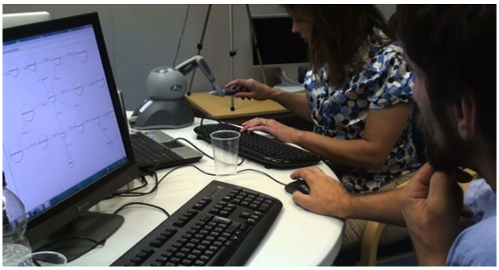


Fig. 2. Sighted and visually impaired colleagues editing a shared diagram

The tool combines a visual diagram editor with auditory and haptic capabilities to allow simultaneous visual and non-visual interaction. That is, two coworkers collaborate on shared diagrams by accessing and editing them through the visual modality (for the sighted user) and the combination of audio and haptic modalities (for the visually impaired user). We deployed this collaborative tool in various workplaces including a local government office and a charity organisation where visually impaired and sighted coworkers access and edit diagrams as part of their daily jobs [13]. In the following, we describe two examples of applying the activity pattern approach retrospectively to analyse data that we gathered from the field studies and identify potential patterns for cross-modal design.

Example 1: Consistency of Interaction Steps. In this example, a visually impaired manager (VI) and their sighted assistant (S) at a local government office edit an organisation chart to reflect recent changes in managerial structures. At one point during the interaction, the pair decides to create a connection between two nodes on the chart diagram to highlight a relationship between an existing and a new position. They do this while discussing how the tool should be used to create this relationship.

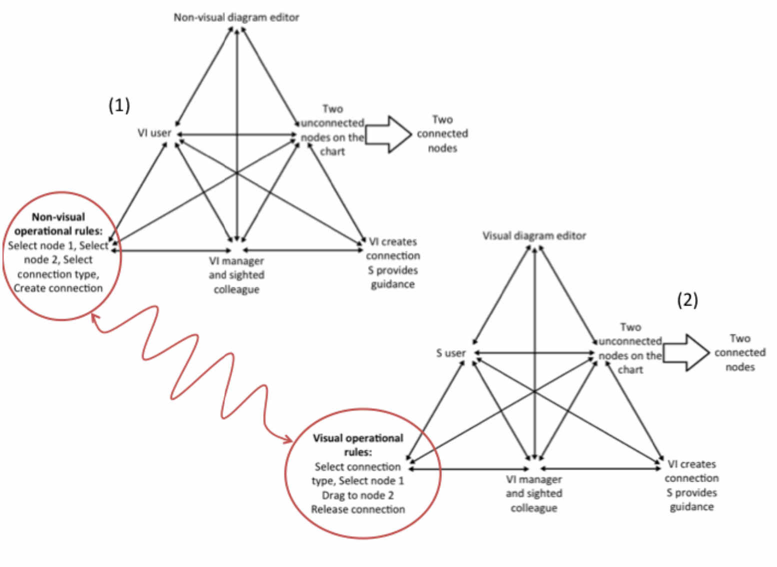


Fig. 3. Two activity systems for creating a connection between two nodes using the non-visual (1) and the visual (2) editors. Contradictions are highlighted with red circles and wavy arrows.

To create a connection between two nodes using the non-visual audio-haptic editor, the visually impaired user must 1) browse the chart to locate the first node and select it, 2) browse the chart again to locate the second node and select it, 3) select the type of connection they wish to use and 4) issue a command to create the desired connection. To do the same in the graphical editor, the sighted user must 1) select the type of connection they wish to use from the graphical tool bar, 2) select the first node on the chart, 3) drag the connection towards the second node using the computer mouse, and then 4) release the mouse to create the connection. Following the activity patterns approach, the actions of creating a connection between two nodes using the visual and non-visual editors can be represented as the two independent activity systems shown in Figure 3, which highlights a contradiction between the operational rules in the two activity systems; there is a mismatch between the interaction steps that each collaborator has to follow in order to create a connection between two nodes on the chart. Modelling the collaborative action of creating a connection in this manner has therefore uncovered a potential design flaw – which manifests itself as contradictions – that could hinder collaboration. Addressing this design flaw could lead to a design pattern that can eliminate the issues raised by the contradictions. For instance, the potential design pattern could describe the need to reconcile the two mediating artefacts in this context by ensuring equivalence and consistency of interaction steps between the visual and non-visual modalities.

Activity Patterns in the Requirement Phase. While we used the activity patterns approach to drive retrospective analysis of data gathered from evaluation studies, we used it in a second instance to drive the design process from its initial requirement phase. In a second project, we conducted a participatory design workshop (PD) with visually impaired audio producers, engineers and musicians in order to understand how they go about their work and what kind of difficulties they come across. Participatory design activities generate a huge amount of data and patterns could help with the process of organising the themes that emerge from this data. Here too, the concept of contradictions can be a useful guide to identify the tensions that exist in activity systems constructed to model the scenarios described by workshop participants, and hence could lead to insights about design solutions that could resolve such contradictions.

Fig. 4. Participatory design workshop with visually impaired and audio producers

Scenario 1: Exchanging Formats between Colleagues. In a first scenario, while discussing his experience of working with sighted colleagues and clients, a visually impaired producer described his frustration with inaccessible plugins on digital audio workstations. The visually impaired producer explained how his work with sighted colleagues often involves exchanging projects back and forth in order to complete sub tasks involving the manipulation of audio captured using inaccessible formats or requiring interaction with inaccessible audio plugins. In some cases, these accessibility issues have led to his exclusion from potential collaborative projects because the standard formats used are not readily accessible or would take too long to work with.

Scenario 2: Accessing a Visual Programming Language. In a second scenario, a visually impaired participant who specialises as an accessibility trainer described a similar experience with inaccessible visual tools. In this case, the issue was specific to working on projects that were coded using an audio programming language known as Max/MSP[1], which is a visual programming language that uses diagrammatic representations as its main programming components. The visually impaired participant highlighted how inaccessible such programming languages are even though they are used to code audio, which could be considered a natural working modality for visually impaired individuals.

Both scenarios above could be captured by the activity system shown in Figure 6. Here, there is a clear contradiction between the subjects of the activity, i.e. the visually impaired audio producer and accessibility trainer, and the tools available to them as mediating artefacts in the context of their activities. Capturing these contradictions allows us to think about possible design solutions to eliminate them, which could eventually lead to the development of fully articulated design patterns that embody such solutions. For instance, the diagrammatic nature of the Max/MSP programming language could be made accessible using the potential design pattern described in Example 2 above, in which the spatial properties of a diagram are translated into hierarchical structures that support non-visual inspection and navigation.

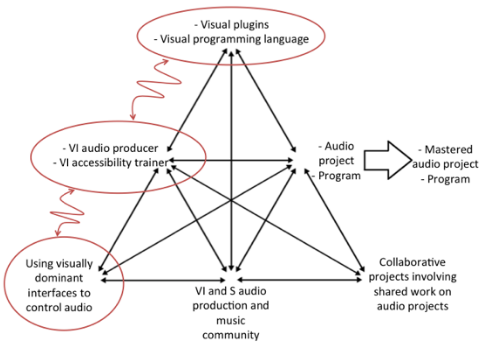


Fig. 5. An activity system showing some contradictions captured through the PD workshop. Contradictions are highlighted with red circles and wavy arrows.

We chose the above examples and scenarios to demonstrate how the potential design patterns they uncover could be structured around different hierarchical levels of activity. Each design pattern falls within one level of the hierarchy; example 1 was concerned with the operational rules for connecting two nodes on an organisational chart thus modelling activity at the *operational* level; and both scenarios 1 and 2

[1] See: `http://cycling74.com/products/max/`

relate to a higher level *activity* involving the organisation of collaboration between sighted and visually impaired stakeholders in sound production activities and the impact of currently available tools on such organisation. Patterns can therefore be written for different levels of activity and structured into a coherent pattern language, which can facilitate both the creation of new patterns and navigation of a patterns' space to help readers find their way through the design knowledge material.

5 Conclusion and Future Work

The design of cross-modal collaborative systems presents a unique set of challenges because such systems must allow individual users to equally contribute to the shared tasks while accommodating their individual perceptual differences. To date, no research has examined how to capture the knowledge required to design technology that makes cross-modal collaboration easier. Design patterns and pattern languages have a potential to facilitate the capture, presentation and communication of design knowledge. We proposed that designers and application domain experts in cross-modal collaboration could benefit from using design patterns as a uniform representation for expert knowledge. To this end, we explored the question of how potential patterns can be uncovered from an iterative design process and suggested that activity patterns could be used as a structured method to address this question. We also described how the activity patterns approach matches the domain in which individuals with differing perceptual abilities work together to achieve a common desired outcome through the mediation of a cross-modal tool.

One of the key benefits of using activity patterns to identify patterns is the conceptual tool of contradiction, which can be a useful guide for designers to identify the tensions that exist in their designs when used in context and modelled as activity systems. We exemplified how this approach was useful for us in the evaluation phase of a cross-modal collaborative tool as well as a technique for managing requirement data from participatory design workshops. We plan to use this approach to both generate and articulate an initial set of patterns to form a pattern language for designing cross-modal collaboration, which we will then validate by applying the patterns in future design iterations and incorporating them in future participatory design activities.

References

1. Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language: Towns, Buildings, Construction. Center for Environmental Structure Series. Oxford University Press, New York (1977)
2. Arvola, M., Larsson, A.: Regulating prominence: A design pattern for co-located collaboration. In: 6th Inter. Conf. on the Design of Cooperative Systems, COOP 2004, pp. 115–130. IOS Press (2004)
3. Bertelsen, O.W., Bødker, S.: Activity theory. In: HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science, pp. 291–324 (2003)

4. Borchers, J.O.: A pattern approach to interaction design. In: Cognition, Communication and Interaction, pp. 114–131 (2008)
5. Coad, P.: Object-oriented patterns. Communications of the ACM 35(9), 152–159 (1992)
6. Dearden, A., Finlay, J., Allgar, E., McManus, B.: Using pattern languages in participatory design. In: Binder, T., Gregory, J., Wagner, I. (eds.) Proc. of PDC 2002, Palo Alto, CA (2002)
7. Dearden, A., Finlay, J.: Pattern languages in HCI: A critical review. Human–Computer Interaction 21(1), 49–102 (2006)
8. Engestrom, Y.: Expanding learning at work: Toward an activity theoretical reconceptualization. J. of Education and Work 14(1), 133–156 (2001)
9. Erickson, T.: Lingua Francas for design: sacred places and pattern languages. In: Proc. of the 3rd Conf. on DIS: Processes, Practices, Methods, and Techniques, pp. 357–368. ACM (2000)
10. Guy, E.S.: Appropriating patterns for the activity theory toolkit. In: Inter. Workshop on Activity Theory Based Practical Methods for IT Design, ATIT 2004, Copenhagen, Denmark, September 2-3 (2004)
11. Martin, D., Rodden, T., Rouncefield, M., Sommerville, I., Viller, S.: Finding patterns in the fieldwork. In: Proc. of the 7th Conf. on ECSCW, pp. 39–58. Kluwer Academic Publishers (2001)
12. McGookin, D., Brewster, S.A.: An initial investigation into non-visual computer supported collaboration. In: CHI 2007: Extended Abstracts on Human Factors in Computing Systems, pp. 2573–2578 (2007)
13. Metatla, O., Bryan-Kinns, N., Stockman, T., Martin, F.: Supporting Cross-modal Collaboration in the Workplace. In: BCS HCI 2012, pp. 109–118 (2012)
14. Tidwell, J.: Common Ground: A patterns language for Human-Computer Interface design (1999), http://www.mit.edu/~jtidwell/interaction_patterns.html (accessed January 2013)
15. Winberg, F., Bowers, J.: Assembling the senses: towards the design of cooperative interfaces for visually impaired users. In: CSCW 2004, pp. 332–341 (2004)

Cross-Communicability: Evaluating the Meta-communication of Cross-Platform Applications

Rodrigo de A. Maués and Simone Diniz Junqueira Barbosa

Informatics Department, PUC-Rio
Rua Marques de Sao Vicente, 225/410 RDC
Gavea, Rio de Janeiro, RJ, 22451-900, Brazil
{rmaues,simone}@inf.puc-rio.br

Abstract. Evaluating cross-platform systems is challenging due to the different constraints and capabilities of each platform. In this paper we extend the Semiotic Inspection Method (SIM), a Semiotic Engineering evaluation method, to evaluate cross-platform systems. We introduce the term cross-communicability to denote the quality of the meta-communication of the system as whole, taking into account the user traversal between the different platforms. To assess cross-communicability, we describe a novel approach to conduct the SIM, which introduces a contrastive analysis of the designer-to-user meta-communication messages of each platform, based on a semiotic framing of design changes initially proposed for End-User Development. The results from an analytical study indicate that this approach is capable of identifying and classifying several potential communication breakdowns particular to cross-platform systems, which in turn can inform the design or redesign of a cross-platform application.

Keywords: Cross-platform, user interface design, communicability, semiotic inspection method, semiotic engineering.

1 Introduction

People are increasingly using applications and services that span a wide variety of computing devices, such as PCs, smartphones, tablets and digital TVs [1, 2, 3, 4, 5]. However, developing such systems can be a challenge because it is necessary to know not only how but also when to incorporate features available in one platform into the others, since each platform may differ in its capabilities and constraints (*e.g.*, screen size and resolution, input mechanisms) [4, 5, 6, 7]. Hence, it can be difficult and costly to maintain the quality of the system across each platform.

Evaluating cross-platform systems is as challenging as designing them [1, 4]. Traditional methods for assessing quality attributes of systems need to be adapted for the cross-platform context. The evaluated quality of separate parts (*i.e.*, system versions on each platform) does not necessarily correspond to the quality of the whole cross-platform system [2, 4]. People alternately use the "same" application in different platforms, frequently switching from one to the other. When traversing between system

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 241–258, 2013.

versions of a cross-platform system, a person should be able to reuse his or her knowledge of the available functions and of how tasks are performed. The usability aspects related to these traversals are introduced in [2] as inter-usability.

Recent works have investigated the development [6, 7, 8, 9] and the evaluation [2, 5, 10] of cross-platform systems by focusing mostly on their cognitive aspects, such as (inter) usability [2, 4, 10] and user experience [5]. In this paper, we shift the focus from evaluations based on cognitive theories to one based on Semiotics [11, 12]. Thus, instead of assessing the system's usability, we assess its communicability [12, 13], as described in the next section.

In this paper, we introduce the concept of cross-communicability as the system communicability across platforms. To evaluate cross-communicability, we use a semiotic framing originally proposed for End-User Development [15] to extend the application of the Semiotic Inspection Method (SIM) [12, 14], enriching the evaluation results and discussion.

This paper is organized as follows. We begin by presenting the theoretical background for this work. We then present the details of the application of the extended SIM to evaluate cross-platform systems and the concept of cross-communicability. Next, we describe our analytical study and present the results. Finally, we present our conclusions and future works.

2 Background

2.1 Semiotic Engineering Basics

Semiotic Engineering [11, 12] is a reflective and explanatory (as opposed to predictive) theory of human-computer interaction (HCI) that focuses on communicative rather than cognitive aspects of HCI analysis and design. It views the user interfaces of interactive systems as meta-communication artifacts, *i.e.*, through the user interface, designers[1] convey to the users their understanding of who the users are, what they know the users want or need to do, in which preferred ways, and why. Users then unfold and interpret this message as they interact with the system. This meta-communication message can be paraphrased as [12]:

"Here is my understanding of who you are, what I have learned you want or need to do, in which preferred ways, and why. This is the system I have therefore designed for you, and this is the way you can or should use it in order to fulfill a range of purposes that fall within this vision." (p.25)

The designer-to-user message is comprised of signs [11, 12, 16]. A sign is anything that stands for something else, to somebody in some respect or capacity [17]. Signs compose one or more signification systems that arise from culturally (and, in the case of HCI, also artificially) encoded associations between content and expressions [11, 12, 15]. For example, words and images typically come from signification

[1] In this paper designers should be interpreted as whoever speaks for the design team of a given application.

systems that exist in a culture outside the specific context of HCI, whereas mouse pointers belong to signification systems that are native to computer applications.

Semiotic Engineering classifies the signs in three different types [12, 16, 18]: static, dynamic and meta-linguistic. Static signs express and mean the system's state. They are motionless and persistent when no interaction is taking place (*e.g.*, icons, text areas, buttons at a given moment, menus). Dynamic signs express and mean the system behavior and emerge during interaction. They represent the transitions between the system states (*e.g.*, a save button becomes enabled after entering the name of a client in a registration form). Meta-linguistic signs refer to other interface signs and are used by the designer to explicitly communicate to users the meanings encoded in the user interface and how they can be used (*e.g.*, instructions and explanations, error and warning messages, hints). They are most commonly present in help systems, online/offline documentation, user manuals and promotional materials.

Based on this theoretical framework, the quality of a user interface is given by its communicability, which is "the system's property to convey effectively and efficiently to users its underlying design and interactive principles" [12, 13]. On the one hand, when a user can comprehend how the system works because the designer expressed himself properly through the user interface (communicability), it becomes easier to learn how to use it (usability) [11, 12]. On the other hand, when the user fails to understand the communication intended by the designer, a communication breakdown takes place that may hinder or even preclude the use of the system [11, 12]. Semiotic Engineering has two qualitative and interpretive methods to assess the communicability of a user interface [11, 12, 16]: the Communicability Evaluation Method (CEM) [12, 13], which involves user observation, and the Semiotic Inspection Method (SIM) [12, 14, 16, 18], which involves user interface inspection. While CEM focuses on the reception of the meta-communication message by the users, SIM focuses on its emission by the designer. Our paper extends SIM to cover the first part of the communication process – the emission of the message – in cross-platform systems.

2.2 The Semiotic Inspection Method

The Semiotic Inspection Method (SIM) [12, 14, 16, 18] is a qualitative inspection method grounded in Semiotic Engineering that allows the evaluation of the communicability of a computer system through the analysis of its static, dynamic and meta-linguistic signs. The goal is to identify communication breakdowns that may take place during the user-system interaction and to reconstruct the designers' meta-message conveyed by the user interface. Like other inspection methods (e.g. Heuristic Evaluation [19] and Cognitive Walkthrough [20]), SIM does not involve the participation of users [11]; it is the evaluator who examines the interface in search of ambiguities and inconsistencies in the signs chosen by the designer. Thus, as in any theory-based method, the more knowledge the evaluator has, the better the evaluation results will be.

Heuristic evaluation and cognitive walkthroughs focus on identifying usability issues, mostly related to the ease of use and learning of the system operations, while the strategic aspects of the system (the logic underlying the design decisions) conveyed by the user interface are not covered. Instead of assuming users always adopt a plan-based problem-solving method, semiotic engineering is more aligned with Suchman's work on situated action [21], which suggests that users are constantly reevaluating their current circumstances and adapting their "plan sketches" via communicative actions performed at/through the user interface. By inspecting the user interface language, we are able to predict some of these breakdowns, which may lead to user interaction problems. Rather than relying on a cognitive plan-based analysis, our inspection draws on Linguistics and Semiotics, which provide us with clues on how people make sense of and negotiate meaning.

SIM requires a preparation phase, in which the evaluator defines the purpose of the inspection, performs an informal inspection by navigating through the system to define the focus of the evaluation, and finally elaborates the inspection scenarios. Next, the evaluator proceeds to the execution of the inspection. To properly execute the method, the evaluator must assume a "user advocate" position. The execution of the method is carried out in five distinct steps: (1) a meta-linguistic signs inspection; (2) a static signs inspection; (3) a dynamic signs inspection; (4) a contrastive comparison of designer-to-user meta-communications identified in steps (1), (2), (3); and, finally, (5) a conclusive appreciation of the overall system communicability. In the first three steps, the evaluator inspects the signs within the scope of the evaluation and reconstructs the meta-messages being conveyed by them at each level, filling out the template of the designer-to-user meta-communication and identifying potential communicability problems at each level. Because each sign level has distinct expressive possibilities, the generated messages may not be identical. Thus, in step (4) the evaluator contrasts the meta-messages generated in the previous steps and checks for real or potential inconsistencies or ambiguities among them. Finally, in step (5), the evaluator qualitatively assesses the system communicability by unifying the meta-communication messages and then generates an evaluation report.

2.3 Impermeability and Computer Manipulations of Signs

Before we proceed with the extended application of SIM to cross-platform systems, it is important to briefly present two semiotic concepts that will be used in our analysis: impermeability and computer manipulations of signs.

In [15] those concepts were used to provide a semiotic framing for End-User Development (EUD). Both are related to the identity of an application, which is an important matter for both EUD and cross-platform development. While for EUD the main concern is how to empower users to customize or extend a system without losing the designed system identity, the main concern of the designer of a cross-platform system is how to design a different version for each platform while preserving the system identity across them, so as not to hinder the alternate use of the various versions. By assuming that traversing between platforms bears resemblance to traversing between customized versions of a system, we propose to apply these concepts beyond the scope of EUD for the first time, to help in the semiotic inspection of cross-platform systems. It should be made clear that the designer did not necessarily make these modifications, but based on the differences between supposedly equivalent signs we can assume which manipulations are capable of transforming a system version into another. From the nature of the manipulation, we can identify potential cross-communication breakdowns.

The concept of impermeability [15] is related to the encapsulation of signs, *i.e.*, the sign meaning is in an atomic capsule and therefore it cannot be altered. Thus, the originally encoded meaning of impermeable signs is always preserved. Impermeable signs can be essential or accidental. Essential impermeable signs can only be used in monotonic manipulations, *i.e.*, those that do not destroy the basic meanings (identity) of the application. Accidental impermeable signs may not be changed either, but they may be subtracted from the application by means of an operation we call *pruning*. For instance, let us consider a "skip to next track" button. It is an impermeable sign and thus its meaning of skipping to the next track on a playlist should not be changed, otherwise it may confuse the user (*e.g.*, if the next track button is used to skip to the next playlist instead). If this sign is essential (*i.e.*, part of the system identity), it should also be present in any other platform. However, if it is not part of the system identity (*i.e.*, it is an accidental sign), it may be pruned in any platform.

From a computer symbol-processing point of view, two different kinds of modifications can be made to signs [15]: meaning-preserving modifications (types I, II, and III) and meaning-changing ones (types IV through VII).

Meaning-preserving manipulations affect only impermeable signs and preserve the application identity. Type I changes correspond to *renaming and aliasing* operations that affect only lexical components. Changing the label of a button and changing an icon appearance are examples of renaming. Renaming should be limited to lexical items that are not part of the application's identity to avoid confusing the user about its meaning. Type II manipulations involve making changes only to syntactic components, such as *reordering*: changing the layout sequence of user interface elements or the order in which operations are performed. However, this reordering cannot change the meaning expressed by the individual reordered components. Finally, a type III manipulation (also known as *pruning*) corresponds to the possibility of selecting non-essential (accidental) components to be included in or excluded from the application. Pruning must preserve not only the identity of the application, but also the impermeability of the pruned components. Thus, an important design decision is to select which of the impermeable signs can be pruned if the application identity is to be preserved.

When porting the system to a more constrained platform, the designer of cross-platform systems usually has to leave features out (type III – *pruning*) or adjust them according to the space available in the screen (type II – *reordering*). Also due to space constraints or specific platform patterns, the designer may have to change the appearance of some components (type I – *renaming and aliasing*).

Every application requires that users learn a new and unique signification system that is used in HCI [15]. Therefore, the application's identity and the notion of impermeability are important to keep the user's semiosis sufficiently tied to the designer's vision, so that productive interpretations can be motivated, and unproductive ones discouraged [15]. This is especially important when it comes to cross-platform systems, since you may have many different versions of the same system.

Meaning-changing manipulations, however, can threaten the application identity, and therefore should be avoided whenever possible in order to minimize conflicts between system versions. A type IV manipulation, for example, involves using existing signs to mean something different from what they were designed to mean (*e.g.*, the same button triggers different actions on each platform). Hence, when transitioning from one version to another, users may feel frustrated because of this conflict of meanings. A type V manipulation makes changes to meanings and lexical items but not to grammatical structures. Type VI may involve reordering components or even eliminating components that change the conveyed meaning. And finally, type VII makes changes to meanings, grammatical structures and lexical items, consisting of the biggest threat to the application identity.

In our analytical study, we give examples of some of the manipulations we found and discuss about their potential consequences concerning cross-platform systems.

3 Using SIM to Evaluate Cross-Communicability

When evaluating each version of a cross-platform system separately, each one may have high communicability, *i.e.*, it may efficiently and effectively convey to the users its underlying design intent and interactive principles. However, when traversing between platforms, the user carry understandings from one version that may not be applicable to another, creating conflicts and communication breakdowns. Besides, the meta-communication message of each version should also convey the composition of the cross-platform system, *i.e.*, how functionalities are organized in and across platforms [5]. The quality of the meta-communication conveyed by the cross-platform system as a whole, including all versions, is what we call *cross-communicability*.

The traditional SIM alone is not suitable for evaluating cross-communicability, since we must also systematically contrast the meta-communication messages across system versions. Hence, after the preparation phase required to the traditional SIM described previously, we divide the execution of the method applied in two phases: vertical (within-platform) analysis and horizontal (between-platform) analysis.

3.1 Vertical Analysis

The first phase, called vertical analysis, is designed to evaluate the communicability of the system in each platform individually. Hence, it consists of conducting the same five steps of the traditional SIM, with one difference: the evaluator should also highlight the signs that denote any compositional aspects of the cross-platform system (*e.g.*, a sign that acknowledges the existence of another system version).

Because the inspections must be conducted individually, different evaluators may conduct the inspection of each version. There are three main benefits of this approach: it distributes and reduces the cost of the inspection; it avoids contaminating the inspection of one version with another; and it allows designating each system version to the most appropriate evaluator, a professional with solid technical HCI knowledge in that platform.

3.2 Horizontal Analysis

As stated before, when evaluating a cross-platform system it is important to contrast the meta-communication messages of each system version. Thus, after completing the communicability assessment of each version, the next phase consists of performing a horizontal analysis, which is divided in three steps: (1) identification and collation of sign manipulations between the evaluated platforms; (2) an analysis of the impact of each manipulation on the meta-communication of each version; and finally (3) a conclusive appreciation of the quality of the overall system cross-communicability.

The signs collated in the vertical analysis may undergo modifications across the evaluated system versions due to the constraints, advantages or patterns of each platform. Thus, the evaluator should (1) identify the meaning-preserving and meaning-changing manipulations in each pairwise combination of the evaluated platforms; (2) examine the manipulations collated and categorized in the previous step to assess how they could negatively affect the horizontal meta-communication messages by intensifying already identified vertical communication breakdowns (e.g., causing ambiguities) or even by creating new ones; and (3) qualitatively assess the system cross-communicability by unifying the meta-communication message obtained in each previous step, judging the costs and benefits of the identified manipulations made between platforms, and then he generates an evaluation report.

4 Analytical Study

For the application of SIM on a cross-platform system, we chose an Internet radio and social networking website named 8tracks,[2] which revolves around the concept of streaming user-curated playlists composed of at least 8 tracks. Besides being a popular cross-platform system, 8tracks was chosen because it has different versions available on different devices (desktop and mobile) and also has different versions available on the same device (mobile native apps[3] and mobile web app), therefore covering a broad range of combinations to analyze in our study.

4.1 Preparation Phase

The goal of this analysis is to identify different communication breakdowns that may occur individually on each platform version of a cross-platform system and their relation to other breakdowns that arise when these versions are put together. We defined the context of use for the evaluation through the following scenario:

"Rodrigo will throw a party and needs to set a good playlist, but has no idea how to start. Gabriel, Rodrigo's friend, suggested he use 8tracks system, where he can discover and listen to several different mixes. Rodrigo had already created an account in the past, but he has never really used the system. He thinks this might be the perfect opportunity to give it a try. He then searches 8tracks for mixes with the latest hits.

---

[2] `http://8tracks.com/`

[3] 8tracks has two different mobile native apps available, an iOS and an Android version, but only the Android version was evaluated in our study.

He saves his favorite mixes, to be able to play them later at the party, if he so wants. He also creates his own mix only with the favorite songs from the mixes he has played. So as not to forget which ones they are, he will use another feature that allows saving only individual favorite tracks to be bought later, instead of saving full mixes."

Although it was observed by a previous informal inspection that not all the functionalities in the scenario would be available in all the evaluated system versions, we decided to let it be this way to evaluate how each version conveys information about the system composition as well. Next, we briefly present the main results from the vertical and horizontal analyses.

4.2 Vertical Analysis

A single evaluator conducted the study. Due to space constraints, in this paper, we do not describe the whole vertical analysis of the 8tracks versions, which consists of a typical application of SIM. Instead we present synthesized versions of the meta-message, focusing on some communication issues found when analyzing the system communicability on each platform. Although the vertical analysis was conducted separately for each platform, some references will be made to similar characteristics across the versions to avoid repetition and also due to space constraints.

Desktop Web App. Besides the hints and tips throughout the system, there are many internal pages (*e.g.*, about, licensing, FAQ) and a blog available where some meta-linguistic signs can be found. This documentation mainly explains the system purpose and its underlying concepts, instead of how to use it. Therefore, it is implied that the system is in general thought to be easy to use and to learn. The available explanations describe many aspects of the legal rules that the service must follow to allow the free streaming of music. These rules limit some functionalities of the system, but expecting that users will probably not want to read the documentation before using the system, the main rules are only briefly explained through dynamic messages to let users become aware of them by trial and error when interacting with the system.

Some aspects of trial and error, however, can be frustrating due to the lack of appropriate communication. When attempting to skip to a track one too many times, for instance, the following message is shown: “Apologies for the inconvenience, but our music license requires us to limit the number of tracks you may skip each hour.” Since there is no static or dynamic sign indicating how many tracks you have already skipped or may still skip, or how long it has been since you exceeded the number of “skips,” and also because the skip button is always enabled, the user may often try to skip a track, fail, and then receive the infamous message over and over again.

The meta-communication message derived from the documentation and help content tells us that the service is designed for music-lovers that want to create, share and discover new mixes in a radio-style way, *i.e.*, like a non-interactive music player. Static signs on the mix page (Figure 1) appropriately convey this radio-style music discovery experience depicted in the meta-linguistic signs: the music progress is represented by a very thin and discreet line, indicating that it is not so important and that users should not worry about it, just like on the radio; there is no bar over this thin line to control the music progress; and the only control buttons available are a button to play/pause the music, a button to skip to the next track or to the next mix.

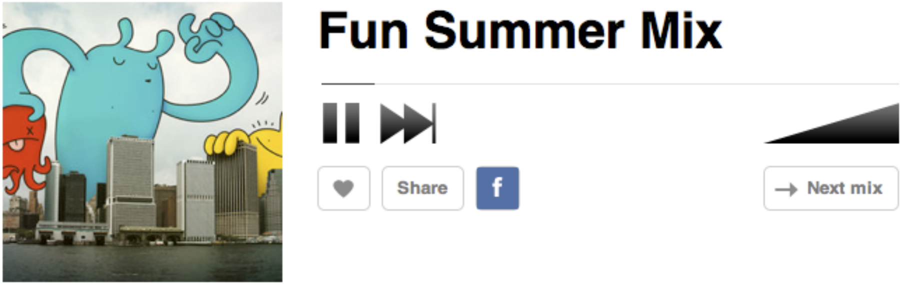


Fig. 1. Part of the desktop web version's mix page

Regarding the search for mixes, a communication breakdown was found. Once a search has been done, the user may sort the results by hot, new or popular. Although it is easy to infer the outcome of sorting by newest mix, the same cannot be said about the other two criteria, and there is no explanation on the system about the difference between popular and hot mixes.

Meta-linguistic signs on some of the internal web pages of the system and on the 8tracks blog almost exclusively express the information regarding the system composition. References are made to both the mobile web and the mobile native versions. There is also a static sign at the bottom of each system page, a button, referring to the mobile native version.

Mobile Web App. Besides the fact that the meta-linguistic signs available on documentation of the desktop web version discussed previously are also available to any other system version, in the mobile web app there is a direct link to the desktop web app's "FAQ" and "About" pages, as if they were its own. Because of this and the fact that very little information about the mobile web version can be found in the system documentation, it can be assumed that this version has the same features of the desktop version. However, during the analysis of the static signs it became clear that not all the functions depicted in the meta-linguistic signs are available in this version. Hence, if a user consults the documentation he may be misled and frustrated after searching the whole app for a function that does not even exist in it.

Although it is harder to present hints on a mobile web system because it is not possible to hover over user interface components, in general there are really few messages and instructions in this platform, possibly meaning that the designer believes that this interface is also fairly easy to use and no further explanation is necessary.

In addition to the same meta-linguistic signs available at the desktop version, in this version two static signs can be found on the bottom of the dashboard page referring to both the mobile native and the desktop web versions.

Another communication breakdown was found on the system search mechanism. Although there is a tip on the kind of text the user should type to perform a search, in the results page there is no indication of which criteria were used to search the mixes. Hence, the user does not know whether the text he typed was matched with the mixes tags, with the mixes names, with another criterion or with all of them. Another potential breakdown found on the results page is the lack of static or meta-linguistic signs evidencing what happens when the user selects one out of the three options (most recent, hot this week or popular this month) on the search results page. Even with the aid of the dynamic signs, there is no clear evidence as to whether you are sorting or filtering the search results by one of these criteria.

Users can also directly search mixes by selecting tags in the dashboard. Although the static signs indicate that you can search only by one tag at a time, it is actually also possible to search by a combination of tags if you perform a long press on each desired tag at a time, and then you press a go button that will appear next to the tags (Figure 2). However, since this is not explained to users, it is a communicability breakdown, and users may never perceive this search possibility even if they intensely interact with the system through trial and error.

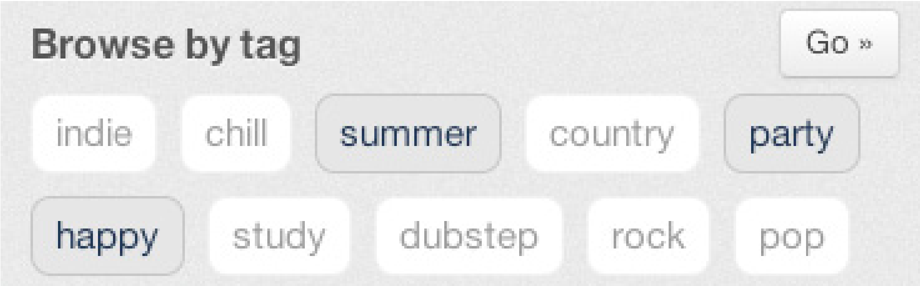


Fig. 2. Example of the mobile web version of a search where the user combined multiple tags as the search criteria

Mobile Native App. Although there is no static sign referring to the other versions, this version contains an interesting new meta-linguistic sign regarding the system composition on its own internal FAQ: it clearly states that, if you want to create a mix or preview the songs you have added to your favorites, you need to use the desktop version. Hence, this sign not only recognizes the existence of another version but it also indicates the system distribution of functionalities, revealing more clearly the redundancy across these system versions. Nevertheless, there is no reference about the mobile web version on this platform.

There are additional meta-linguistic signs available on the app's Google Play page and also on an Update Log page internal to the app, clearly explaining the purpose, functionalities and limitations of this version. Both pages also display the changes made to each release of the app, which is important because it helps the user to interpret new features or modifications made to previously known signs.

The static signs indicate that the user can filter the mix search results by tags, but since this label is also next to the hot, new and popular buttons, the user may also wrongly infer that these buttons can filter the results (Figure 3). However, when inspecting the dynamic signs we notice that the result count is not affected; therefore these buttons only sort the results. Still, a distracted user may never notice this dynamic sign, since it is quite subtle.

This version also introduces a new sign: a hidden left-hand menu, used to hide navigation options until someone taps a button to expose them (Figure 3). This off-canvas pattern is very common and useful in mobile applications, since it avoids occupying the small screen of the device all the time. However, the way this menu is accessed in this platform introduces a temporary failure [12]: it uses the sign of a home button, which naturally would make the user infer that he is going to a home or dashboard page of the system, which does not exist in this version. Some other problems previously discussed on the inspection of the desktop version were also found on this version, such as the hot/popular ambiguity.

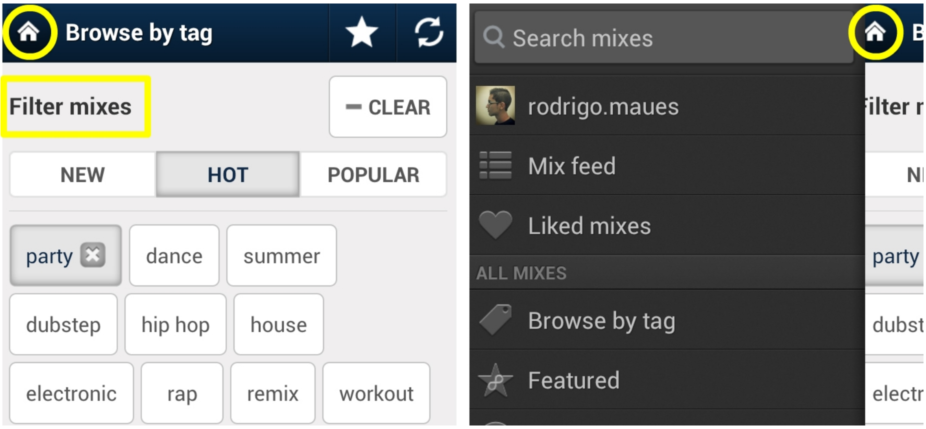


Fig. 3. Search results filtering page before (on the left) and after (on the right) pressing the home button, which actually triggers the hidden left-hand menu to appear on the screen. The highlighted signs are responsible for different vertical and horizontal breakdowns.

4.3 Horizontal Analysis

Having concluded the vertical analysis, we conducted the horizontal analysis to assess 8tracks's cross-communicability. First, we contrasted each evaluated version with the two others, resulting in six tables containing the modifications that could have been made at the user interface signs in going from one version to the other, classified by the nature of the manipulation. Manipulations from type I to V were identified, but most of them were meaning-preserving ones. Due to space constraints, only a sample of them is presented, classified by type to discuss the main communication breakdowns found across versions and to illustrate the benefits of this analysis.

It is worth mentioning that, in a cross-platform system such as 8tracks, *i.e.*, a system whose functionality and content is available in multiple devices, many of the unchanged characteristics found across the platforms may indicate the identity of the system and some of its impermeable signs. For instance, the control buttons on the mix page, the buttons for liking the mix and for sharing the mix (Figure 1) are present with their meanings unchanged in all the platforms. This indicates that they probably are some of the essential impermeable signs and part of the system's identity. However, these relations are still not so clear.

Type I Manipulations. Only a few type I (*renaming*) manipulations were found: the label "Created" on the mobile versions is written "Published" on the desktop web version (Figure 4); there is a slight difference between the desktop and the mobile versions of the mix feed icons (Figure 5); and the label "recent" in the mobile web version is written "new" in the other versions (Figure 6). Although they may not lead the user to a failure, those unnecessary inconsistencies should be avoided.

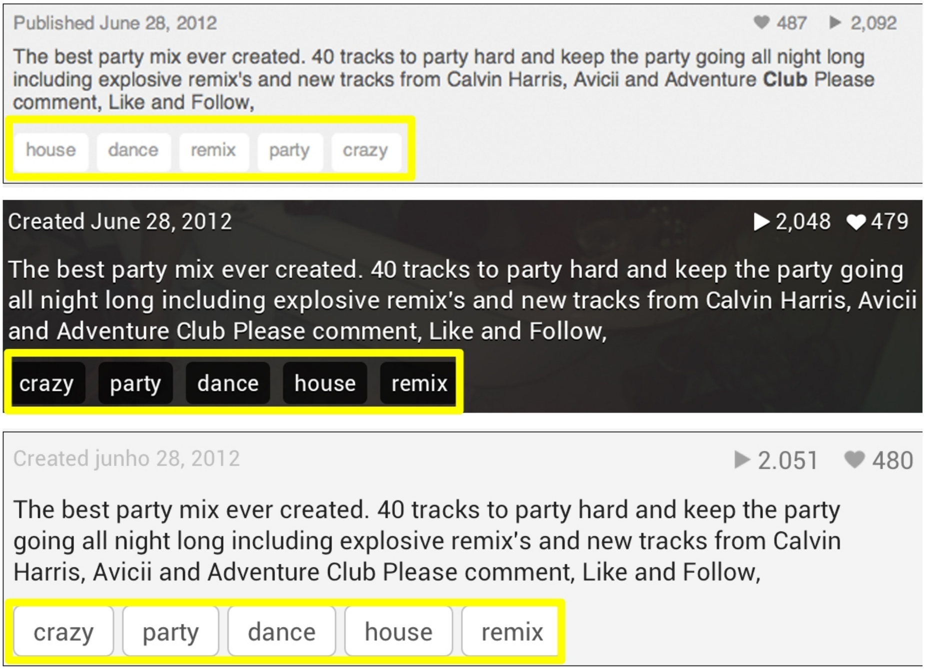


Fig. 4. From top to bottom, the mix description appearance on the desktop web, mobile web and mobile native versions, respectively. The highlighted tags are not clickable only in the mobile web version.

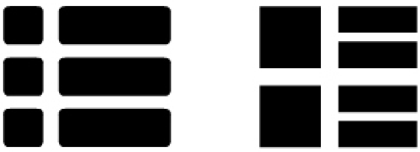


Fig. 5. Example of type I manipulation and unnecessary inconsistency: difference in the appearance of the mix feed icon used in the mobile versions (on the left) and in the desktop web version (on the right)

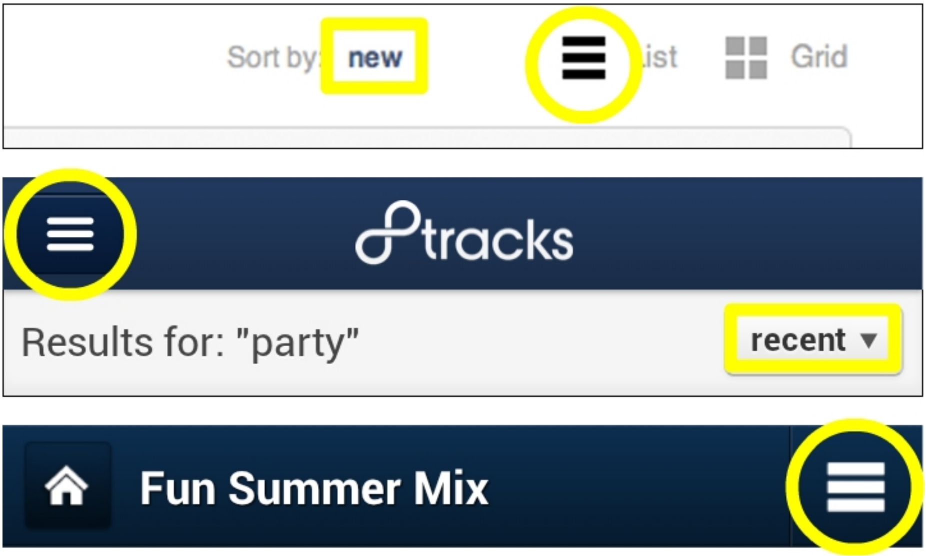


Fig. 6. From top to bottom, part of the search results page on the desktop and mobile web versions and part of a mix page on the mobile native version, respectively. The highlighted elements reveal type I (rectangle) and type V manipulations (circle).

Type II Manipulations. Type II manipulations (*reordering*) are most common and inevitable when it comes to cross-platform systems working in different devices. This happens because the display size and resolution of each device force designers to reorganize the layout to fit all the interface elements, or even to separate elements into two or more screens to avoid pruning some essential impermeable signs. Hence, it is inevitable to break the interface layout consistency to fit everything.

However, it is possible to find on 8tracks some type II manipulations that resulted in unnecessary inconsistencies. For instance, in the mix description (Figure 4), the labels for the mix like and play counts are displayed in one order in the desktop version and in the reverse order in the mobile versions, even though there are no space constraints. Although the latter example may not constitute a serious communication breakdown, it may negatively affect some cognitive aspects. For instance, a person who is already used to the order of the labels for mix like and play counts in a platform will have to learn the order of the elements again in the other platforms. If instead of labels the interface elements were buttons, for example, the same reordering would represent a bigger issue, since a distracted person might press the wrong button because the order (that the person was already used to) has changed.

Another reordering problem found was that the same groups of mixes are arranged in the desktop web in a different order from the mobile web. Since the order of elements may also convey a relation of importance between them, there is an ambiguity on what the designer thought was more important to the user: the desktop web prioritizes the users' mixes (*i.e.*, their own mixes, liked mixes and mix feeds from whom he follows) while the mobile web prioritizes general featured and hot mixes.

Type III Manipulations. After the vertical analysis, the evaluator already has a clear vision of which features and components are present or not in each version, which makes it easy to identify type III (*pruning*) manipulations. Based on the collated type III manipulations, it became evident that the versions are complementary, *i.e.*, although they share a set of data and functions, some of them provide access to data or functions that are not available in the others.

Some of the observed pruning operations from the desktop to the mobile versions may be justified by either reduced platform capabilities or little utility. For instance, the functionality of creating mixes is not available in the mobile versions, maybe because of either platform limitations (slower internet connection when using 3G to upload each track and limited processing power in comparison to a desktop); or little utility, since the user will be using the system mostly on the go and therefore will not be interested or even capable of spending too much time looking at the device screen. Thus, the 8track's designer may have intended that the users should only listen in their mobile devices to what had been produced on their desktop computers.

In the desktop version, since some mix arts may contain inappropriate content to be shown during the service use in a public or work environment, the user may flag some arts as Not Safe For Work (NSFW) and configure them not to be shown. This feature is not available on the mobile versions, probably because Smartphones are highly personal; therefore there is little utility in avoiding displaying any type of inappropriate content, since only the user would see it on his device.

Type III manipulations can indicate some interesting communication breakdowns across platforms, because the lack of some features between certain versions may represent an ambiguity for the user. In our study, one example is the feature of adding a track to the favorites. To do so, the user must press the track's star button placed either next to where the current track is being shown or on the tracklist, which shows only the mix tracks already played. The vertical analysis of the mobile web version revealed that, on the tracklist page (Figure 7), there is only a button to buy each track and you can neither see the tracks you have added to your favorites nor add them; therefore you can only add a track to your favorite (or know that it has already been added to your favorite tracks) while it is still playing. When contrasted to the desktop and the mobile native versions, this issue is amplified, because the user in the desktop version can buy the track or add it to the favorites using buttons on the tracklist, and in the mobile native the user can only add but not buy the track in the tracklist (Figure 7). Therefore a person used to these versions may get frustrated when trying to add a track to the favorites on the mobile web version after he has listened to the whole song, only to discover that he cannot do it. Since in 8tracks the user cannot go back to a previous track and the tracklist is shuffled each time the user listens again to a mix from the beginning, he would be forced to go to another version to see the full tracklist so far and then add the track to his favorites. Since pruning the button to buy a track from the tracklist does not represent a problem, but removing the star button (to add a track to the favorites) does, it is possible to infer that the tracklist and the star button together make up an impermeable sign, and therefore should not be separated.

Fig. 7. From top to bottom, a part of the tracklist page on the desktop web, mobile web and mobile native versions, respectively. The highlighted signs are responsible for a horizontal communication issue.

Still regarding pruning manipulations, removing the dashboard page on the mobile native version per se would not lead to a communication problem across platforms. However, confusion arises because there is a similar home button that shows the left-hand menu (Figure 3) instead of leading the user to a dashboard page (Type IV manipulation – changing the meaning of a sign), which happens in the other platforms.

Pruning manipulations can also be observed on the 8tracks meta-linguistic signs. Adding a new FAQ to the mobile native version results in interesting observations. First, its benefit to the mobile native version is clear, once it avoids the same communication breakdown as the mobile web version regarding the confusion about which functionalities are available or not. Second, the mobile native's FAQ contains a piece of information about the purpose of the favorite tracks feature (*i.e.*, save tracks to buy them later), which is missing from the meta-linguistic signs in the desktop version. Hence, if a user learns about this feature from the native app, he may understand it better than another user who inferred its behavior by its name on the desktop version.

A similar situation can be observed regarding the feature of sorting search results by hot, new or popular. As mentioned before, a common breakdown in the desktop and mobile native versions is the lack of clarity on the difference between the hot and popular criteria. However, if a person has used this feature on the mobile web version, he can eliminate this doubt, since the labels there are more detailed: "hot this week" and "popular this month". In the mobile web version the breakdown regards not knowing whether a sorting or a filtering operation was applied to the results after selecting one of the criteria. This problem can be either avoided or transformed into another issue, depending on which version was used prior to the mobile web one. If the user has already used the desktop version, he will know the results are sorted by one of those criteria. If the user has had previous experience only with the mobile native, he may mistakenly assume a similar behavior, i.e., that the results were filtered (Figure 3). Finally, contrasting the desktop and mobile versions altogether may lead to another breakdown, since the user may no longer be sure whether each version is behaving differently or similarly, and in case they are indeed doing the same operation, they might not know whether it is sorting or filtering the results.

The latter breakdown is probably due to one of two reasons: an unfortunate type I manipulation, because as stated before you should not change a label if you are changing an impermeable sign, and therefore you mislead the user into thinking that something means something else (sort *versus* filter); or a type II manipulation that went wrong, *i.e.*, the label in the mobile native app was referring only to the tags but was misplaced next to the sorting options as well.

Type IV Manipulations. Regarding meaning-changing manipulations, one type IV manipulation (only the meaning of a sign is changed) is responsible for a potential issue when the user selects a mix to go to its page: in the desktop and mobile web apps, when entering the mix page, the mix does not start playing right away (the user needs to press a play button), but the same is not true in the mobile native app. The fact that the mix will start playing once selected in the latter version may frustrate a person used to the other meaning of selecting a mix on the web versions of the system, and vice-versa.

Type V Manipulations. On the mobile web version, a button (Figure 6) on the header of the search results page (which triggers the interface to go to the home or dashboard page) can cause a breakdown because it does not resemble a home button and therefore its meaning cannot be easily interpreted or even learned by the user. When contrasting this version with the mobile native and desktop web versions, this button can cause additional breakdowns, because a very similar sign is used in those versions (Figure 6) with different meanings. Thus, we have the occurrence of a type V manipulation (meanings and lexical items are changed).

In the mobile native the meaning of this sign is to display the mix tracklist, while in the desktop web the meaning is to display the search results in a list (instead of in a grid). The same sign did not create a vertical communication problem in the desktop web and mobile native versions; it created a vertical communication problem in the mobile web version; and it also created a horizontal communication breakdown, since the user may get confused about which action is associated to the button on each platform, leading to temporary failures.

Another type V manipulation is responsible not for a new communication issue, but for amplifying a breakdown identified during the mobile version's vertical analysis: the tags on the mix description (Figure 4) resemble buttons, but they are not actually clickable. Since in the desktop web and mobile native versions the same tag buttons (although different in color) are clickable (triggering a search by a tag when clicked), this manipulation may lead the user to frustration, since he is likely try to click on the tags on the mobile web version if he is already used with the other system versions.

5 Conclusion and Future Work

This paper described a first effort to apply Semiotic Engineering and its methods to the evaluation of the quality of cross-platform systems, which we defined as cross-communicability. We presented an analytical study using a new approach to SIM to assess how well the designer-to-user meta-communication message gets across to users through each system version separately and also through the cross-platform system as a whole, taking into account the possible ambiguities and communicative breakdowns when traversing between versions.

The proposed evaluation of the cross-communicability of a system consists of two analyses: a vertical analysis, which is mostly an application of the SIM to each system version; and a horizontal analysis, in which the signs and meta-messages from the previous analyses are contrasted and the differences are classified according to the type of sign manipulations that would be necessary to transform one version into another. This classification is based on a semiotic framing originally proposed for EUD.

We could identify through our analytical study some interesting facts when contrasting the meta-messages of the different system versions: ambiguities and new communication breakdowns may arise in one or more system versions; individual breakdowns may be transmitted from one version to another; individual breakdowns in one version may be amplified by individual characteristics or issues of other versions; and also some individual breakdowns in one version may be overcome due to signs present on the other versions.

Thus, we believe that by looking at the results of a SIM evaluation informed by our extended approach, the designer of cross-platform systems may become more aware of several communicative breakdowns regarding composition, continuity (*i.e.*, how tasks state and actions are migrated across platforms) and consistency aspects, which can help them to better design or redesign each system version. The proposed sign organization by manipulations may also help to improve existing cross-platform development processes, as in the graceful degradation [6] method, for instance. Results from a horizontal analysis could serve as a basis to refine the set of transformational rules applied during the graceful degradation approach, as well as to validate the priority ordering between the rules in the method.

As future work, we plan to triangulate the results obtained with endogenous and exogenous sources. We also plan to conduct a study comparing the extended SIM with more traditional inspection approaches, such as heuristic evaluation and cognitive walkthrough. Also we intend to investigate more closely the relation between the system's identity and the quality of a cross-platform system, as well as how we can properly define which signs are impermeable and essential to one or more platforms, and to the system as a whole. We also plan to formulate a set of heuristics as a result of consistently applying SIM to multiple cross-platform applications. And finally, we plan to extend CEM to cross-platform systems as well.

References

1. Antila, V., Lui, A.: Challenges in designing inter-usable systems. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 396–403. Springer, Heidelberg (2011)
2. Denis, C., Karsenty, L.: Inter-usability of multi-device systems - a conceptual framework. In: Seffah, A., Javahery, H. (eds.) Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces. Wiley & Sons (2004)

3. Paternò, F.: Designing multi-device user interfaces: how to adapt to the changing device. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 702–703. Springer, Heidelberg (2007)
4. Seffah, A., Forbrig, P., Javahery, H.: Multi-devices "Multiple" user interfaces: development models and research opportunities. Journal of Systems and Software 73(2), 287–300 (2004)
5. Wäljas, M., Segerståhl, K., Väänänen-Vainio-Mattila, K., Oinas-Kukkonen, H.: Cross-platform service user experience: a field study and an initial framework. In: Proc. MobileHCI 2010, pp. 219–228. ACM (2010)
6. Florins, M., Vanderdonckt, J.: Graceful degradation of user interfaces as a design method for multiplatform systems. In: Proc. IUI 2004. ACM (2004)
7. Lin, J., Landay, J.A.: Employing patterns and layers for early-stage design and prototyping of cross-device user interfaces. In: Proc. CHI 2008, pp. 1313–1322. ACM (2008)
8. Ghiani, G., Paternò, F., Santoro, C.: On-demand cross-device interface components migration. In: Proc. MobileHCI 2010, pp. 299–308. ACM (2010)
9. Paternò, F., Zichittella, G.: Desktop-to-mobile web adaptation through customizable two-dimensional semantic redesign. In: Forbrig, P. (ed.) HCSE 2010. LNCS, vol. 6409, pp. 79–94. Springer, Heidelberg (2010)
10. Serna, A., Calvary, G., Scapin, D.L.: How assessing plasticity design choices can improve UI quality: a case study. In: Proc. EICS 2010, pp. 29–34. ACM (2010)
11. Barbosa, S.D.J., Silva, B.S.: Interação Humano-Computador. Campus-Elsevier (2010)
12. de Souza, C.S.: The Semiotic Engineering of Human-Computer Interaction. The MIT Press (2005)
13. Prates, R.O., de Souza, C.S., Barbosa, S.D.J.: Methods and tools: a method for evaluating the communicability of user interfaces. ACM Interactions 7(1), 31–38 (2000)
14. de Souza, C.S., Leitão, C.F., Prates, R.O., da Silva, E.J.: The Semiotic Inspection Method. In: Proc. IHC 2006, vol. 1, pp. 148–157. SBC (2006)
15. de Souza, C.S., Barbosa, S.D.J.: A semiotic framing for end-user development. In: Lieberman, H., Paternò, F., Wulf, V. (org.) End User Development: People to Flexibly Employ Advanced Information and Communication Technology, pp. 401–426. Springer (2006)
16. de Souza, C.S., Leitão, C.F.: Semiotic Engineering Methods for Scientific Research in HCI. Morgan & Claypool (2009)
17. Peirce, C.S.: The essential Peirce (Vols. I and II). In: Houser, N., Kloesel, C. (eds.). Indiana University Press (1992)
18. de Souza, C.S., Leitão, C.F., Prates, R.O., Bim, S.A., da Silva, E.J.: Can inspection methods generate valid new knowledge in HCI? The case of semiotic inspection. International Journal of Human-Computer Studies 68(1-2), 22–40 (2010)
19. Nielsen, J.: Heuristic Evaluation. In: Mack, R., Nielsen, J. (eds.) Usability Inspection Methods, pp. 25–62. John Wiley & Sons (1994)
20. Wharton, C., Rieman, J., Lewis, C., Polson, P.: The Cognitive Walkthrough Method: a Practitioner's Guide (1994)
21. Suchman, L.: Plans and situated actions: The Problem of Human-Machine Communication. Cambridge University Press (1987)

On-Line Sketch Recognition Using Direction Feature

Wei Deng, Lingda Wu, Ronghuan Yu, and Jiazhe Lai

Academy of Equipment, Beijing 101416, China
dengwei@whu.edu.cn

Abstract. Sketch recognition is widely used in pen-based interaction, especially as the increasing popularity of devices with touch screens. It can enhance human-computer interaction by allowing a natural/free form of interaction. The main challenging problem is the variability in hand drawings. This paper presents an on-line sketch recognition method based on the direction feature. We also present two feature representations to train a classifier. We support our case by experimental results obtained from the NicIcon database. A recognition rate of 97.95% is achieved, and average runtime is 97.6ms using a Support Vector Machine classifier.

Keywords: Sketched symbol recognition, NicIcon database, multi-stroke shapes.

1 Introduction

Sketching is a natural form of human communication. Sketch-based interaction is a fast and efficient means of capturing information by automatically interpreting hand-drawn sketches and can be an important part of the early design process, where it can help people explore rough ideas and solutions in an informal environment. Sketch recognition refers to the recognition of predefined symbols or free-form drawings (e.g., an unconstrained circuit drawing); in the latter case, the recognition task is generally preceded by segmentation in order to locate individual symbols. This paper focuses on the recognition of hand-drawn isolated symbols. With the growing popularity of devices with touch screens, there is increasing interest in building sketch-based user interfaces.

However, many challenges remain in terms of intra-class compactness and inter-class separation due to the variability of sketching. Because it is likely that different people will have different drawing styles, such as the stroke-order, -number, and non-uniform scaling, as well as complex local shifts. Moreover, the style of the same individual at different times may differ. A good recognition algorithm should place few drawing constraints on users. Related research is that of handwriting recognition, such as handwritten digit and Chinese character recognition with many effective algorithms.

This paper presents a new approach to sketched symbol recognition based on the direction feature. It is insulated from stroke order and number. We also present two feature representations, which obtain good experimental results in the NicIcon database.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 259–266, 2013.

2 Related Work

One common approach to sketch recognition focuses on building structural shape descriptions. Its basis step is stroke segmentation and primitive recognition using temporal and spatial features. A sketched symbol can be represented as a tree or graph and the similarity between two sketches can be calculated by structural matching. Hammond and Davis [1] developed a hierarchical language to describe how diagrams are drawn, displayed, and edited. They used this language to perform automatic symbol recognition. The attributed relational graph (ARG) is an excellent statistical model to describe both geometry and topology of a symbol [2], and is insensitive to orientation, scaling, and drawing order. The advantage of structural methods is distinguishing similar shapes. Therefore, high accuracy in stroke segmentation and primitive recognition is necessary.

An alternative approach looks at the visual appearance of shapes and symbols [3] instead of the relation between geometric primitives. Ouyang and Davis[3] proposed a visual approach to sketched symbol recognition, which is similar in spirit to our proposed method. It used a set of visual features that captured on-line stroke properties like orientation and endpoint location. Almazan et al. [4] and Escalera et al. [5] described a framework to learn a model of shape variability based on the Active Appearance Model (AAM), and proposed two types of modified Blurred Shape Model (BSM) descriptors(DBSM and nrBSM) as basic shape features to learn the model. Eitz et al. [6] described a large scale exploration of human sketches. They analyzed the distribution of non-expert sketches of everyday objects and developed a bag-of-features sketch representation. Willems et al. [7] explored a large number of on-line features, which were sorted in three feature sets due to different levels of details. Recently Delaye and Anquetil [8] presented a set of 49 features, called HBF49, for the representation of hand-drawn symbols for use as a reference for evaluation of symbol recognition systems.

This paper presents an on-line sketched symbol recognition method using the direction feature. The key contributions of our method are:

1. A framework to recognize sketched symbols using the direction features. It is insensitive to the variation in the order and number of strokes.
2. Two feature representations. They are statistical vectors rather than structural representations of geometric primitives. This allows it to be robust to different drawing styles.
3. Obtaining high recognition accuracy in the NicIcon database using Support Vector Machine (SVM) classifier in real time.

3 Our Approach

We designed our approach primarily using the direction properties of symbols. An overview of the recognition process is shown in Fig. 1. First, preprocess the input trajectory, including resampling. Then the direction of every point is decomposed to four directions, horizontal, vertical, diagonal, and anti-diagonal. Feature representation is performed using the grid and direction feature. Finally, an SVM classifier is trained to perform classification.

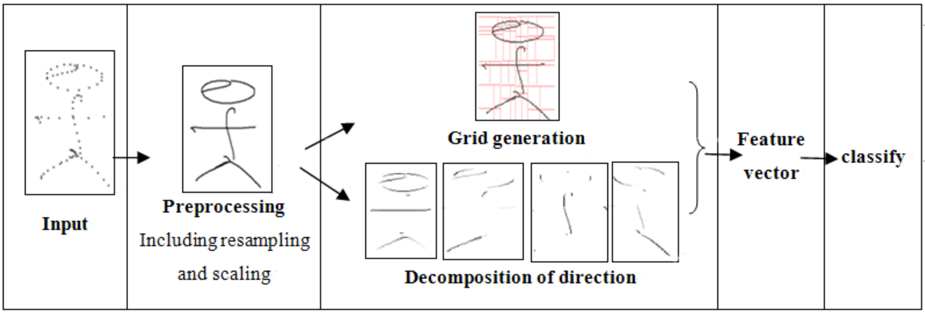


Fig. 1. System overview

3.1 Preprocessing

The preprocessing of the input sketch directly facilitates the pattern description and affects the quality of the description. We choose to normalize patterns in scale and translation, and to apply a trajectory resampling strategy. These operations are simple to perform and guarantee a better stability of extracted features, for any type of input pattern.

Since on-line strokes are typically sampled at a constant temporal frequency, the distance between neighboring points varies based on the pen speed. This produces more sampled points where the pen is typically slower. In order to make feature extraction more reliable, we resample each stroke at a constant spatial distance. In our experiments the sampling interval is set to one. Next we remove differences of scale and translation. The coordinates of stroke points are simply shifted and scaled such that all points are enclosed in a standard box. In experiments we set $x, y \in [0,100]$. It means translating maximal dimension of a symbol to 100 with aspect ratio preserved [10]. We do not translate the on-line trajectory to image (such as in [3]). Therefore direction features are accurate with the absence of jagged edge.

3.2 Decomposition of Direction

The so-called direction feature is widely used in character recognition. Similar to Ouyang and Davis [3], we decompose direction vectors of resampling points into four directions, at 0, 45, 90 and 135 degrees. If the direction of a point lies between two standard directions, it is decomposed into two components in the two standard directions, as shown in Fig. 2. One major advantage of this representation is the independence of stroke direction. A stroke drawn left to right is the same as one drawn right to left. This is essential for free hand-drawn symbols.

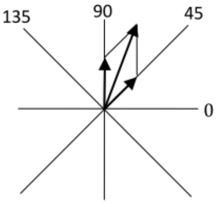
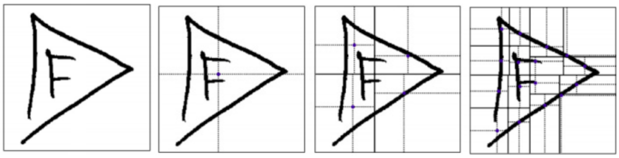


Fig. 2. Decomposition of direction



(a) (b) (c) (d)

Fig. 3. The procedure of grid generation [4]. (a) original shape points. (b) four sub-regions split by geometric centroid. (c) and (d) iterative procedure to obtain grids of 4x4 and 8x8 respectively.

3.3 Grid Generation

Grid is used to blur the features, for directly overcoming the local non-uniform variability. We learned that blurring should be considered as a means of obtaining latent dimensions(subspace) rather than as a means of reducing computational cost, though the effects might seem similar [9]. Moreover, the mesh size of 8x8 used in statistical approaches was determined by the optimum blurring parameter in light of the Shannon sampling theorem [9]. We refer to the grid generation by Almazan et al.[4]. First, we calculate the geometric centroid of all resampling points. Then the region is split into four sub-regions. Therefore, the points are split into four sets. Next, in each sub-region the iterative procedure works until we get a grid of size 8x8, as shown in Fig. 3.

3.4 Feature Representation

The representation schemes of the input sketch and the model database are important since the classification method depends strongly on them. We present two feature representations. The size of the grid is 8x8, and the sub-direction feature is $S=\{x_i, y_i, dir_{1i}, dir_{2i}, dir_{3i}, dir_{4i}\}$, i=1,2,…,n, where n is number of resampling points; $\{x_i, y_i\}$ is coordinate of point P_i; dir_{1i}, dir_{2i}, dir_{3i} and dir_{4i} describe sub-direction strength of P_i in above four directions, respectively. One simple feature representation is to cumulate sub-direction strength in each grid cell. That is

$$V=\{v_j\}\,,\;\; v_j=\{\sum_{p_i\in m_j} dir_{1i}\,,\sum_{p_i\in m_j} dir_{2i}\,,\sum_{p_i\in m_j} dir_{3i}\,,\sum_{p_i\in m_j} dir_{4i}\}$$

where i=1,2,…,n , j=1,2,…,8^2; and m_j is the j^{th} cell in the grid, P_i is the i^{th} point in sampling points, v_j is direction feature of m_j. So we can form a 8x8x4 dimensional vector, which can be used in a statistical classifier. We denoted this representation as 'Dir' , compared with below 'DirBSM'.

Another feature representation is presented, called Directional Blurred Shape Model (DirBSM), is presented here. BSM[4,5] is a shape descriptor , whose main idea is to describe a given shape by a probability density function encoding the probability of pixel densities of a certain number of shape sub-regions. Given an image of a particular symbol, each pixel contributes to compute the BSM descriptor. This is done by dividing the given image in a kxk grid (cells). Then, each cell receives votes not only from the shape pixel located inside its corresponding cell, but also from those located in the adjacent cells. Thereby, every pixel contributes to the density measure of its sub-region cell, and its neighboring ones. This contribution is weighted according to the distance between the point and the centroid of the cell receiving the vote.

However, in our approach the shape is not an image but a set of shape points with four direction attributes. We have modified the BSM descriptor to DirBSM. For each direction attribute, we calculate a BSM vector, as illustrated in Fig. 4. Then four vectors can form a final shape descriptor, which is the feature representation of a symbol.

Input: A grid of $k \times k$ cells and $S = \{x_i, y_i, dir_i\}$, where $dir_i \in \{ dir_{1i}, dir_{2i}, dir_{3i}, dir_{4i} \}$, $\forall i = 1,2,\ldots,n$

Output: the BSM of one sub-direction

1. Denote grid as $R = \{r_1, ..., r_{k^2}\}$, with c_j the centroid for grid cell r_j.
2. Denote $N(r)$ as the neighbor cells of cell r, Let r^p be the cell which contains the point p. Initialize the probability vector v as $v(j)=0$, $j = 1, 2, ..., k^2$.
3. For each point $p(x_i, y_i)$

    $D=0$

    For each $r_h \in N(r^p)$, $(h \in \{1, 2, ..., k^2\})$

        $d_h = d(p, r_h) = \| p - c_h \|^2$

        $D = D + 1/d_h$

    End for

    Update the probability vector v as $v(r_h) = v(r_h) + dir_i / (Dd_h)$

  End for

4. Normalize $v(h) = v(h) / \sum_{j=1}^{k^2} v(j)$, $\forall h = 1,2,\ldots,k^2$

Fig. 4. Getting BSM vector of one direction feature

To modify the feature distribution for improving the classification performance, all the measurements in the feature vectors are transformed by variable transformation $y=x^p$ [10]. In our experiments we tested different values of p to obtain optimal parameter.

4 Experiments

In this section, we show the performance of the proposed method for sketch recognition tasks using above two feature representations.

4.1 NicIcon Database

The NicIcon database [11] is composed of 26,163 sketched symbols of 14 classes in the domain of crisis management and incident response systems. An example is shown in Fig. 5. The database comes from 34 different writers and it is commonly used for on-line symbol recognition, but off-line data is also available. The database is already divided into three subsets (TRAINING, TEST and EVALUATION sets) for both writer dependent and independent settings. Approximately, 36%, 24% and 40% of symbols are contained in the training, test and evaluation sets, respectively. We have selected the on-line data that is writer-independent, because it is the most similar to human-computer interaction.

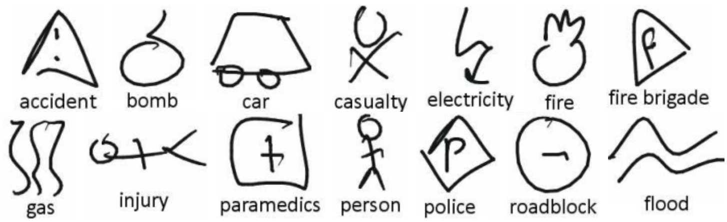


Fig. 5. Samples of 14 different classes of NicIcon database

We compare two kinds of feature representations (Dir and DirBSM). Training samples are used to train a classifier and test samples are used to test recognition accuracy. We select the LibSVM toolbox with default parameters.

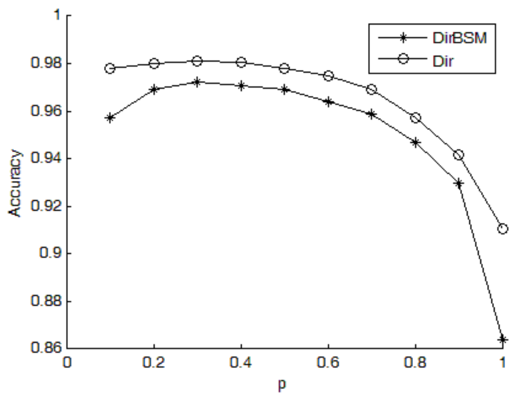


Fig. 6. Getting optimal p using training set and test set

In order to get optimal meta-parameter p, we train a classifier using the TRAINING set and test its performance using the TEST set. Fig. 6 shows the accuracy with parameter p. The optimal p is p=0.3. So all the measurements (both in traing and testing) in the feature vectors are transformed by variable transformation $y=x^{0.3}$ [10]. Then we use both the TRAINING and the TEST sets for training the classifier and the EVALUATION set for measuring the recognition performance (T+V/E). The accuracy rates are 95.57% (DirBSM) and 97.95%(Dir) , respectively.

4.2 Results and Discussion

In Fig. 6, when p is small, the accuracy of Dir is better than that of DirBSM. The reason for this may be that the BSM descriptor blurs the shape and cannot distinguish shapes in details.

As Table 1 indicates, on this dataset we achieved an best accuracy rate of 97.95%., compared to other results in the literatures [4,7,8].

Table 1. Accuracy compared with other methods

Methods	Accuracy
(a) combination of features[7]	92.63% *
(b) NRAM+nrBSM(SVM)	94.29% *
(c) NRAM+nrBSM(NN)	95.18% *
(d) HBF49[8](SVM)	97.44%
our approach(SVM)	95.57 (DirBSM) 97.95 (Dir)

*The values of accuracy rate in (a), (b) and (c) come for reference[4]. And the experiments in (b) and (c) used off-line data in NicIcon database.

For improvements in the recognition system, we analyzed misrecognition factors in the experiments. The top four false-positive symbols are accident, bomb, paramedics and electricity. It means there are other symbols that were misrecognized as them. And the top four false-negative ones, which were misrecognized as other symbols, are fire, casualty, police and fire brigade. Factor-1 is the large distortion in sketching. Some of these sketches cannot even be recognized by people, as shown in Fig. 7. Factor-2 is the presence of extra strokes or missing ones. Factor-3 is the similarity in direction features of sketches, mainly between the symbols gas and flood. Some distorted symbols are similar to different patterns, when they are represented by direction features. These may be resolved in terms of introducing other useful features.

(a) (b) (c) (d) (e) (f)

Fig. 7. Misrecognized samples. (a) too large distortion; (b) and (c) distortion of triangle becoming a rectangle;(d) and (e) extra stroke; (f) distortion of rhombus becoming a trapezium

4.3 Runtime Performance

Finally, we evaluate the runtime performance of our approach. We use Matlab to program this method on a 3.10*2 GHz machine (Inter Core™ i5-2400). The average runtime of recognizing one symbol is 97.6ms, including extracting feature and classifying with SVM-RBF. This method should therefore realistically be usable in real-time human-computer interaction.

5 Conclusions

This paper presents an on-line sketch recognition method based on the direction feature. It is a statistical method, which uses the visual feature of shape points instead of regarding the shape as an image or temporal sequences of geometric primitives. We also present two feature representations to train an SVM classifier. Experiments show that in the NicIcon database fine results are achieved, including high accuracy rate and real-time effects. This method is insensitive to stroke order and number. In future, we will do more experiments using larger databases with more classes, and introduce more feature to enhance accuracy rate.

References

1. Hammond, T., Davis, R.: LADDER, a sketching language for user interface developers. Computer & Graphics 29(4), 518–532 (2005)
2. Lee, W., Kara, L.B., Stahovich, T.F.: An Efficient Graph-Based Recognition for Hand-Drawn Symbols. Computers & Graphics 31, 554–567 (2007)
3. Ouyang, T.Y., Davis, R.: A visual approach to sketch symbol recognition. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009) (2009)
4. Almazan, J., Fornes, A., Valveny, E.: A non-rigid appearance model for shape description and recognition. Pattern Recognition 45, 3105–3113 (2012)
5. Escalera, S., Fornes, A., et al.: Blurred shape model for binary and grey-level symbol recognition. Pattern Recognition Letters 30, 1424–1433 (2009)
6. Eitz, M., Hays, J., Alexa, M.: How do humans sketch object? ACM Transactions on Graphics. In: Proc. SIGGRAPH (2012)
7. Willems, D., Niels, R., Gerven, M.V., et al.: Iconic and multi-stroke gesture recognition. Pattern Recognition 42, 3303–3312 (2009)
8. Delaye, A., Anquetil, E.: HBF49 feature set: A first unified baseline for online symbol recognition. Pattern Recognition 46, 117–130 (2013)
9. Fujisawa, H.: Forty years of research in character and document recognition—an industrial perspective. Pattern Recognition 41, 2435–2446 (2008)
10. Liu, C.L., Nakashima, K., Saho, H., et al.: Handwritten digit recognition: investigation of normalization and feature extraction techniques. Pattern Recognition 37, 265–279 (2004)
11. Niels, R., Willems, D., Vuurpijl, L.: The NicIcon database of handwritten icons. In: 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), Montreal, Canada, August 19-21 (2008)

Beyond Rhetoric to Poetics in IT Invention

Annie Gentes[1] and Ted Selker[2]

[1] Telecom ParisTech, 46, rue Barrault, 75634 Paris, France
[2] Carnegie Mellon University Silicon Valley, Mountainview CA
annie.gentes@telecom-paristech.fr, ted.selker@sv.cmu.edu

Abstract. Two kinds of discourse typically define scientific productions: logical (epistemology of science) and rhetorical (sociology of science). We suggest that research projects can also be analyzed as poetical productions. While rhetorical strategies anticipate controversies and deploy techniques to defend projects and findings, poetical practices deepen the cultural and symbolic dimensions of technologies. Based on use cases that show different ways the poetics come to bear on research and development projects in information technology (IT), we discuss the play on words and images and how they contribute to the definition and creation of a new technology within research projects. Three cases of poetical practices are presented: naming technologies, christening projects, and designing logos. We give examples of naming and project identity formation to underscore how such a poetic stance impacts projects. Images and words help people imagine what the technology is about by giving imaginary traits and cultural substance. This paper's analysis is a call for further work exposing the value of conscious use of poetical approaches to deepen the framing of IT projects.

Keywords: Rhetoric, Poetics, Project Definition, Design, Sociology of Science, Sociology of Technology, Invention.

1 Introduction: From the Rhetoric to the Poetics of Science

Science endeavors to create and test hypotheses [1]. Scientific contribution is then based on logical demonstrations that distinguish it from social and other common knowledge discourses. The epistemologies of sciences detail the methods of validation of what they build as facts, either true or false. Inductive and deductive methods are discussed. But science is also about strategies to defend a territory, to plead for a cause, to affront enemies, to gather followers. This strategic part of science has been extensively analyzed by the sociology of science after Latour [2]. Actors build a full-fledged rhetorical strategy to support their project. The sociology of sciences and technologies [3] has shown that making science implies defending it and therefore elaborating a rhetoric that is going to convince the stakeholders. The father of rhetoric, the Greek philosopher Aristotle, describes rhetoric as an art of persuasion that every citizen should learn so as to debate about things in the city. In society, things are being discussed and judged. Winning the audience is the focus point of rhetoric. Latour's contribution has been to analyze the form and role of scientific rhetoric.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 267–279, 2013.

In particular, he shows that scientific papers are built around a defensive strategy that finds allies as a way to prevent dismissal. He describes how researchers use rhetorical arguments from theoretical background, "enrolling" other papers that have been written so as to prevent attacks. Latour speaks of "stratification" [2] that is to say the building of a system of authority within the paper. This system also excludes those researches that do not follow the same direction. Research and scientific activities are therefore described like a court of law, where participants build up the defense of their case. For Latour, this rhetorical strategy makes it difficult for scientists to produce certain types of discourses. In particular, he points out that popularization is difficult because scientific texts are written to exclude people not to include them. What is more, scientific controversies need strong positions that do not leave space for dreams or imagination.

However, certain texts (including images like drawings, photographs and videos) that are produced in the course of a research project cannot be totally described by either their deductive or inductive building of facts, nor their rhetorical function. Why call a technology Bluetooth after a king of Denmark? Why choose a cartoon character – Popeye - to represent an adhoc mobile communication research project? And why tell all kinds of stories when one might expect to communicate technical information by concentrating on algorithms and material implementations? We suggest the following answer: research done to study an area and engineering done to create technologies are not only based on logical demonstrations, or by negotiation between actors, they benefit by being subjective and cultural associations**,** poetic activities related to humanistic and literary practices too. They are using and working on symbolic representations that expand the observations and definitions that researchers obtain within their methodological framework and constitute an alternative to the rhetorical inscription in society by a poetical invention of concepts and narratives. Though we describe practices that have been experienced by most researchers in HCI and engineering, these practices can be analyzed as an integral part of the research process rather than discarded as communication, or valorization practices. Our standpoint is that these practices that focus on the subjective and cultural associations need to be considered as an essential part of the technical research as an invention. We consider invention along Simondon's definition. The philosopher of science and technology defines invention by articulating imagination to cultural and social context through a tangible artifact [4].

This article contrasts the poetical and the rhetorical, and their role in the emergence of new technologies. We contend that engineers, designers and scientists expand their work with their literary and humanistic skills that we propose to call their poetic stance. Ignoring the poetic aspects of their work is detrimental to a clear understanding of how it comes about. It could also be detrimental to strategies of innovation that have to take into consideration the whole gamut of activities around invention and creation to make sure that the potential of technologies are explored.

Our goal here is to show the value of poetics in IT projects. We first start from the poetic stance from Aristotle's founding text and a recent body of literature on poetics. We then present several use cases that are samples of this poetic stance: naming of technology, christening of a project, and design of logo. The analysis of these use cases shows interactions between the poetic naming and narrative process (with words or images) and the projects themselves.

2 A Poetic Stance

Before analyzing some of these poetic productions of science, we propose to start on the basis of Aristotle's definition of "poetics" and how it has been developed in literary theory in order to understand what is at stake in these images and stories that use more fictional metaphors and concept visualizations than either fact based demonstrations or rhetorical arguments [5]. Then we proceed to support our argument for poetic practices with examples of these poetic productions.

Like rhetoric, poetics is an art of production. While over the centuries, poetics has been reduced to a collection of stylistic traits and can still be considered as such (the poetics of literary genres, like science fiction for instance) more recently, poetic studies have focused on the creative process itself and its specificity [6]. In particular, the linguist Jakobson proposes a model of language that takes into consideration its different functions. Among them, the poetic function of language refers to itself, in other words is its own material. It is opposed to the referential function of language that points to things [7]. Referential and poetic functions of language are further developed by semioticians, who distinguish denotation and connotation. Denotation is about the literal meaning and essentially descriptive while connotation is the subjective and cultural associations [8]. The poet and literary analyst Paul Valéry described this poetic stance happening when the purpose is not to use language as a persuasive means to create a social consensus but to transcend the here and then, because the writer or artist unfolds either latent or possible aspects of the world [9]. For Meschonnic too, specialist of poetry and translation, the creator questions what could be [10]. This production can be called prospective, not because it foretells the future but because it produces new meaning. Aristotle uses the word *Dunamis* or *dynamis* that means possibility or capability. It is not an abstract vision. The design work consists of processing, working on, and with a variety of materials: either language (styles, genres [11], or pigment or sound [12]. Playing with words, inventing stories, making images are not only there to teach, or to defend, or to demonstrate. They expand the way we think about the world. The artist taps into and reveals the potential of a medium to transform and interpret the world.

Such inventive practices within research projects have recently been the subject of social studies of scientific texts [13], in ethnography for instance [14], and of imaging and visualization. Scientific images, for instance, have been noted for their aesthetic ambition [15]. Sociologists of science, Burri and Dumit point out the "engagement" that images produce and in particular how they are made instrumental in the production of scientific knowledge [16] While Latour [2], Lynch and Woolgar [17] stress how images tend to objectify knowledge and to produce yet another argument in the rhetoric of science, we follow those who note the fecundity of images with respect to creativity and invention. Images in this respect are considered as epistemic creations [18].

While the theme has been thoroughly studied in different fields of science, we think that it needs further analysis for at least two reasons. First we need to identify the diversity of poetic productions - not only images of science, but a whole gamut of other productions like names and logos. Second, we need to question their relationships and their impact on the project.

3 Case Studies

3.1 Methodology

To study the poetic stance of inventors, designers and engineers we combined two methodological traditions. The authors participated and observed research projects from their inception to their completion and could therefore study the way names and narratives played a role in the maturation of the invention. These observations were eventually completed by in-depth interviews. These methods from social sciences, such as ethnographic observations and participations as well as interviews were complemented by methods from the humanities: literary analysis of texts, semiotic study of documents, analysis of media. These analyses focus on meaning-making as defined in semiotics [19], anthropology of material culture [20] and more recently design studies [21]. They study the references to certain literary or visual traditions that are used within videos, texts, photos. These analyses show how words and tales not only can describe the technology, but also carry the imagination of the technology rather than its facts, and in doing so influence the way researchers look at their projects.

Several aspects of this poetic process are exposed in this article: the question of naming the technologies in the view that they will become common names; the christening of projects with proper names, and the role of logos.

We defined a corpus of texts related to research projects that either author was taking part in. Semiotics scholars define a corpus as a group of texts (linguistic and visual) that belong to the same genre [17]. Texts of the same genre share structural and pragmatic characteristics in so far as they are not only similar in style, narrative structure, or visual organization, but also as they are discussed by social actors as targeting the same audiences, for the same purposes, with similar textual strategies. The constitution of the corpus does not pretend to be exhaustive. It is always but a segment of the actual production. The purpose is to gather enough material to answer a research question. The method therefore consists in defining a corpus of textual elements and considering what are the shared elements and their purpose in the text. In this particular instance, the authors selected only part of the poetic elements of the research practice". They focused the semiotic analysis on condensed forms of identity of the technology, of the group, through naming or representing by a logo over a diversity of situations: naming a field of research, a device, a communication norm. The purpose is to emphasize the similarities over different examples rather than to focus on the differences that should be further explored.

3.2 Scientific Puns: Play on Words and Definitions

In 1982, the three volumes AI Handbook attempted to present all of published AI projects of the time [22]. At that time the corpus consisted of 50 or so named research systems[1]. They were named with words or acronyms that punctuate the index with a variety of poetic heritages. WHY is a question answering system, BUGGY is a system to help students debug their understanding; some 30 such as WHY, BUGGY or Scholar use functional and denotative names. 20 more of the research project names such as CRYSTALIS or MYCIN are clearly connotative whimsical names words to make them memorable or celebrate some joke that have no bearing on the function of the programs. 8 of the names such as SOPHIE could be easily given a person. As this simple retrospective of the AI handbook shows, first there is a variety of naming options, and second many of them are visibly not pushed by purely descriptive purposes.

More generally, some words like 801.11 describe the way the technology came into being. Some words like "wi-fi" for supposedly "wireless fidelity" are the result of brand work. Others like " mesh networks" rely on metaphors that seems to relate to a physical network. Still other names completely drop out from the technical description to refer to an apparently arbitrary choice as the term "Bluetooth". In, this article, the central methodology follows the ethnographic approach of identifying concepts from an analysis of specific example projects. Here we study three examples that seem to belong to different naming strategies: “Bluetooth” as a private joke within a community, “Artificial Intelligence” as relating a technology to artificial beings in literature, “TrackPoint” as renaming in a lineage of products.

Bluetooth. Bluetooth as a name could have come from many places around the time of its creation. Blue lasers and LED’s were the high-tech discoveries that were being used to sterilize things, to drive higher bandwidth on laser disks and optical fibers. But none of these references had the least to do with the short range Bluetooth radio standard. The term "Bluetooth" came out of meetings between researchers who were attending normalization meetings on the issue of short-range radio frequency. A leading actor of the "Bluetooth" norm, Jim Kardach, posted on his blog that the diversity of technical names was such that it led to use a code name from the culture of some of the participants [23]. The name King Harold Bluetooth of Denmark, who succeeded at unifying the various Danish regions in the 10th century, had emerged in discussions and post conferences.

---

[1] The 48 prominent AI research projects of the time are described in the1983 AI handbook [22]. The names of these research projects can be classified as, A: a person’s name, B: connotative/whimsical or arbitrary, or C: functional.
A. 8 of the 48 names could easily be used as a person’s name without notice:
DEDALUS, GUIDON, HAM, NOAH, PECOS, SOPHIE, TEIRESAIS, WUMPUS.
B. 20 of the names are connotative whimsical or arbitrary:
ACRONYM, CRYSTALIS, CASNET, DEDALUS, GUIDON, HAM, HARPY, IRIS, LIBRA, MYCIN, MEMOD, MOLGEN, NOAH, PECOS, SAFE, SOPHIE, STRIPS TEIRESAIS, WEST, WUMPUS.
C. 28 of the projects are named with functional or denotative names:
ABSTRIPS, ACT, Alpha beta, AM, AQ11, BUGGY, CHI, CONGEN, DENDRAL, EPAM, EXCHECK, EXPERT, HACKER, HEARSAY, HWIM, INTERLISP, LEX, Meta-DENDRAL, MACYMA, Minimax, NLPQ, PROSPECTOR, Protosystem, PSI, SCHOLAR, Student, The Programmers Apprentice, WHY.

"Bluetooth was borrowed from the 10th century, second King of Denmark, King Harald Bluetooth (King Harald I of Denmark) who was famous for uniting Scandinavia just as we (Intel, Ericsson, and Nokia) intended to unite the PC and cellular industries with a short-range wireless link." [21]

Rather than keeping complicated names or acronyms that were far too descriptive and technical the participants chose the name Bluetooth that started as a joke and finished as the expression for the norm....

"*…it became apparent the need to have a single name; as Intel would talk to people about "Biz-RF", Ericsson about "MC-Link", and Nokia about "Low Power-RF", which also created confusion…*" [21]

While it was not meant to be the final name, it ultimately ended up becoming the official name of the standard.

TrackPoint. Another example is the "TrackPoint" device. According to its inventor, Ted Selker, it was the subject of multiple debates. At IBM, several years were spent developing a way of moving a cursor without moving hands from the home keyboard position. The project was first called the "Jstick" for the use of the right index finger on the j key as a joystick. Some people picked up this solution, included it in a few laptop computers in the mid 1990's, and marketed it as "Jmouse".

Due to problems with discoverability, installation, button overloading and handedness, the device moved out from under a key to between the g and h, and got christened with a functional name: "pointing stick" in press release and presented at Interact1990 [24]. Various people proposed that if a 6 inch handle that users wiggled was named a joystick this little one could be named "joy nub". As IBM's design consultant Richard Sapper chose red for the device, many thought of calling it a clitoris or a nipple. Scientific naming conversations focused on an easy way to remember the descriptive "objective" name, but product naming conversations focused on a descriptive "subjective" name. In the objective case, its origin in some early experiments under the j had people calling it the "Jstick", the underlying technical system being an isometric joystick and its use as a moving a cursor or focusing on a point on the screen got it called a "pointing stick". In the subjective case, people described the appearance and the gesture and considered it as obviously erotic. As the device was about to go to product, the naming department of IBM gave it a different name which turned out to have a trademark problem after the launch team was already being trained. There were tense conversations with the naming person who thought naming this little feature was trivial as he had whole product lines and products to name… With only weeks till public communications were to begin, he chose a recycled name: TrackPointII. The name TrackPoint had been trademarked as a descriptive name for a mouse product that IBM made in 1991 [25] that could be turned over to be used as a trackball that a user rolled their hand over to move the cursor. The term track also changed the connotation: from a person's actions of rolling a ball to the person's goal of tracking down a point or pointing at the screen. In both devices, the "point" concentrates on the effect on the screen.

Artificial Intelligence. Finally, in interviews with Selker, the late Professor John McCarthy explained how he picked up the name for a new field: "Artificial Intelligence" in the 1950s [26]. McCarthy was a huge fan of science fiction throughout his life. In many conversations he would state that he had thought Artificial Intelligence (AI) would captivate people's imagination more than the other term he had been considering: "Machine Intelligence". Embedded in his choice of name was his love of science fiction but also the fact that he supposed that people would be more interested in exploring human intelligence as well as more capable machines and he wanted to connect the two. Instead of simply describing the work as making yet another engineering field, he liked the Artificial Intelligence expression for its poke at anthropomorphizing the quest.

3.3 Naming Projects: Acrobatic Acronyms

Technologies that are invented need to be given a name. Similarly, though for a shorter period of time, research projects are also baptized. There too, one can witness a variety of inspirations: eventually project participants may forget what the name of the project stands for. One of the reasons is that they are made of acrobatic acronyms. In fact, people often keep the acronym but change its explanation. Some names are more telling than the others: PLUG [2](Play Ubiquitous Games and Play More) that was chosen for an RFID mobile game sounded both relevant and evocative. "Transhumance[3]", "Safari[4]" that focused on adhoc networks, were suggestive of mobility, but did not give any real clue to how they were dealing with it. And of course, "Adam[5]"

or "Popeye"[6] are particularly obscure.Gentes could observe the naming process for a research project carried out by two departments of social sciences (Center for the Sociology of Innovation in Ecole des Mines and the Co-design Lab at Ecole des Telecom) as well as a legal research lab (CERNA). The team wanted to study the uses of peer-to-peer but also mobility and ad hoc networks. They wanted to understand how these technologies were spreading in our society. In fact, a good part of the discussion was about the object of research itself. For a part of the team, peer-to-peer was a good starting point, but it focused on applications, while the other members wanted to focus on infrastructures. Finally, the common denominator between the two groups was the fact that they were interested in Distributed Architectures. That sealed at least 2 letters of the acronym that the team was looking for: AD. But Distributed Architectures is a much broader field than a project around communication could own. Starting with AD, they played on a number of combinations based on the sonority and the connotations of the words: AD + mobiles information: ADMIN that gave the project the connotation of computer management. The team tried: AD+ Evolution_of Relations:

[2] PLUG - play ubiquitous games- ANR/RIAM -2007-2009.

[3] Transhumance (Ad Hoc Services: Fulfilling a Platform for Peer to Peer Applications on Mobile Ad Hoc).

[4] Safari - Mobile Ad Hoc Networks – ANR/RNRT- 2003-2005.

[5] Architectures Distribuées Pour Applications Multiples, `http://adam.hypotheses.org/`

[6] POPEYE (Professional Peer Environment Beyond Edge Computing) – STREP/IST – 2006-2008.

ADER. But the group promptly remarked that it sounded like the inventor Clément Ader who was a major inventor in the aviation field and also for the telephone. AD + Ingénierie des Relations_(engineering of relations): ADIR. But none of these names were “fun” and the group finally came up with ADAM that would stand for “Architectures Distribuées Pour Applications Multiples” (distributed architecture for multiple applications), that did not mean much more than the other options but that combined applications with infrastructure and a biblical name, knowing that at the time there was only one man in the team.

3.4 Scientific Logos: A Question of Identity

The logo is a sign of branding permeating the logic of research as a commodification process [27] that “mediate the junctions between technological producers and consumers”. Our view is that this trend of naming is not only functional, memorable and strategically positioning the project but also poetically adding to the identity of the project.

We would like to go back to one particularly successful example that was, in effect, realized by a professional designer, while most of the time new project teams and research groups try to find the resources within their team, students, or friends who agree to do it for free…

The project is called ILHAIRE: Introducing Laughter in Human Avatar Interaction, Research and Experiment [26]. The logo was realized by one of the research partners, the French animation company: La Cantoche. Though it is not a communication company and it does not do logos for a living, they have graphic designers and artists who could work on the project. The whole process is interesting in as much as it is quite different from an ordinary process of branding. According to the project manager, one of the main differences is that the team includes “a large variety of disciplines and interests in the project and there is no formal hierarchy between the members [interview with J.]” so making a decision was perhaps more complicated in that respect. Another difference is that the logo is defined in relation to the theme of the project (and not as could be the case in other circumstances to position the “product” within a market). What makes it really precious then is that, according to J., “it focuses on the identity of the project, the “heart” of the project”.

The exchanges between the different participants showed that this identity was a subject of controversies: social scientists contended that laughter was social, for others it was important to see that laughter was necessary to health and well being. Engineers focused on the models of gestures and acoustic dimensions. The team was also worried that talking about laughter would not be taken seriously. They wanted to make it clear that it was research, which was finally why the baseline was chosen: “the science of laughter”. The production of the logo was therefore one of the productive and possibly rare occasions where the different stakeholders explained and compared their viewpoints.

What makes it a real poetical work is that the participants had to learn about the process itself, the impact of images, the role of signs and words. The role of the project manager was particularly important in sharing this vision about the making of a logo. For instance, they had to understand that a logo is "not an illustration". In particular, it was tempting to use an avatar as part of the logo but then it would work as an example, rather than a sign with a more general purpose. They also had to work on how to share values through the choice of a logo: "warm colors", "playful", or showing: "the collaboration between different fields of research (engineering + human sciences) [28]".

The process went from a culture of words with denotative value to a process of images with an emphasis on culture- from trying to deploy the acronym, to gradually incorporating the culture of cartoon, with references to the field of computer science.

The final logo shows a mastery of graphical techniques: see for example, the bubble and emoticon that are left open by the frame of the rectangles.

But more importantly, it played a role in the representation of the project for the participants. It crystalized a partnership, an identity, and the representation of a technology.

4 Findings

Despite the differences between giving names to technologies, to projects or designing a logo, the examples in this paper emphasize the relationship between a technology and the different ways to invent it.

Names are caught between a denotative and connotative tension. In the naming process, the example of the TrackPoint naming story as told above is an epic journey of descriptive functional names (pointing stick) and descriptive joking names with sexual innuendos. Despite the differences, the top three options ("pointing stick", "joy nub", "TrackPoint") struggled with an accurate description of the device. The name either points to the underlying technology or to its surface properties, how it works or how it is perceived from a user perspective. The difficulty to choose between denotative options: pointing stick and TrackPoint, shows that to describe something might not be so easy in the first place. As demonstrated by Bucciareli [29], designing is about agreeing on a definition; different stakeholders have a different definition of what the technology is about. What this example as well as the Bluetooth example show is that the denotative choice is not solving the problem and that names are produced also with connotations in mind. This is obvious in both "Artificial Intelligence" and "Bluetooth" as well as the discarded "joy nub". Connotations deepen the cultural relationship of the technology with other artificial products because they evoke linguistic and social codes, rhetorical and ideological conventions and therefore enlarge the cultural references of the innovation. Play on words are used to explore and exploit connotations and support associations of ideas in the case of naming technologies or projects. As related by its inventor, "joy nub" was probably the more striking and memorable name, but was abandoned for reasons of propriety. Artificial Intelligence was picked by its author with a background of at least two centuries of literature on artificial creatures. What was at stake was to stimulate the imagination and the research thanks to the evocative metaphor of intelligence and the open ended nature of the word artificial that no longer related the expression to an actual artifact but acted as far more diffused.

Names are also compared to other names and finding the right name might have to do with the necessity to break from or on the contrary to establish a connection with other names. Trackpoint (at the beginning called Trackpoint II) was chosen to establish such a connection. It was thought of in relation to other existing products or devices. The naming process therefore focused on founding the new device within a family, and gave it a sense of continuity, an inscription in the history of tracking devices. Part of the imaginative process was about the lineage of technical artifacts. On the other hand, Bluetooth does not relate easily to any other existing technology and is strikingly original as it decidedly departs from any denotative and descriptive naming tradition.

Giving a name to a project seems to be slightly different from giving a name to a technology. While the inventors of a technology aim at creating a common name, the groups who baptize their project inscribe their activity in history and identity through a proper name. Names given to project can be chosen to be descriptive, to remind people of the roots of or aspirations for the project, or simply be a memorable word or sound that will attach to people minds to help the project. According to Kripke [30] this proper name stands for the dubbing ceremony that accompanies its creation. As proper names, they have some relationships with the culture of their authors: Bluetooth is chosen by people mostly coming from Nordic countries, Adam is a private joke about the only man of the research team. The connotations are then exploited to signify not so much the technology as the collaborative work of the team. But the word then does not so much refer to anything (who really knows who Bluetooth was and what he did?) than build on the originality of the choice of words that appear truly poetic if not totally random: a tooth: something tiny but piercing, associated with the unlikely colour blue that might refer to a range of other technologies of the time (Blue Ray for instance). While proper names are the subject of debates between descriptive or causal historical theories [31], an important point is that they are also context-sensitive [32]. This means that they can take on meanings very differently according to the context of utterance. In fact, a proper name invites different viewpoints on the same thing because it is not purely descriptive: for each partner of the ADAM project, the identity of the project remains open to interpretation. In fact, the combination of Distributed Architecture on the one hand, and Multiple Applications on the other while originally chosen for evocation purposes, also forced the members of the team to think about the relationship between infrastructures, applications and services. It expanded the conceptualization of these systems. As a proper name, they benefit from a certain "looseness", as defined by Searle [33]. They do not describe and give attributes to an object (or person) but still refer to it in ways that open to a whole family of possible descriptions. A literary analysis of these names shows that they try to mix technical features with a symbolic representation of the technology.

Designing logos apparently shares some of these issues of group identities and technological descriptions as the example of Ilhaire tends to show. The image is chosen to represent the technology without using the usual tools of the trade, like scientific schemas. Rather the example shows that several cultural relationships are designed within the iconic choices. The logo is clearly related to the culture of cartoons. It also alludes to emoticons and finally to scientific images through the representation of a sound curve. It functions as a metonymy: the project is part of a wider scientific, cultural and literary whole. Similarly to the expression "Artificial Intelligence", it therefore opens up to larger realm of possible research and applications.

5 Conclusion

In this article, examples bring forward the value of imagination and strive to promote the study of poetic practices in research project. These examples show that we can use methods that combine sociological methods and methods from humanities to produce thorough analyses and comparisons and come up with typologies of poetic practices integral to all IT project. The primary function of these writing practices is not only descriptive. Names of technologies, acronyms for projects, logos and narratives, are motivating, aspirational and evocative practices that support the expansion of the scope of the technology.

Poetics can be understood in production and in reception. In production, the poetic work is an aesthetic and intellectual work that looks for meaning. These productions, rather than working on precise attributes, operate what the logician and philosopher of language Saul Kripke, calls the "opening of possible worlds". Even in the case of technology naming that tends to limit the possible connotations, we have seen that the exploration of names becomes an expansive part of a project with results hovering between denotation of what the project is about and connotations of how people want to feel about it. In reception, the proper name of a project is also poetic, because like other proper names, it supports a context-sensitive diversity of interpretations. The Artificial Intelligence example above shows how the diversity of interpretations might be designed into a name to help the project to continually expand its goals.

Poetics as part of the aesthetic experience depends on the process of open interpretation. A word or an image that seems to close the object on its properties does not support the same interpretive leeway. We can use the difference that Umberto Eco [18] analyzes between a traffic sign and a work of art. The traffic sign is no longer a traffic sign if it is freely interpreted and discussed. But a work of art constitutes "an open product on account of its susceptibility to countless different interpretations which do not impinge on its unadulterable specificity. Every reception of a work of art is both an interpretation and a performance of it, because in every reception the work takes on a fresh perspective for itself [18]. Therefore, what is at stake in these poetical practices of engineering is to explore the cultural meanings of the technology, then to open it to a diversity of interpretations.

This paper focuses on the semiotics of condensed forms of identity through name and logo. But poetic practices include other literary and visual forms that can affect research projects. In particular, we can concentrate on narratives such as use-cases or personas and the way they contribute to build and expand the scope of research projects. We could also compare these findings to research projects over time. Placing this research within a historical perspective would help us to consider different poetic practices, different genres and styles over time. For instance they would help to figure if the growing multidisciplinarity in research projects is a key factor in the production of figures of speech that gather the different disciplines. In further research, we can also explore how words and images that accompany the emergence of a technology not only help researchers anchor their technology in culture, but also contribute in special ways to the involvement of potential users. In particular, when users are asked to test technologies, how do they interpret them based on a diversity of factors including their names and other representations. We should further enquire into how these representations help structure a design for debate around its value and use [34]. This research program can help to establish a taxonomy of poetic practices according to fields of research, actors, types of research products. Beyond the taxonomy, a theory of a poetic of research can help the process of invention by articulating findings to an artifact situated in and incorporating social and cultural context.

References

1. Chalmers, A.F.: What Is This Thing Called Science?, 3rd edn. Open University Press (1999)
2. Latour, B.: Science in Action: How to Follow Scientists and Engineers Through Society. Harvard University Press (1988)
3. Hackett, E., Amsterdamska, O., Lynch, M. (eds.): The Handbook of Science and Technology Studies, 3rd edn. MIT Press (2007)
4. Simondon, G.: L'invention dans les techniques. In: Cours et Conferences. Seuil (2005)
5. Sachs, J. (trans).: Aristotle: Poetics. Focus Publishing/ R. Pullins Co. (2005)
6. Preminger, A., Brogan, T.V.F., Warnke, F.J. (eds.): The New Princeton Encyclopedia of Poetry and Poetics, 3rd edn. Princeton University Press (1993)
7. Jakobson, R.: Linguistics and Poetics. In: Sebeok, T. (ed.) Style in Language, pp. 350–377. M.I.T. Press (1960)

8. Barthes, R.: Elements of Semiology. Lavers, A., Smith, C. (trans.). Jonathan Cape, London (1967)
9. Valéry, P., Oeuvres, G.H.J. (eds.): Collection Pléiade, Variétés, Paris, vol. 1 (1957)
10. Meschonnic, H.: Pour la poétique, Gallimard - Nouvelle Revue Francaise, vol. 1 (1970)
11. Genette, G.: Introduction à l'architexte. Seuil, Paris (1979)
12. Gentès, A.: The Design of Emotions: How the Digital is Making us More Emotional. In: Pelachaud, C. (ed.) Emotion-Oriented Systems, Hermès, London (2011)
13. Olohan, M.: Scientific and Technical Translation. In: Millan-Varela, C., Bartrina, F. (eds.) Routledge Handbook of Translation Studies, Routledge, London, New York (2012)
14. Clifford, J., Marcus, G.E. (eds.): Writing Culture: The Poetics and Politics of Ethnography. University of California Press (1986)
15. Sicard, M.: Chercheurs ou Artistes? Entre Art et Science, Ils Rêvent Le Monde, Autrement (1995)
16. Burri, R.V., Dumit, J.: Social Studies of Scientific Imaging and Visualization. In: Hackett, E., Amsterdamska, O., Lynch, M. (eds.) The Handbook of Science and Technology Studies, 3rd edn., pp. 297–317. MIT Press (2007)
17. Lynch, M.E., Woolgar, S. (eds.): Representation in Scientific Practice, 1st edn. MIT Press (1990)
18. Cambrosio, A., Jacobi, D., Keating, P.: Ehrlich's 'Beautiful Pictures' and the Controversial Beginnings of Immunological Imagery. Isis 84(4), 662–699 (1993)
19. Eco, U.: The Open Work. Harvard University Press (1989)
20. Keane, W.: Semiotics and the Social Analysis of Material Things. Language & Communication 23, 409–425 (2003)
21. Krippendorff, K.: The Semantic Turn, A New Foundation for Design. Taylor & Francis, Boca Raton (2006)
22. Barr, A., Feigenbaum, E.: The Handbook of Artificial Intelligence, vol. 1. William Kaufmann, Inc. (1981)
23. http://www.eetimes.com/electronics-blogs/other/4394006/Bluetooth-how-it-almost-didn-t-happen-
24. Rutledge, J.D., Selker, T.: Force-to-Motion Function for Pointing. In: Proceedings of INTERACT: IFIP International Conference on Human-Computer Interaction, Amsterdam, Netherlands, pp. 701–706 (1990)
25. Trackpoint (Model L40 SX) Manual, Publication no. 1397090 (1991)
26. McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html
27. Fishman, J.R.: Manufacturing Desire: The Commodification of Female Sexual Dysfunction. Social Studies of Science 34(2), 187–218 (2004)
28. http://www.ilhaire.eu/
29. Bucciarelli, L.L.: Designing Engineers. MIT Press (1996)
30. Kripke, S.A.: Naming and Necessity, New edn. Wiley-Blackwell (1991)
31. Brock, S.: The Ubiquitous Problem of Empty Name. The Journal of Philosophy 101(6), 277–298 (2004)
32. Burge, T.: Reference and Proper Name. The Journal of Philosophy 70(14), 425–439 (1973)
33. Searle, J.R.: Proper Names. Mind (67), 166–173 (1958)
34. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural Probes. Interactions 6(1), 21–29 (1999)

Storytelling in Visual Analytics Tools for Business Intelligence

Micheline Elias[1,2], Marie-Aude Aufaure[1,3], and Anastasia Bezerianos[4]

[1] MAS Laboratory, Ecole Centrale Paris, France
[2.] SAP Research, Business Intelligence Practice, France
[3] AxIS Research Group, INRIA Paris, France
[4] Univ Paris-Sud & CNRS, INRIA, France
michelleliasgmail.com,
Marie-Aude.Aufaure@ecp.fr, anab@lri.fr

Abstract. Stories help us communicate knowledge, share and interpret experiences. In this paper we discuss the use of storytelling in Business Intelligence (BI) analysis. We derive the actual practices in creating and sharing BI stories from in-depth interviews with expert BI analysts (both story "creators" and "readers"). These interviews revealed the need to extend current BI visual analysis applications to enable storytelling, as well as new requirements related to BI visual storytelling. Based on these requirements we designed and implemented a storytelling prototype tool that is integrated in an analysis tool used by our experts, and allows easy transition from analysis to story creation and sharing. We report experts' recommendations and reactions to the use of the prototype to create stories, as well as novices' reactions to reading these stories.

Keywords: Visual Storytelling, Business Intelligence.

1 Introduction

Stories are one of humanity's communication structures and storytelling a means of passing on wisdom and culture. Individually and collectively, stories help us make sense of our past and reason about the future. Johnson 17] and MacIntyre [21] argue that story narrative also goes beyond communication, it is also a process of extracting meaning from events, that is central to human experience and conduct.

It is thus not surprising that intelligence analysts, who make sense of data, identify links between disparate pieces of intelligence, and communicate their findings to decision makers, are interested in storytelling. Their analysis process is supported by the construction of stories and narratives, both during sense-making and during presentation of results. Bier et al. [3] point out that a story is a powerful abstraction to conceptualize threats and understand patterns as part of the analytical process, and story structures and storytelling is the means to present the analysis results. As analysts continue to work with increasingly large data sets, data visualization has become an incredibly important asset both during sensemaking analysis, and when communicating findings to other analysts, decision makers or to a broader public [15, 27].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 280–297, 2013.

Given the importance of storytelling in different steps of the analysis process it is clear there is a need to enhance visual analysis tools with storytelling support. Nevertheless this process is not simple [33, 20], as analysts need to work within very large data resources and highlight and explain items or events of importance and their connections to their audiences. Despite the growing use and research work on storytelling and narrative visualization in the visualization domain [26, 9, 15], related research on the domain of BI has not equally progressed. Our work attempts to rectify this.

In Business Intelligence (BI) analysis, the most popular visualization tools are dashboards. Dashboards [8] are collections of visual components (such as charts or tables) on a single view [12], that permit analysts to explore their data and quickly view different aspects of complex datasets. Nevertheless, simple collections of visual representations cannot be interpreted by untrained audiences; to become meaningful they require interpretation and explanation, often presented in a story narrative.

We attempt to answer the following research questions: What are the actual practices of BI experts in creating and communicating visual stories to their audiences, and do current BI visualization tools support well this story creation and storytelling process? How can we enhance BI visual analysis tools with narrative capabilities, and are these capabilities effective in communicating analysis stories to others?

Our work makes the following contributions:

(1) Interviews with expert BI analysts (story ”creators” and ”readers”), provide a better understanding of current practices in creating BI stories. BI stories are an asynchronous, visual, and interactive means of transmitting and sharing information on data between analysts and their audience. They include visualized data (dashboards, charts or tables) and detailed explanations on the story structure in the form of presentation(s), detailed textual annotations and external resources such as wikis. (2) Current tools fail to support the storytelling process, which remains cumbersome and requires frequent switching between software (analysis tools, screen captures, text editors, presentation tools, etc.). We emphasize the need for storytelling support, and extract requirements for enhancing BI tools with visual storytelling capabilities. (3) Following these requirements and a user-centered design approach, we implement a storytelling prototype incorporated in an existing visual analysis dashboard, to fluidly transition from analysis to story creation and reading. (4) We report on feedback from BI experts on the usefulness of the prototype as a communication and teaching medium, and from BI novices reading a story on its effectiveness for story communication.

2 Related Work

Stories are series of ordered events and facts, and their connections (Oxford English Dictionary). In intelligence analysis, it is furthermore an abstraction used by analysts to conceptualize threats and understand patterns in the analytical process [3], and communicate their findings to others [2].

Stories in Business. In recent years, organizations and their leaders have identified the importance and value of narrative and anecdotal information conveyed in the form of stories [29], to present either complex ideas, the context and details of crucial information, or personal interpretation of that information [28].

Research conducted to date has demonstrated the value of storytelling to improve organizational structure [24] and collaboration quality [23, 6], socialization and adaptation of new employees [7, 18, 19], organizational and financial success [4, 5], innovation and new product development [22], and teaching and learning [16].

The majority of this work is a meta-analysis of the effect of storytelling within an organization, rather than identifying the storytellers' needs in terms of supporting and enhancing the storytelling process as is our case. Moreover, the stories themselves discussed in this work, relate to the transmission of information and knowledge within an organization, mostly in textual or verbal form, rather than in visual form. The widespread use of *visualization* dashboards in the domain is a more recent development [8], and so is the transmission of knowledge *between organizations* (dedicated BI analysis organizations and their clients). Thus storytelling needs in the domain have evolved. In Section 3 we explain current practices in visual knowledge transmission and we take a step at characterizing current problems and needs more precisely.

Stories in Sense Making. Baber et al.[2] point out that contemporary theories of sense making rely on the idea of 'schema', of a structure to organize and represent factual information, as well as the knowledge, beliefs and expectations of the people who are engaged in sense making. They can thus be considered as a collection of narratives. They further discuss the formalism of stories in sense making, and how the most effective stories are organized around the actors in the stories, their actions and rationale, events, their context, and most importantly the relationships between these. As argued by Bier et al.[3], for effective collaboration and communication we need to use less text, and organize knowledge around entities (people, places, things, times etc.) rather than free form text. Similarly, Pirolli & Russell [25] propose the mapping of intelligence facts and insights into frames, that can be expressed in a variety of forms including stories, maps, organizational diagrams or scripts. It is thus clear that conducting intelligence analysis, communicating findings, and organizing knowledge in stories, has a strong visual component that represents entities and their connections.

Stories in data visualization. Often text or audio transmit the main story, while visualizations support the story or provide details. Comics and flowcharts are special narratives relaying mostly on visuals rather than text. Recently, we have seen an increase in integrating complex visualizations into narratives in many news organizations [11], journalism reports (e.g. New York Times, Washington Post, the Guardian), and television reports (e.g. Gapminder[1]). Segel et al.[26] explore different aspects of narratives from a variety of sources and identify distinct genres of narrative visualization.

In the business intelligence domain the main visualization tools are dashboards [8], collections of multiple visual components (e.g. charts, tables) on a single view [12]. BI dashboards permit users to interpret data at a glance, and are very popular (e.g. [Dundas[2], Oracle bi 10g[3], Xcelsius, Spotfire[4], Tableau[5]]). But as Wojtkowski and

[1] `http://www.gapminder.org, 2010`
[2] `http://www.dundas.com/microsite/dashboard`
[3] `http://www.oracle.com/technetwork/middleware/bi-enterprise-edition/`
[4] `http://spotfire.tibco.com/`
[5] `http://www.tableausoftware.com/`

Wojtkowski [33] point out, dashboards and other visualization tools used to analyze complex data cannot simply tell stories. They need to be "tailored" to accommodate storytelling by better highlighting items of importance within very large data resources [14], in a way that is efficient for the storyteller and clear for the audience.

Some visualization systems began to integrate storytelling [20]. GeoTime [9], a geo-temporal event visualization system, integrates a story system that shows events in space and time, hypertext linked visualizations, and visual annotations to create an environment for both analytic exploration, and story creation and communication. The design of the tool was based on real user needs and has been evaluated with intelligence analysts. Wohlfart and Hauser [32] present a system for demonstrating to audiences the path followed in analyzing 3D volumetric data, while providing audience with limited interaction with the 3D data. Both these systems deal with a single visualization/chart seen over time or from different views. Thus their designs are not necessarily applicable to the multi-dimensional and multi-chart BI visualizations where the connections and links between data are not as clear.

Storytelling tools in BI are not yet as advanced. Systems like Sense.us [15] and Tableau[5] allow analysts to visualize data, conduct analysis, and store a history of the exploration, that can serve as a step towards creating a story. Many Eyes [31], Tableau Public[5] and Sense.us [15] publish interactive visualizations online, and permit collaborative analysis through comments on a single visualization, creating an evolving analysis. This collaborative annotation can be seen as a step towards a collaborative knowledge narrative, where an analysis story could be extracted from the visualizations and comments. Nevertheless, these tools do not provide explicit means to indicate story progression and highlight relationships between multiple visualizations (seen in dashboards), that are key to intelligence analysis and communication [2].

3 Interviews

To support visual storytelling in BI, we first investigate current practices in transmitting BI analysis results, and identify challenges experts face currently when creating and sharing their stories, or reading and interpreting stories of others. We interviewed 5 BI experts in a leading business intelligence development company, with experience from 6 months to 12 years. Participants' job descriptions included development project manager, project manager, development director, data warehouse engineer and delivery manager. Three experts used dashboards and communicated their analysis or read analysis from others daily, while two several times a week. Interviews were held in person or by phone and lasted over 1 hour (Table 1). We report next our major findings. Note that they hold true for all BI reporting tools available today, as indicated by our experts' experience with multiple tools, and our own investigation.

Table 1. Experts', experience, dashboard use and interview duration

Experts	P1	P2	P3	P4	P5
Experience	7 years	6 months	6 years	3 years	3 years
Dashboard use	daily	weekly	weekly	daily	daily
Duration	90 min	120 min	60 min	60 min	60 min

3.1 Current Practices: BI Reports

All experts communicate their analysis or read analysis from others in the form of BI reports. These contain an entire dashboard, often accompanied by several single charts and tables. Their purpose is to help monitor aspects of business performance, by highlighting with charts key performance indicators (KPIs) that indicate success of project management and progression of development teams. Our experts were report creators, but occasionally acted as readers of others' reports.

A BI report can be a single page, with only a title and a dashboard (Fig. 1.). This one-page report summarizes the most important data, and can serve as a starting point for longer reports, up to 50 pages in length, that give more details. Details can be additional visualizations, tables, annotations, links to the data used in the visualizations, and finally block text, although all participants preferred limiting text to 5% of the space in a report. When possible, creators want to make their reports "live", with interactive visualizations for the audience to explore. Thus BI reports have a very strong visual component, with little text added for explanations. The one page report is preferred by clients like company managers, that want an "at-a-glance view" of a project status, and that see many such reports during a day. Longer detailed reports are used to communicate findings to other analysts, project teams and product owners.

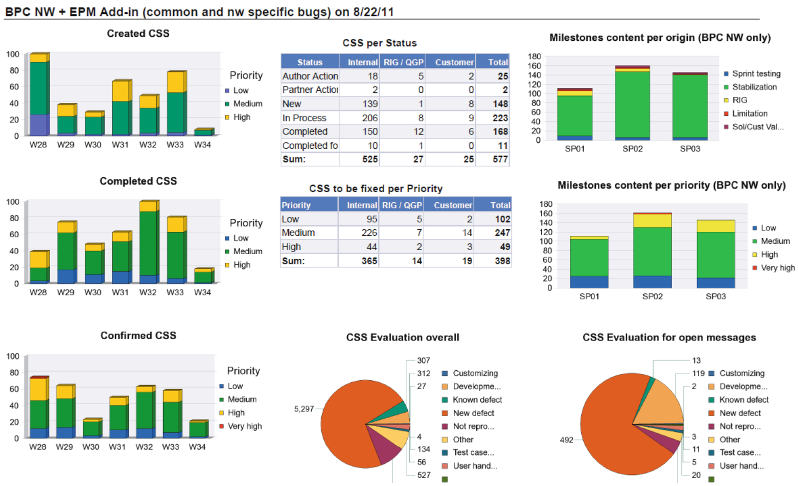


Fig. 1. One page BI report from an expert: dashboard presenting bug status in a project. Note the lack of annotations/explanations, aside from chart titles and data dimensions.

Experts mentioned that BI reports are used to (i) answer specific questions (e.g. What is the code development progress in project X the last 3 weeks?); (ii) investigate specific data (e.g. Investigate the increase of bugs in section Y of project X); (iii) manage conflicts and highlight problems (e.g. Team X completed less use cases than planned, because test team Y did not complete testing); and (iv) interpret past data and predict future trends (e.g. Given last year's sales, what is this year's projection?).

BI reports are nevertheless only one of the tools used to communicate analysis findings, and are not a complete BI story.

3.2 Current Practices: Supporting Material for BI Storytelling

Experts explained that reports are difficult to understand without detailed explanations from the creator. Thus they don't represent a complete BI story on their own. Before publishing a new report, the creator provides an introduction session to report readers. During this (usually one hour long) session, she explains with a presentation the entire story and goal of the report, the meaning of each chart, the relations between different KPIs in different charts, as well as in what sequence to read it. Thus their verbal explanations and presentation slides are ways for creators to explain their analysis path. These presentations often show relational diagrams, text explanations and interpretations, highlight specific visualizations or parts of visualizations using colored highlights, arrows and other symbols. The report creator often draws by hand on a chart the ideal data values, to help compare the current situation to target goals.

The audience can ask questions during or after the training sessions, through emails or arranged meetings. Content similar to the presentation, is often put on a wiki page to answer follow-up questions. The session recording, wiki page, and presentation slides, are made available to the audience as reminders and reference material. Experts explained that this material is not a complete story either, as it does not include the visualization and data details shown on the report.

Thus a *complete BI story* is a collection of visual representations of the most important data followed by further data details (BI report), accompanied by instructions on how to read the visualizations (order, connections, importance) in the form of presentations and verbal or textual instructions on a wiki. Although creators present a desired way to view data, this structure is not enforced: the audience can pursue the story in a different sequence and dig for data details in the report. Thus a BI story differs from a simple fixed sequence presentation that prohibits exploration. Its goal is to communicate analysis findings and supporting evidence.

3.3 Current Practices: Teaching BI Storytelling

According to our experts, all the material making up a BI story is also a tool for training analysts. As a recently trained creator explained (P2), this material taught her the key aspects of BI reports and how to interpret them. When analysts start out they read in detail older reports and their supporting material to understand how to analyze visualizations, see relations, and identify important points and their link to KPIs.

Experts mentioned that designing a new report is hard and requires a lot of experience, thus they often use a template that they modify according to their needs. Three (P2,P3,P4) stressed the influence of a senior creator (P1), and their reliance on her BI report templates to create their own. They still occasionally contact the senior creator when facing difficulties or are unsure of the clarity of their message in a report.

The supporting material is also often reused. All experts explained that data (and thus reports) change, but often the structure of BI stories repeats itself. Moreover, they sometimes create a sequence of reports on dynamically evolving data (e.g. sales over several years). These stories change very little. Thus material on how to read reports can be re-used by adapting it to new stories, nevertheless, experts pointed out it can be a tedious and repetitive process.

3.4 Current Practices: Collaboration in BI Stories

Experts explained that report creation follows several iterations of communication with a client (usually a decision maker like a manager or executive), who is one of many possible report readers. Thus the focus and structure of the story changes and evolves through the communication between the story teller and the story reader over time, and stories become the collaboration product of story tellers and readers. This process does not appear in most other visual analysis domains and visualization storytelling, where the reader's needs do not directly factor into the creation of the story.

This interactive report evolution process can be heavyweight (through phone calls, emails, etc.) and take several iterations to iron out, while the communication details are often lost between different versions of the report. At each iteration, creators explore the desired data using interactive visualization dashboard tools, create initial visualizations, provide additional information and details about the analytical techniques used, and finalize reading paths (i.e. how to read a report).

Once the story is finalized, it serves to clarify connections in the data and answer specific questions analyzed by BI analysts (see section on BI Reports). These answers often result in more detailed or tangent questions from the audiences, and the need for analysts to conduct side data investigations and generate new stories. As in other visualization domains, BI stories are thus a visual communication medium between storytellers and their audiences, but contrary to other domains, there is an open communication loop, where audiences can continue to ask for new stories.

3.5 BI Storytelling Challenges

All experts use an in-house reporting tool, that gives access to different data sources and can extract interactive dashboards, charts and tables from their dashboard analysis system. It also provides the capability of adding text. Three (P1,P3,P5) also used Excel for creating less complex reports. Our experts identified several issues with this process, which are similar to other BI analysis and reporting tools.

When extracting interactive visualizations from their analysis tools, metadata and annotations from their analysis are not extracted and have to be recreated. Moreover, the annotation capabilities of the report creating software are very limited. They cannot annotate specific data points, while sequences and connections cannot be displayed graphically but have to be explained in one of the supporting material. Finally the report creation tool can be in-house software. Thus to share reports, creators often extract static snapshots of the tool output and save it in PDF format for their readers.

Some report creation tools give access to interactive visualizations and dashboards through hyperlinks. Our experts like this option and add these links to their non-interactive reports (PDF, Excel) and presentations. Readers within the organization may have access to the in-house reporting tool, and can then interact with individual visualizations or dashboards (e.g. query and filter data, or drill down/up). This interactivity is lost when reports are communicated outside the company.

To overcome the shortcomings of reporting tools, creators are forced to provide the supplementary material (presentations, wiki pages) and when possible links to interactive visualizations for in-company clients. They described this process as limiting and requiring a lot of work duplication.

We requested clarification on two points regarding the possible integration of analysis and storytelling: (i) Given that analysis dashboards are exploratory environments that allow users to interact fully with data, should such interactions be allowed in the storytelling? (ii) If creators had the opportunity to create their entire story with all support material in one place, how would they prefer the story to be visualized?

Interactivity. Due to the constraints of current reporting tools, shared reports are often non-interactive when accessed outside the organization. When creators were asked if they want the visualizations in the reports to be completely interactive and encourage readers to interact with them (e.g. using drill down/up, filter, link & brush), four (P1,P2,P3,P5) of our experts prefer to have interactive visualizations that permit linking and brushing (i.e. data selection). But they would limit the more advanced interactions such as drill down/up or filtering. They felt that all the data needed to tell the story should be displayed clearly in the report without the need to explore the data further. The fifth (P4) would not be opposed to fully interactive visualizations. Thus authors feel business stories should be mostly author-driven and constraint, known to work best when the goal is storytelling or efficient communication [26].

Story Templates. Going from the current practices of storytelling (BI report, presentations, wiki) to a dedicated storytelling tool is not straightforward. We thus showed our experts a group of story templates identified by Segel et al. [26] to see if they met their needs. All chose the "Annotated chart" as the preferred template (Fig. 2), with the modification that it should have multiple charts on the same page (dashboard) that they can annotate. Four experts (P1,P2,P3,P5) identified "Partitioned poster" as a potential template, where the side list of charts display details that support the main chart in the central region. Three (P1,P2,P5) mentioned that the "Slide Show" template is useful both as a means to focus on attention on each chart and zoom to details, and as a step-by-step presentation that clarifies the analysis path and the ideal reading sequence. One expert (P5) found the "Flow Chart" useful for showing some business scenarios, like following bugs during development (discovered, tested, fixed etc.). Another (P3) found the "Comic strip" useful template, but with added annotations.

So besides the templates identified by Segel et al.[26], for BI stories we need a new template that consists of an annotated dashboard. Our users attempt to do this with their current reporting tools, but they are limited (cannot annotate detailed points and relations), or they create it manually in their presentations.

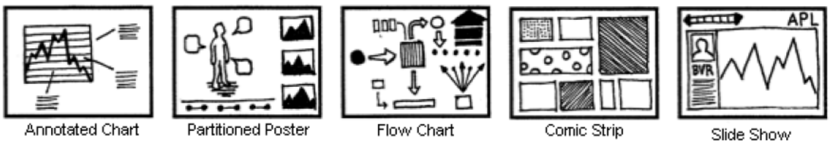


Fig. 2. Genres of visual narratives templates from Segel et al. selected by our BI experts

4 Participatory Design Session

We brought back the senior BI expert that trained other analysts (P1), to identify the best design for a BI storytelling tool. She provided us beforehand with one of her old reports and supplementary material. During this session we gave the expert a copy of her material, and several cut-outs of the entities in the report (charts, titles or tables). She was also given ”narrative aids”, such as arrows, lines, numbered items, grouping containers of different colors, and annotation bubbles. We asked her to construct a stand-alone story of the report on an A3 paper, with accompanying audio if necessary, in a way that it can be interpreted by a general audience with no training. The session lasted 2 hours and was recorded.

Fig. 3 shows the final version of the manually created story by the expert, where we can see the intense usage of annotations and explanations, connected by arrows and lines. Other structures used by the participant were chart grouping rectangles (enclosures for placing charts that need to be read together), sequence markers (what needs to be read first), and highlight markers (to draw attention to part of the story).

We asked our expert how she would prefer the readers to see the story, as a static image or an animated presentation. She explained that both are needed, the static representation shows the ”entire story and gives context”, whereas the animation ”focuses the audience where I want”. She then played out for us how she wanted the animation to be presented, explaining when and how to zoom to specific areas of the story. This play-back was recorded using static shots.

Fig. 3. Story created by an expert during the design session

5 Requirements for Creating BI Stories

Based on expert self-reporting, BI reports, the main communication medium of BI analysis, consist of snapshots of visualizations and textual descriptions, made in a different environment than the one used for analysis. These reports are fact-based and can be interpreted only by an expert audience. To increase their audience, report creators use supplementary material, like wiki pages, presentations, and when possible links to interactive visualizations. Based on our experts' interviews and a participatory design session, we identified a set of requirements for enhancing BI analysis tools with storytelling capabilities:

R1. Fluid Transition. Analysis tools used to explore data and create visualizations are different from report creation tools. Exporting visualizations from the first to the second to create a BI story costs time and effort, and limits the possibility of embedding meta data or annotations created during the analysis. To ensure that creators do not recreate information, they require a fluid and integrated way to transition from their analysis and meta data associated with it, to report creation.

R2. Integration. To tell their stories, BI creators need tools that combine all materials used currently in their story creation: BI reports, interactive visualizations, ways to indicate story structure, highlighting capabilities, presentation of the story in sequence, and textual or audio explanations.

R3. Narrative visual aids. Report creators need to add focus expressions to draw attention to specific visualization data, such as highlighting, coloring, annotating and zooming. They also require ways to indicate reading sequence (e.g. vectorial references, like arrows [13][30]). These are not available in reporting or BI analysis tools.

R4. Interactive visualizations. Visualizations on shared reports are often non-interactive when read outside the organization. A storytelling tool should have completely interactive visualizations, although the way that readers interact with the data should be limited (by default to brushing and linking) and be controlled by the creator. This balance has been identified as a challenging aspect of storytelling [20].

R5. Appropriate BI Story templates. BI stories have specific structure not necessarily shared by other story narratives identified by Segel et al.[26]. Our experts identified templates of interest (Fig. 2) and highlighted the need for a new template that consists of an annotated dashboard.

R6. Reuse. Although BI reports and data changes from analysis to analysis, often the underlying structure of BI stories remains the same. It is thus important to be able to easily reuse the structure of stories created within the tool both for stories of evolving data and similar future stories.

R7. Optional playback. Readers should be provided with a static representation of the entire story to get an overview and explore it on their own pace, as well as a guided playback to help them follow analysis paths that not easy to discover and understand the important points according to the creator.

Such storytelling support would facilitate the creation and learning to read BI stories. While the scientific visualization storytelling tool 32] supports R3,R4,R7, and GeoTime [9] supports R1 and R3, we are not aware of any BI dashboard or other multi-chart visualization system that provides *all* the above functionalities.

6 BI Narrative Prototype

Following these requirements, we extended an existing dashboard exploration tool to support the creation of BI stories.

6.1 Exploration/Annotation Dashboard (Fig. 4)

The user starts from a traditional analysis dashboard, a collection of coordinated (synchronized) charts connecting one or more data sets. It provides advanced exploration capabilities such as data selection and filtering, and OLAP functionality (e.g. drill-up/down) [1]. We build our narrative tool on top of an existing visualization dashboard system [10] that supports annotations on "data targets", such as charts or tables (e.g. a bar-chart) or parts of them (e.g. a specific bar in the chart). Annotated data targets are highlighted in the dashboard and an icon indicates the number of attached annotations. A list of all annotations is also available on the right.

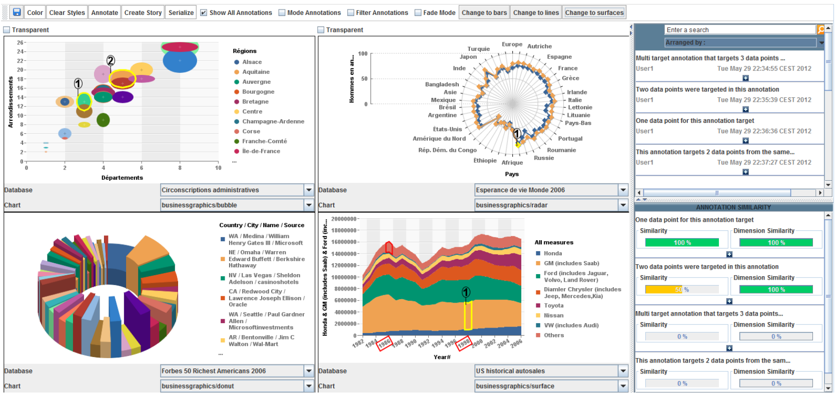


Fig. 4. Dashboard with visualizations created during analysis on the left, and list of annotations added to different data contexts while exploring analysis visualizations on the right.

After or while conducting her analysis on the dashboard, the analyst can create a story. The menu option "Create story", opens up the narrative board window. All visualizations and their annotations (annotation text + data targets highlighted) are placed in the narrative board. Thus the analyst can transition fluidly (**R1**) from analysis to story creation. Because dashboards can present evolving data, but a story can be an instance in time, by default each visualization is placed on the narrative board with a time-stamp to indicate when the visualization was taken from the data.

6.2 Narrative board (Fig. 5)

Here users create a BI story. Visualizations and annotations from the dashboard appear on the main window, arranged by default as in the dashboard (to support the dashboard template identified in our interviews **R5**). Users can resize and relocate all story entities freely, or choose to organize them using other appropriate narrative templates (**R5**). They can edit visualization titles, annotations of entities, and add new text entities. Users can add visual entities that indicate relations, grouping and emphasis (**R3**) to help create stories, available at a tool palette on the right. The story can be seen in: a static representation, where all entities remain one screen; and a playback representation, where entities are highlighted sequentially according to an author-defined sequence (**R7**). We categorize the entities available on the narrative board as:

1. Information entities: visualizations, text, annotations. These are mainly imported visualizations and annotations created during the analysis. For annotated visualizations, the annotation text, the visualization, as well as the annotated data target (specific part of the visualization that is annotated), are all defined as story entries. Users can also place text anywhere on the narrative board to further explain their story.

2. Relational entities: arrows, lines, html links. A story can include relationships between entities, such as causality. We provide visual arrows, lines, html links and other vectorial references to define relations between entities (Fig. 5.c).

3. Organization entities: visual grouping and sequences. Previous work [25] and our interviews emphasized the need of visual grouping of entities that are to be seen together in a story. We support this with entity grouping borders. Our experts also indicated the need to define a reading sequence for each entity, to help readers move through the story. We achieve this by allowing users to define the order of appearance of all entities through a sequence list (Fig. 5.c). Users can change the entity's sequence in the list, delete, or rename any entity. They can also define a playback time for each entity, a time for the entity to be in focus in the playback presentation mode. Audio can also be recorded for any entity, to be played during the story playback.

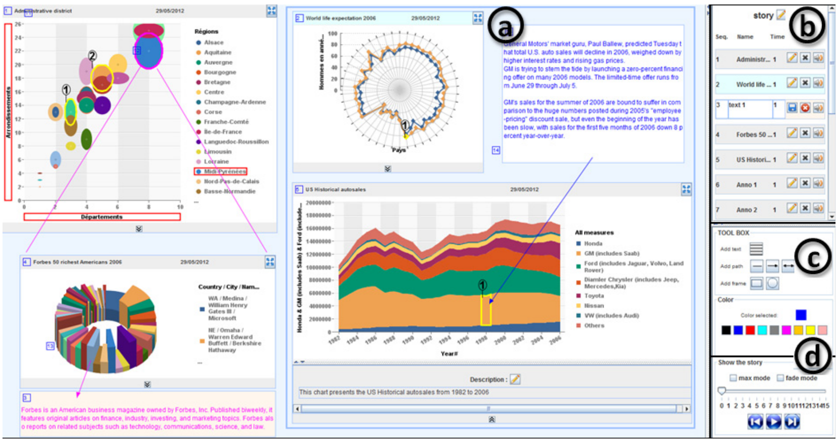


Fig. 5. (a) Narrative board containing all story entities arranged by the story teller: information, relational (arrows), organization (groupings) and emphasis entities, and numbers indicating author reading sequence. This sequence appears in a list (b), where authors can define playback properties and add audio commentary. Below is a pallet of relational and organization entities (c), and the Playback panel (d) to control playback through the story time line.

4. Emphasis entities: highlighting and zooming. To focus reader's attention to a specific story entity, users can add color highlights to any entity (e.g. a visualization, an arrow, or text entity), or select and highlight parts of visualizations (e.g. a bar in a bar chart visualization). This color highlighting can be present in the static story presentation (always visible), or during playback (the color highlight appears at a specific point in the story timeline). Besides color highlighting, users can add zooming highlighting to any story entity, to take place in a specific point in time during playback.

Playback. As discussed in (**R7**) our experts desire 2 ways of showing their stories to their readers, a static overview version that provides context and allows free exploration, and a playback that shows a recording of the author suggested path. Our narrative board acts as the static version. We provide three options of animated playback, giving focus to entities in sequence for a given duration. In the "color highlight", entities in focus change color to grab attention (default playback). In the "max playback", entities in focus are zoomed-in to the maximum possible size, taking up the entire narrative board. Finally, in the "fade mode", during playback, all entities except the entities in focus fade out (Fig. 6). As mentioned, authors can record audio to accompany the playback. Readers can pause the playback at any point to explore the story.

Fig. 6. Two playback modes: Max (a), and Fade (b)

Interactive Visualizations and Exploration. Our visualizations are imported with time-stamps referring to a "snapshot" of the dashboard data at the time of the story creation. Nevertheless, they are still connected to that version of the data and are still live. Thus users can still interact with them to perform brushing and linking actions (**R4**) and explore them further. By default we deactivate advanced interactions like drill-down/up, but authors can reactivate them through the visualization's properties.

The creators may also decide to take a new snapshot of the data (i.e. change the timestamp for the visualizations, and so the values of the data). Thus they can reuse the story structure for a new version of data (**R6**), which is particularly important for evolving dynamic data. Finally, users can select and replace any visualization with another, even from another dashboard loaded in the system. Thus they can reuse the story structure not only for dynamic data, but also for completely different datasets.

Thus our system integrates all the material used currently in BI story creation (**R2**): BI reports (visualizations and annotations), interactive visualizations, ways to indicate story structure, highlighting, optional story presentation in sequence, and text or audio explanations. It can then be shared with readers with access to the prototype. With respect to other BI systems, ours is one based on a user-centered design based and expert user requirements, and is unique in supporting annotations that link multiple points in the story, different story playbacks, and providing templates for BI stories.

7 User Feedback

We conducted two user feedback sessions to assess the usability and effectiveness of the system, both from the creator's and the reader's perspective.

7.1 Story Creation by Experts

We invited back two of the interviewed experts (P1,P2), and showed them the system in individual one hour sessions. Expert P1 was the experienced analyst (7 years) that has trained others. Expert P2 has been an analyst for 6 months and was trained by P1. They explored the prototype, saw story examples, and created stories themselves.

General Feedback and Recommendations. Both experts were very enthusiastic with the prospect of having access to such a system for their work. They found it easy to use and helpful, especially as it is integrated in the exploration/annotation dashboard tool. As P1 mentioned it "saves me from recreating any charts or annotations to present to others". It also saves time and effort not just in report creation, but also in communicating reports: as P1 explained, by using the prototype, there is no need for an explanation presentation to clients, for wiki pages or recreating interactive charts.

Both creators reaffirmed that story reading must be guided by the story creator, else the goal of the story may be lost. Currently they enforce this by the order of explanations used in their BI reports. They commented that the prototype supports this well with the numbered sequence for entities, while the playback (with fade out and zoom) guides readers through the story and keeps their focus on one entity at a time. But they appreciated that they can create and show a story on one page (even with scrolling), as opposed to the current multi-page reports, because as P1 mentioned "people tend to read only the first page, and explore less the following pages". And because it gives readers flexibility "the reader can always understand the position and link between parts and the overall story" and "look for more details when she wants".

Participants liked the ability to use the annotated dashboard template, saying that "this is how we want to present our analysis story, similarly to our dashboard". They both commented that annotations attached to data points are very important in pointing out to readers important data values "I add a lot of manual arrows to point to annotations that refer to specific data points on a chart, and their ideal value".

Both experts suggested our system should support two types of BI narrative stories: *(i)Fixed stories,* that present snapshots of datasets at specific points in time, yet are interactive (e.g. for filtering), to be shared as tutorials, explanations, or reports. This is the default case in our system. And *(ii) Online stories,* that present dynamically evolving data, and can have the same analytic scenario regardless of data values. Thus stories may have the same chart descriptions (e.g. what data is shown), the same KPI relations, and the same reading sequence. Here visualizations in stories are no longer snapshots, but are updated with data changes. We have implemented this extension.

Collaboration and Communication with BI Stories. Expert P2 commented on how this storytelling prototype can also be used as a means to evolve stories. Multiple analysts can integrate their own comments and knowledge in the story, encouraging peer learning, but also collaboratively creating more complete and detailed stories.

Expert P1 mentioned that she could envision using the system to iterate the definition of the story directly with her clients (a process conducted now by email or phone-calls). She envisioned clients adding themselves on the story further explanations on specific entities and their relations (to explain patterns), possible summaries of decisions they took based on the report, or highlighting what information is missing. This goes beyond collaborative analysis and storytelling: it directly empowers readers and becomes a medium to communicate what they want from the story.

In both cases, the prototype moves from a one-way communication to a collaboration medium, where the authoring of a story opens-up and evolves with the contributions of many users, and acts as an archive of knowledge and different points of view. Such a system, our experts explained, needs to clearly differentiate between the contributions of individual authors. We are currently exploring this extension.

Finally the story can be archived and used by new analysts that learn how to create stories (comment from our trainer analyst). P1 stressed how important such an archiving is for knowledge passing between analysts. The recently trained P2 stated the tool can help him further improve his reports by looking at the story structure of others.

7.2 Story Reading by Novices

We then ran a second session to evaluate the prototype from the reader's perspective, and thus close the story communication cycle. We conducted 40 to 50 minute sessions, with 5 BI novices. All were IT professionals, knew what a dashboard is but had never worked with one. Two had heard of BI reports but never used one. Participants were asked to (1) Read a BI report created by one of our experts, (2) Read a BI story (created by an expert in the previous session), and (3) Explore our prototype.

The story presented the progress of a development project from different perspectives: General (how is the project evolving in terms of finished code in a sprints time line, how many code components are added to a waiting list, and how many critical, major or minor code components are done each week); and Detailed (the progress of each development team in coding and testing each code component). The development is not progressing according to plan because many bugs fixed are not critical, whereas new bug reports coming in add critical bugs to the waiting list. The bottom right chart in the report highlighted the problem and the rest provide details.

Reading the Report. When given the report, all readers read the first page dashboard from left to right and up to down (which is not the "author" suggested order). They understood the goal (progress of a project) and what the charts displayed (e.g. bugs in waiting list) with the aid of chart titles and legends. But they all struggled to find the problem illustrated in the report. Only one participant noticed that the project development is not progressing over time, but she could not understand why. This supports our experts' comment that reports cannot be read without the supporting material.

Reading the Story. Participants found that reading a story was easier "it showed the facts in an understandable manner". All 5 readers found the system easy to use, understood the story, and were able to answer correctly comprehension question related to the story content. We report here our main observations and readers' comments:

Memorization: All users remembered even detailed aspects of the story and only went back to the system when answering detailed quantitative questions (e.g. to retrieve numbers from charts), while all general comprehension questions were answered without going back to the story visualizations and comments. They tended to answer using the same terms used by the expert analyst in the story. When asked similar questions on the report, they went back to it each time to search for answers.

Confidence: Readers were not confident in their understanding of the report as they had to draw their own conclusions. They expressed worry that they may have misunderstood or not noticed important points. While when reading the story, all felt confident, as their interpretations are confirmed by the analyst's comments.

Guidance: All readers appreciated the guidance in reading the story, both static and playback. They commented on the importance of both modes: ”the static mode permits to understand the whole story” and ”dig for facts”, while in playback it ”is easier to follow” the story sequence and focus on important data. Four participants preferred the fade playback, and three the max mode when focusing on an entity.

Understanding: All users understood the story using our tool, and found it easy to read and interpret. They commented on how it was easy to find answers both to qualitative and quantitative questions. The story structure showed clearly what is the problem, how to analyze it and how to find the cause. While they described the report as ambiguous, as they couldn't identify the relation between charts or KPIs.

Transmitting knowledge: All readers found annotations very helpful in explaining the relationships between charts and KPIs, and in teaching them the analysis logic they should follow. Four mentioned that the system can ”aid in transmitting different knowledge in the company between different users”, and two would like to use the system to communicate with their team their own data (even if they are not analysts).

Engagement: Three readers got very engaged with the story, as ”stories are more encouraging than static reports”. They began searching in charts to find how this problem can be solved, exploring the story outside the suggested structure.

The comparison of BI reports without supplementary material to our story prototype is clearly unfair and our goal is not to prove its superiority in terms of communication value. It serves to illustrate how our stand-alone prototype could be an effective means of communicating BI stories without additional material. A comparison of our tool against the collective BI story material used today is future work, but our experts already identified its superiority in terms of saving time on report creation, the sharing of interactive stories, and being a medium for collaborative story evolution.

8 Discussion and Conclusions

We identify for the first time the current practices and needs of BI storytelling. BI analysts often organize visually and communicate their analysis story to others. Nevertheless their tools do not allow easy transition from visual analysis to storytelling. They often use multiple tools, replicate work, and train their audience to understand their analysis. This reinforces the need for explicit storytelling support, missing in most existing visual analytics systems even in other domains, and provides insights on how to address this need for dashboards and other coordinated view systems.

Based on interviews and a paper prototyping session with expert BI analysts, we derived requirements for extending BI visual analysis dashboards to support storytelling. These can be adopted as-is by designers of BI systems, or inspire and inform visual storytelling research in other domains. For example providing an author suggested story structure, seeing the story in either a playback or it in a static overview, and imposing exploration constraints to interactive visualizations, can apply to other domains. Others, like the need to reuse story structure, may be BI specific and need to be reexamined for other domains. Using these requirements we implemented an extension to a BI dashboard, allowing transition from analysis to storytelling.

We evaluated our prototype with story creators and found the requirements and prototype meets their needs. More importantly, they highlighted the potential of storytelling tools as (possibly asynchronous) two-way communication channels, were authors communicate their findings, and readers also pose questions directly on the story. This empowers readers, informs authors of limitations of their story, and captures the evolution of the story. In BI it also accelerates story creation, that is highly client driven, and whose details are often lost in emails and phone calls. Tools can also act as collaborative platforms, where multiple authors refine, create variations of, and archive stories. Our findings open questions regarding the archiving and navigation mechanisms of story versions, trust in authoring changes, and maintaining a clear story focus. These questions open up exciting new research avenues.

We then presented a story from an expert to novice readers, to test the understandability of stories in our system, and its potential as a stand-alone tool for communicating BI stories. Readers understood the stories without training, and answered complex comprehension questions. Few previous works (e.g. [9]) evaluate visual storytelling systems from the reader's perspective, a crucial step in the communication cycle.

Our system was identified by experts as having great potential for reaching a broader audience, as little to no training is required to read stories, and for training analysis. Although designed as a stand-alone tool, it does not aim to replace expert analysts. As an expert mentioned, ours and other storytelling tools can aid novice analysts or readers to quickly read analysis results and focus their questions to experts on more complex aspects, such as methodology, goals, content details, etc.

We are finalizing our prototype to give to BI analysts. We aim to verify requirements gathered through expert self-reported with more direct observations of the use of the tool and its impact on current analysis and communication practices.

References

1. OLAP and OLAP server definitions. The OLAP Council (1995)
2. Baber, C., Andrews, D., Duffy, T., McMaster, R.: Sensemaking as narrative: Visualization for collaboration. In: VAW 2011, U. London College (2011)
3. Bier, E.A., Card, S.K., Bodnar, J.W.: Entity-based collaboration tools for intelligence analysis. In: IEEE VAST, pp. 99–106 (2008)
4. Boje, D.M.: The storytelling organization: A study of story performance in an office-supply firm. Administrative Science Quarterly, 106–126 (1991)
5. Boje, D.M.: Stories of the storytelling organization: A postmodern analysis of disney as "tamara-land". Academy of Management Journal, 997–1035 (1995)

6. Boyce, M.E.: Collective centring and collective sense-making in the stories and storytelling of one organization. Organization Studies, 107–137 (1995)
7. Brown, M.H.: That reminds me of a story: Speech action on organizational socialization. University of Texas at Austin (1982)
8. Cutroni, J.: Google analytics. OReilly Media, Inc. (2007)
9. Eccles, R., Kapler, T., Harper, R., Wright, W.: Stories in geotime. In: IEEE VAST, pp. 19–26 (2007)
10. Elias, M., Bezerianos, A.: Annotating bi visualization dashboards: Needs & challenges. In: ACM CHI 2012, pp. 1641–1650 (2012)
11. Farhi, P.: CNN hits the wall for the election (2008), `http://bit.ly/cnn-wall`
12. Few, S.: Show me the numbers: Designing tables & graphs to enlighten. Analytics Press
13. Forceville, C.: Educating the eye? kress and van leeuwens reading images: The grammar of visual design (1996). Language and Literature 8(2), 163–178 (1999)
14. Gershon, N., Page, W.: What storytelling can do for information visualization. Commun. ACM 44(8), 31–37 (2001)
15. Heer, J., Viégas, F.B., Wattenberg, M.: Voyagers and voyeurs: Supporting asynchronous collaborative visualization. Commun. ACM 52, 87–97 (2009)
16. Helmer, J.E.: Between horses: An ethnographic study of communication and organizational culture at a harness track. U. of Illinois at Urbana-Champaign (1989)
17. Johnson, M.: The moral imagination. University of Chicago Press, Chicago (1993)
18. Louis, M.R.: Surprise and sense making: What newcomers experience in entering unfamiliar organizational setting. Administrative Science Quarterly (1980)
19. Louis, M.R.: Organizations as culture-bearing milieu. Organizational symbolism: Monographs in organizational and industrial relations, pp. 39–54 (1983)
20. Ma, K.-L., Liao, I., Frazier, J., Hauser, H., Kostis, H.-N.: Scientific storytelling using visualization. IEEE Computer Graphics and Applications 32(1), 12–19 (2012)
21. Maclntyre, A.: After virtue. University of Notre Dame Press, Notre Dame (1981)
22. McWhinney, W.: A health confluence (1995) (unpublished paper)
23. McWhinney, W., Battista, J.: How remythologizing can revitalize organizations. Organizational Dynamics, 46–58 (1988)
24. Mitroff, I., Kilmann, R.H.: Stories managers tell: A new tool for organizational problem-solving. Management Review (1975)
25. Pirolli, P., Russell, D.M.: Introduction to this special issue on sensemaking. In: Human-Computer Interaction (2011)
26. Segel, E., Heer, J.: Narrative visualization: Telling stories with data. In: IEEE InfoVis (2010)
27. Shrinivasan, Y.B., van Wijk, J.J.: Supporting the analytical reasoning process in information visualization. In: ACM CHI 2008, pp. 1237–1246 (2008)
28. Snowden, D.: The art and science of story or 'are you sitting uncomfortably?' Business Information Review, 215–226 (2000)
29. Sole, D., Wilson, D.G.: Storytelling in organizations: The power and traps of using stories to share knowledge in organizations. Harvard Graduate School of Education (2002)
30. Tversky, B., Heiser, J., Lozano, S., MacKenzie, R., Morrison, J.: Enriching Animations, volume Learning with animation. Cambridge University Press (2007)
31. Viegas, F.B., Wattenberg, M., van Ham, F., Kriss, J., McKeon, M.: Manyeyes: a site for visualization at internet scale. IEEE TVCG 13, 1121–1128 (2007)
32. Wohlfart, M., Hauser, H.: Story telling for presentation in volume visualization. In: IEEE EuroVis 2007, pp. 91–81 (2007)
33. Wojtkowski, W., Wojtkowski, W.G.: Storytelling: its role in information visualization. In: European Systems Science Congress (2002)

Using Narrative Research and Portraiture to Inform Design Research

Connie Golsteijn[1,2] and Serena Wright[3]

[1] Digital World Research Centre, University of Surrey, Guildford, GU2 7XH, United Kingdom
[2] Department of Industrial Design, Eindhoven University of Technology, Den Dolech 2, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[3] Institute of Criminology, University of Cambridge, Cambridge, CB3 9DA, United Kingdom
info@conniegolsteijn.com, sw639@cam.ac.uk

Abstract. Employing an interdisciplinary perspective, this paper addresses how narrative research and portraiture - methods originating from, and commonly used in social sciences - can be beneficial for HCI and design research communities. Narrative research takes stories as a basis for data collection and analysis, while portraiture can be used to create written narratives about interview participants. Drawing on this knowledge, we show how a focus on narrative data, and analysis of such data through portraiture, can be adopted for the specific purpose of enhancing design processes. We hope to encourage design and HCI researchers to consider adopting these methods. By drawing on an illustrative example, we show how these methods served to inform design ideas for digital crafting. Based on our experiences, we present guidelines for using narrative research and portraiture for design research, as well as discussing opportunities and strengths, and limitations and risks.

Keywords: Qualitative research, methods, narratives, story-telling, narrative research, portraiture, design research, interaction design, craft.

1 Introduction

The use of qualitative research methods originating from social sciences – for example, interviews, ethnography, and data coding – is well established in HCI and design research communities. In practice, however, there is still a significant gap between disciplines, because social science methods adopted in HCI do not always provide a close fit to the method's original ethos, often 'fail[ing] to do justice' to the kinds of insights that such methods can provide [1, p.549]. There is still plenty of discussion about the role of the social sciences and humanities in the inherently 'interventionary' world of HCI, which was again illustrated by a vivid panel discussion about this topic at CHI 2012 [2]. This interdisciplinary paper aims to contribute to the discussion on how HCI may learn and benefit from closer investigation, appropriation, and collaboration with the humanities and social sciences, and enable the diversity of human life to come to life through our research. This paper is aimed particularly at researchers in interaction design who seek to generate qualitative person-centered data to aid both

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 298–315, 2013.

the design process and the understanding of the users, but who do not have an extensive social science background. As such, researchers familiar with social science methods may already be aware of some of the points made in this paper, although we hope that they may still benefit from the discussion around the adoption of such methods in a design process. Specifically, we address how the HCI field can benefit from adopting the practices of narrative research and portraiture from the social sciences, and we illustrate how these practices can be used together in a design research process, firstly, to gather, analyze and present data, and secondly, to inform idea generation activities. Narrative research and portraiture methods provide means to engage creatively and holistically with research participant data. As we will illustrate, this approach is beneficial for increasing understanding in users, their diverse motivations and behaviors, and the context of use. In addition, these insights can then be actively utilized in the design process to create novel and appropriate design solutions which are sensitive to these diversities.

The next section will address narrative research and portraiture in the social sciences, as well as the use of related methods in HCI and design. We then present an example of the use of narrative research and portraiture in a study on craft and ideation for 'digital craft', followed by a section with specific guidelines on how to employ these methods in design and HCI. We conclude with a discussion of opportunities and strengths, and limitations and risks of using these methods.

2 Related Work

2.1 Narrative Research, and Related Methods within HCI

Some twenty-five years ago, Bruner [3] posited the concept of 'life as narrative', arguing that human beings construct meaning, make sense, and engage in 'world making' [3, p.11] through 'narrative' – that is, through creating, telling, hearing, recording, and reading stories. Relatedly, the field of narrative research seeks to engage analytically with the storied ways in which we make sense of meaning and experiences, within the wider context of our social world and those social others within it [3, 4]. As such, narrative research is particularly useful for exploratory research projects, which seek to engage with experience and meaning-making processes of diverse individuals or groups, with such approaches being utilized across a range of subjects in the social sciences, including education [e.g. 5], health [e.g. 6] and crime [e.g. 7], as well as representing a primary staple of ethnographic research in any field [8].

The use of 'narrative' data has long been established and recognized as crucial in the attempt to understand users' needs within the HCI research community. In the main, however, we would argue – as Dourish did with reference to ethnographies [1] – that narrative-based research methods have been read too narrowly within the HCI and design field. For example, narratives and story-telling have featured primarily as an outcome or goal of design research [e.g. 9, 10, 11] rather than being fully embraced as a research approach across the entire process. Narrative research approaches appear almost exclusively within experience-centered design, which 'aims to understand and design digital technologies that support rich, social and meaningful experiences in our everyday lives' [12, p.1506, 13], rather than engaging in depth with the stories and lives of the research subjects. Here, the concept of 'narratives' has, for example, been adapted as a conversational interview technique, through the use of cultural probes [14], whilst other uses of 'narrative' in HCI research focus more on possessions and technologies than on individual lives, e.g. 'deep narratives' [15] and technology biographies [16]. Moreover, whilst some HCI methods share some common ground with narrative research – e.g. contextual inquiry's commitment to increasing understanding of users and their actions 'in situ' [17] – they lack other crucial aspects (that is, CI focuses solely on the user in relation to their work, whilst most narrative studies within the social sciences go beyond the phenomenon of interest, engaging with the personal, social, and cultural life history of users). Similarly, whilst ethnographic approaches to design research [e.g. 1, 18] share some key aims with the narrative research and portraiture methods that we adopted for our research (e.g. deep understanding of individuals; rich descriptions of subjects and their environments; positive bias towards subjects' perceptions), they have been read 'too narrowly', with an over-emphasis on 'implications for design' [1]. In contrast, instead of focusing on 'implications for design', the approach we are advocating in this paper aims to increase the importance of ethnographic data, by importing narrative data, integrity intact, into the ideation phase. Despite the presence of some common goals more broadly speaking, however, this does not render ethnography and narrative research methods synonymous; as noted by Lawrence-Lightfoot, 'key contrasts' exist between the two [19].

In conclusion, whilst there is clearly a tradition of utilizing storied and narrative-centered approaches within the field of HCI research, currently this appears to be primarily limited to the area of experience-centered design. Furthermore, methods designed within the social sciences have not always been adapted in a way that remains faithful to their original ethos.

2.2 Portraiture, and Related Methods in Design and HCI

Whilst a diverse range of methodologies exist for analysing storied data [20], the one that concerns us here is that of the 'research portrait' (often referred to simply as 'portraiture'). A research portrait is a written narrative – for example, about an interviewee – which aims to 'capture the richness, complexity, and dimensionality of human experience in social and cultural context' [21, p.3]. The purpose of this is to attend to 'the aesthetic whole' [22, p.48] of the research subject(s), and as such, a portrait must be strongly specific and contextual [19, 21]. The use of portraiture methodologies is most dominant within the social sciences, primarily within the sociological study of education and educational leadership [e.g. 22], but its use extends to criminology [e.g. 4], psychology [e.g. 23], and health research [e.g. 6].

Similar methods to the research portrait also have some degree of prominence within HCI and design research. However, here they primarily exist as a means for distilling and communicating the results of ethnographic fieldwork. For example, Wright and McCarthy [24] highlight the narrative method of 'ethnographic vignettes' – 'short pen pictures of people in a setting [which] have been used to capture the felt experience of working in a particular place or setting' [24, p.642] – as a means to elicit empathic responses from readers [25]. Whilst Wright and McCarthy's method shares with narrative and portraiture an interest in 'felt experience' and the use of 'pen pictures' (which are, in essence, short research portraits) [24, p.642], it lacks our study's full integration across the design process, e.g. as a dedicated tool in ideation.

Similarly, 'scenario-based design' and 'character-driven scenarios' share with sociological portraiture a focus on capturing users' experiences in a storied form, which then facilitate the creation of fictional 'rounded' character descriptions [26], which act as placeholders for individual users. This is also a key aspect in the creation of 'personas', which are composite fictional characters, based on user data from field research, which embody multiple users' unique characteristics and beliefs [27]. Other methods related to personas are 'pastiche scenarios' (uses fictional characters from well-known cultural sources – e.g. novels, movies, plays – to encourage designers to explore alternative interpretations of technologies) [28], 'extreme characters' (uses fictional characters with exaggerated emotional attitudes, e.g. a drug dealer, or the Pope) [29], and 'design alter egos' (uses fictional characters based on the recollections and reflections of designers) [30]. Much like our own study, these approaches also utilize narrative/storied data within the ideation processes, however – to paraphrase Lawrence-Lightfoot and Davis – what they lack is the commitment to maintaining 'a view of the whole' of each individual user, which is the art and science of portraiture [21]. In the aforementioned portraiture-related approaches, character sketches or personas are most frequently summarized and/or blended descriptions of multiple users into one or more realistic and characteristic – yet importantly, fictitious – users [31, 32]. Such an approach risks a lack of depth, detail, and 'wholeness' that is so central to portraiture, potentially leading to superficial, and even erroneous, assumptions [30]. Moreover, such approaches eschew the important reflexive question of the role of the researcher in producing the data generated, and through doing so, effectively 'neutraliz[e] out of existence' the researcher in each individual instance [21, p.86]. This is highly problematic since it is 'crucial that [the researcher/s] voice be monitored' [21, p.86]. To fail to do so risks the dominance of the researcher's interpretation of the user's data, and the loss of what was originally important to the person being consulted.

Conversely, the portraiture approach addressed in this paper tells the story of each interviewed or observed person in a separate individual portrait, within the context of use and staying true to the real users, with an ever-present eye to reflexivity and researcher 'voice', considering at all times whose perspective is being presented when the researcher relays a point. We believe our approach therefore minimizes the risks of stereotyping and oversimplifying users and their experiences inherent to the aforementioned methods. In addition, we propose importing these holistic descriptions of real users directly in the ideation process. This ensures that attention remains focused on the diversity of the people in the target group throughout the process. This approach therefore contrasts quite strongly with the ways in which narrative data is used within personas, which are usually created between data collection and ideation phases, thus generalizing and summarizing – risking the loss of individual diversities - before ideation has begun. This may cause interesting insights to be lost earlier in the design process, which portraiture aims to prevent. To sum up, the narrative research and portraiture approach addressed in this paper distinguishes itself from other ethnographic data gathering approaches – e.g. ethnography and contextual inquiry – by engaging holistically with users' stories and by providing tools for integrating these stories throughout design processes, and from other data representation approaches – e.g. scenario-based design and personas – by retaining researcher reflexivity, and by carrying forward the stories of real users throughout the design process and into the ideation stages, which helps to retain the diversity of the target group, for both the communication of ethnographic fieldwork results and the generation of broad spectrums of ideas.

3 Illustrative Example: Using Narrative Research and Portraiture to Design for Digital Crafting

In this descriptive section, we present an illustrative example from our own study to demonstrate the ways in which a narrative and portraiture method of data generation and analysis contributed to our research within design and HCI. The section which follows on from this, which is more analytically-focused, will center on the ways in which the lessons we learned can be extrapolated for the benefit of wider design and HCI communities.

Broadly, the study we will address was concerned with 'everyday craft'; that is, the creative processes people engage in to carefully make things [33-35]. More specifically, we wanted to better understand everyday crafting practices with physical materials, in the attempt to initiate design explorations of how characteristics and processes of craft may be extrapolated to the digital realm. The study also looked to developing ideas for new products or systems related to 'digital crafting' (that is, crafting with digital materials and/or tools). Because in this paper we merely aim to illustrate the use of the method through the example of our craft study (and not so much to discuss the topic of craft itself), a literature review of what constitutes 'craft', and related HCI and design work, lie beyond the scope of this paper.

In total we interviewed eight individuals who were involved in crafting/making things with physical materials in a diverse range of settings. In order to explore the breadth of everyday crafting, we recruited individuals with varying levels of expertise, and included professionals, semi-professionals, and amateurs, hoping to cover all types of crafters from the 'certified […] genius', to those individuals who just 'seem[ed] to like making things […] in everyday life' [34, p.75]. Each individual was chosen specifically for their work in craft- or art-oriented disciplines, in the hope that they would inspire our development of the 'digital crafting' concept. As such, the interviewees were: a guitar builder, a jewelry designer, a hairdresser, a paint artist, a glass artist, a silk painter, a wood and metal hobbyist, and a mixed media artist. We felt that the narrative and portraiture approaches were particularly appropriate because we wanted to engage with the meanings these individuals attached to 'crafting' and the characteristics, processes and purposes behind its physical manifestation, in the hope that we could extrapolate these findings to better understand ‘craft’ within the digital realm.

In order to illustrate the procedure of our method we will describe how the first author went about interviewing one of the participants – Paul, a Dutch guitar builder – using a narrative approach, before describing the process of ‘portraiture’ (i.e. writing Paul’s research portrait), and how we used this to generate ideas.

3.1 Interviewing Paul Using a Narrative Approach

Paul’s interview took place in his home in a small town in the south of the Netherlands. Before going to meet him, we had developed an interview guide that would act to encourage storied responses, rather than potentially close such responses down (often an issue with standard semi-structured interview guides) [4] – after all, one cannot work with narratives if one does not provide the conditions for their construction in the first instance. This was therefore a crucial step, one which is explored in detail in Section 4.1 below, to prevent the tedium of repetition. Briefly, the interview guide consisted of a list of topics we were interested in rather than a concrete set of questions, and included: how and why the participant started crafting, the process of crafting, how they learned it, and materials and tools they used. We further left plenty of room for discussion of unanticipated topics that were brought up by the participant.

While the introduction was started in the living room, immediately thereafter the interviewer was invited to Paul’s workshop, which was located in the garage, which had been refurbished and dedicated to the craft of guitar building. Much like it is beneficial for contextual inquiry and ethnographic research more broadly, being in the guitar builder’s workshop also aided the narrative interview for a number of reasons. Firstly, it illustrated some of the topics Paul was talking about, and allowed the interviewer to better understand and document (both through taking notes and photographs) the context of the crafting practice. Secondly, it gave both interviewer and interviewee handles for new topics to address, and thirdly – crucially – it benefitted the narrative character of the interview as Paul naturally (without prompting) started telling stories about materials, tools, and examples in the workshop.

Because the main interest lay in exploring the breadth of the practice, it was important to let Paul talk about his craft and his workshop freely, in order to facilitate the generation of storied data that was personally relevant to him. We began the interview by asking questions that would elicit storied responses [again, see Section 4.1], such as “Can you tell me something about the kind of crafting you do?”, and “Can you tell me how and when you started building guitars?” During the interview, the participant was encouraged to draw on examples and stories of personal relevance to him, generating ideas previously unanticipated by the researchers. A new topic was only introduced by the interviewer when the participant had finished a story. Interviews lasted half an hour to little over an hour (44 minutes on average) and were audio recorded to allow the interviewer to more fully engage with the participant. The few written notes that were made focused mainly on aspects the audio recording would not capture, such as the interviewer’s observations and impressions during the interviews, e.g. on participants’ use of examples, the mood and personality of the crafter, and the appearance of the workshop, e.g.:

"As the interview takes place in his workshop, it gets hands-on by default and throughout the interview Paul keeps walking up and down the workshop, opening drawers, taking things from shelves, and handling tools and materials to show me exactly what he is talking about. I get the feeling the workshop further serves as a mental map to give Paul new handles for things to talk about and he visibly enjoys using the half-finished guitar parts lying around as examples."

3.2 Creating a Written Portrait

After the interview the audio recordings, written notes and photos were used to write up a 'research portrait' about Paul. Because of the narrative character of the interview, rich qualitative data was collected and a two-phased analysis process helped to identify which parts of the interview and which quotes provided interesting insights, and helped to retain and communicate a coherent picture of Paul as a craftsman. In the first phase, notes were taken on interesting comments and observations while listening to the audio recordings, reading notes, and looking at captured photos. During this phase, relevant sections of the recording were transcribed verbatim, e.g. Paul's explanation of why he likes building guitars, or his experiences learning the craft. Also notes were included on the context of the interview and when an example was shown.

In the second phase these notes were written up as a portrait which followed the structure of first introducing Paul's craft and the context of the interview, before looking at when and how he started, and the materials and tools he used, followed by any other interesting themes from the interview, including working by assignment, risks, and teaching guitar building. This meant that the portrait did not need to follow the sequence in which interview questions were asked. The portrait, in which the participant's name was anonymized, was a rich description supported with lengthy quotes from the interview where this was considered useful, e.g. because of the level of detail or the relevance to the research aims. The portrait further combined 'first order' narratives (those of the participant), and 'second order' narratives (the stories the researcher is conveying) [36], including interesting observations, and interpretations that would later become important for ideation, such as Paul's creation of his own tools:

"As I look around the workshop I see, apart from an impressive collection of the obvious tool such as saws, chisels, and files, many devices and tools that are unknown to me. Paul explains to me that he makes these himself to support parts of the process: 'Most of the work involved in building a guitar is precision work and each time you have to measure something there is risk of error, so you start looking for ways to limit this risk and create tools for this.' He modestly adds that the ideas for these tools do not all come from him, but also from colleagues, books and the internet. His self-made tools range from hand-powered tools (for example, a large, round, slightly hollow sanding disc for sanding the top panel of the guitar's belly, and a compass with a chisel to cut out a circular groove for the rosette around the sound hole), to advanced electronic devices (e.g. a sanding machine for shaping the large, thin wooden panels for the top and bottom of the belly, and an intricate-looking device for bending the thin panels for the sides of the belly with

the aid of a heating element). Paul tells me that when you start doing something as a hobby you have to prioritize and choose which devices to get within your financial possibilities. For the rest you have to make do with what you have, and 'what you can do yourself… it is also fun to build that.' He adds: 'sometimes I get so into making a certain tool and when that is finished, you can just sit down, look at it, and enjoy it. That's wonderful. […]'"

Paul also told the interviewer about how he had started teaching guitar building to small groups of students who, like him, *"don't want a cheap guitar; they want the adventure of building it. They want the experience of the development of that thing and feeling what happens with the wood."* He said he was reluctant to teach at first, but when he saw that so many people were interested he decided he wanted to share his hobby and expertise. He likes these sessions with students because they are interested, and the following closing excerpt from the portrait aims to capture how important this interest and appreciation from others is to Paul:

"Throughout the interview I have gotten a strong feeling for Paul's […] appreciation of my interest in his craft. He explains to me that sometimes people come over who just have a glance at his workshop, ask him questions like: 'So, how many guitars do you make a month?' and they leave after 15 minutes. 'They should just stay away,' in Paul's opinion. Not me, however, being a guitar-player myself I would have been unable to hide my enthusiasm and appreciation even beyond the scope of this interview, much to Paul's liking. As I prepare to leave he repeatedly thanks me for listening and chuckles: 'In 30 years' time, when I've made my 200th guitar, come back and I can tell you much more.'"

This further illustrates how researcher reflexivity was included in the portrait. Moreover, in recognizing the appreciation of the first author's genuine interest in guitar-making, it enabled us to use this reflection to understand the importance of teaching, authenticity, and commitment to Paul's crafting, which provided the basis for the 'Online Guild' idea outlined in the following section.

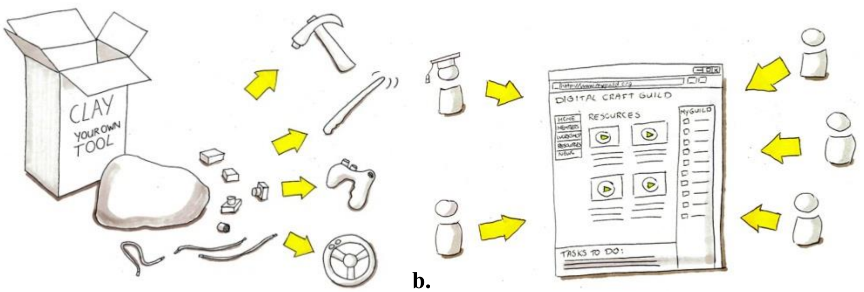
3.3 Idea Generation Based on Paul's Portrait

Because the explored practice of crafting was so broad and there was no predefined direction in which design ideas for its digital equivalent should be sought, the crafter portraits were considered useful to provide a focus to ideation compared to using themes or design directions. The ideation phase consisted of individual brainstorm sessions by the first author in which each separate session focused on a specific crafter. Inspired by our new understanding of what practical, hands-on crafting meant to those engaged in it, brainstorming centered on asking: "If this crafter would instead be crafting with digital media or digital technologies, what could be designed for him or her?" A brainstorm session began by writing down a few key points for each crafter that arose from the portraits. For Paul these points were: knowledge of materials, patience and spending a long time, teaching the interested and appreciative, making his own tools, and the workshop as a mental map. Ideas were generated around these and other themes arising from the portrait, and the portrait was used intensely during the session: sections were read and reread, which triggered new ideas. Ten ideas were generated inspired by Paul's portrait, which included:

- A craft timer that captures the time spent crafting an object in the object itself or on a timeline, and connects relevant occurrences in your life to the crafted object, or captures things you have done or said to be embedded hidden in the object – based on the time and patience Paul has for creating objects, and valuing slow processes.
- A digital craft workshop, such as a desktop or software lay-out, that can function as a mental map by being flexible in how tools and media are organized and by showing small previews around icons of specific tools and programs of what you have been working on last – based on the observation that Paul's workshop seemed to function as a mental map for him.

After this initial idea generation phase, ideas were distilled into a set of four or five idea statements that summarized and highlighted interesting ideas arising from the portrait, interview data, and design ideas, which for Paul included: contextual information implicitly visible in photos (expertise of materials); online workshops and guilds rather than a 'take what you need' mentality (see idea below); and create your own digital tool (see idea below). Finally, two ideas were selected and concept sketches were made for further exploration and discussion within the research team:

1. The Clay Tool (Figure 1a) allows you to create your own computer input device by making use of a set of sensors and actuators and a lump of clay, so you can make the appropriate tool for each task at hand instead of having to rely on manufactured, generic tools – based on Paul's creativity and interest in making his own tools for situations where standard tools do not suffice.
2. The Online Guild (Figure 1b) is a place where interested crafters can get together to share their love for their craft. Rather than being able to download anything, members have to be invested in the guild and contribute to the community. It is a more personal environment than a forum and its functions could include: learning from a remote master, exchanging experiences and skills with peers, browsing a digital workshop, or having tailored sets of tasks within a learning scheme set by another member – based on Paul's desire to teach only those who are interested and committed enough.



a. b.

Fig. 1. Concept sketches for a. the Clay Tool, and b. the Online Guild.

This method of narrative interviewing, portraiture, and ideation was applied to the eight interviews and over ten ideas per crafter were generated in the initial brainstorm rounds. Out of these ideas one or two ideas per crafter were selected, which resulted in a set of thirteen conceptual ideas for new forms of crafting with digital media or digital technology, based on the practical art as constructed within the research portraits we had created. These ideas were not only true to the crafters we interviewed, but also formed a varied set that addressed multiple angles on crafting. For our study, one of the ideas was chosen and developed into an interactive prototype for further evaluation and development, which lies outside the scope of this paper, but illustrates that the method is highly capable of supporting the generation of ideas that are suitable for going into further design development phases.

4 Lessons from Our Example: Applying the Tenets of Narrative Research and Portraiture to Design and HCI

4.1 Generating Narrative Data

Because we were interested in understanding the nature of crafting, our own interview guide – as noted above – included topics for discussion with each crafter (e.g. how they started, and their materials and tools), which were supplemented with a range of discussion 'prompts' (e.g. perfectionism, risks, and social aspects of the craft). This can be applied to any area of research, so readers wishing to adopt this method should similarly choose a list of topics – around ten to fifteen is ideal – which cluster around the area of substantive interest, and a list of related prompts which seek to provide depth in understanding the issue at hand. It is important that deviation from this guide – as happened in our study with the discussion of unanticipated topics brought up by the participant, such as personal challenge – should be considered positive, since it may generate ideas previously unanticipated by the research team.

Of course, having constructed a topic guide is of little use if one cannot ask a question likely to occasion a narrative response, and since the concept of narrative is inherently bound up with that of storytelling, it is crucial that the research methods chosen are capable of eliciting storied data. In this sense, narrative-focused interviews are distinct from run-of-the-mill, semi-structured qualitative interviewing strategies, where even the favoured 'open-ended question' can act to suppress – and even eradicate – the impulse and opportunity for storytelling [cf. 4]. To illustrate this, Hollway and Jefferson [4, p.35] recommended that researchers 'narrativise topics'; that is, 'turn questions about given topics into story-telling invitations'. For example, a question that we might have asked in a more semi-structured interview under the topic heading 'Learning about crafting' – e.g. "How did you learn your craft?" – could easily be answered in one short sentence, for example: 'Through my college degree and placements.' However, by 'narrativising' the question in the way suggested by Hollway and Jefferson above, and instead asking "Could you explain to me the processes and people by which you learned your craft?", we increased the likelihood of eliciting a personally relevant and detailed story about processes, experiences, and interactions with others.

Also crucial to the construction of narrative data is the idea of situated context. For this reason, and based on our own experiences, we would suggest interviews are undertaken within design environments or workspaces relevant to the study and design goals (as noted earlier, this is a common aspect of both narrative/portraitures, and methods of contextual inquiry and ethnography), although we acknowledge that this may not always be logistically viable. Researchers should also seek to ask questions which look to the broader personal histories and social influences behind an individual's reasoning, in order to provide shape and depth to the portrait.

An audio recorder may be more important during narrative-focused interviews than standard interviews, owing to the centrality of rich descriptions and quotes so crucial in the analysis and portraiture stages of the research. Moreover, using a recording device frees up the interviewer's note-taking to focus on the observations, non-verbal data, and reflexive considerations – which provide a depth and wholeness to the interview, and can be used to build up a far more detailed research portrait at a later stage.

In our craft study, it was also considered beneficial that the first author was involved in the whole process from interviewing the crafters, data analysis and writing up the portraits, to the idea generation. This helped to build up a thorough understanding of the crafters, and the characters they represented in the ideation. However, as a basis for idea generation, the portraits are believed to be a powerful tool to support designers even if they have not been involved in the data collection. Because of their richness and realistic qualities, being based on real, individual people, they are able to act as 'substitute whole' for even those without access to the raw data [4, p.70].

4.2 Writing the Portrait

Research portraits, which can be used as analytical tools for any design research involving experiential and storied data, are particularly useful when exploring insights of populations who have traditionally been denied a 'voice' within more mainstream research [21]. However, we also found it to be a useful tool for maintaining the presence of the participants within the analytical process. Therefore, you do not need to be working with an historically 'voiceless' population; simply, you need only an interest in keeping focus on the context (personal, social, and otherwise) within which participants' insights arise, and an acknowledgement of the benefits that can be gained from doing this rather than removing the individual early in the process (as is the usual practice with personas, for example).

A research portrait should include large, verbatim chunks of interview data. It should detail the setting in which the interview took place, and feature the researcher's feelings about the setting and the individual participant. In terms of structure, it should cluster around the key 'narratives', or storylines/plots, that underpin what the participant is telling you, 'documenting their voices and visions – their authority, knowledge, and wisdom' [37, p.51], as illustrated to some extent by the excerpts in the previous section. It should also include observations from the researcher, who acts both as witness to, and interpreter of, participant 'voices', and who engages in sketching the design context, and systematically 'scanning the action' [21, p.87], documenting important contextual observations.

An important tenet of both narrative research and portraiture is that of reflexivity (although not exclusively to them, since it is also central to ethnographic and feminist approaches, for example), which focuses on the importance of researcher(s) reflecting on the research scenario and their interaction within this context [38, 39]. When conducting narrative research, it is important that one always considers the ways in which one's own personal autobiography (think for example about gender, age, social class, educational/employment status and rank) might impact upon the biographies disclosed within the interviews [40], and the same is true of writing a research portrait, which is 'shaped through dialogue between the portraitist and the subject' [37, p.51]. Researchers employing this method should be aware – as we were – that their own backgrounds (in our case, interaction design and criminology), and even their hobbies (e.g. playing the guitar, or not) impact variously on their own views of the topics at hand, the ways in which interviews are conducted, and in which the portraits are written. For example in our craft study, where the first author with a design background conducted the interviews, some participants felt intimidated at first to talk about their craft because they felt the interviewer was 'very creative', and extra attention needed to be given to reassuring participants that we were not assessing their skills but were interested in their stories.

Whilst these interactions can be problematic if one is unaware of them and is writing up the findings oblivious to this, the reflexive researcher acknowledges such phenomena, writes him or herself into the research in order to demonstrate this, and makes clear in writing up the point at which first-order narratives become second-order. We also advocate following Miles and Huberman's advice, who suggest that participants be allowed to read, and comment upon, their own portraits, being sure to question whether our interpretation of lives are 'credible to the people we study' [in: 21, p.246]. This allows us to see where individuals may challenge the interpretations we have made from the data, and provide scope for us to more carefully reflect and consider those instances where our own autobiography may have acted to unwittingly shape that which we had written.

4.3 Ideation and Idea Development

Building the portraits into a tool for design ideation moves beyond their original uses in the social sciences, and employs them as an active part of this innovative phase. Based on our experiences, designing for a broad topic area in which the direction for design ideas was not predetermined, portraits can provide useful handles and focus for idea generation in those cases where 'anything is possible'. Of course, as with any ideation method, the use of portraits is not a guarantee for good ideas, but we found them to be effective when used as described below. We recommend that brainstorm sessions focus on generating ideas around the main question: "If we would be designing [the design goal or topic] for this person, what would that need to be? What would be important to them?" As we have shown, this can be done in a process of first distilling interesting or striking findings from the interview in the form of short statements, generating ideas around these findings, summarizing the key points from these ideas, and selecting or developing the main ideas. However, we have found it important to treat ideation as an iterative process and intensely use the portraits in this process: we recommend reading and rereading the portrait several times during the session because ideas may trigger different interpretations of the portrait and thus new ideas. Also, as in any idea generation approach, it is important to allow all ideas, no matter how unfeasible or 'crazy' they may seem.

In most cases more than one participant will be interviewed and as such each participant should get their own dedicated brainstorm session. We acknowledge that this is time consuming – however, we felt that in terms of outputs generated, this was a fair trade-off. As with most brainstorm approaches it is important to follow up an ideation phase with an idea selection and idea development phase. In this phase ideas are evaluated on originality, feasibility, cost, and any other practical demand the project may pose. For using portraiture in ideation we recommend at this point to put together ideas arising from different participants, look for overlap and possible opportunities to combine ideas, and select the most promising ideas.

Further attention in this phase should go to assessing how specific to each person an idea is and how it may be generalized to a larger audience. Despite the fact that our ideas were generated starting from one specific person, we found that the resulting design concepts could be extrapolated to other target groups and use contexts: through the process of idea generation, selection and development, ideas were generalized, categorized, summarized, and extended to larger target groups, making sure they were relevant beyond idiosyncratic individuals whilst retaining their unique relevance to the interviewee. Similar to the designing for extreme characters approach [29] and pastiche scenario [28] approaches, we found developing or reformulating concepts to reach a broader audience afterwards generally easy to do, and the unique inspiration the individual initially provides, weighted up to this extra evaluation step. On the whole, we found that the embracing the portrait across the research process helped us to maintain the individual inspiration and diversity for idea generation that could have easily have gotten lost if categorization, generalization, and summarization had been done immediately after the interviews, as is the case with personas. As such, using portraiture for design takes into account the diversity of different people within a target group throughout the whole design process, and retains this attention for the individuals until the conclusion of the ideation process.

5 Discussion

Drawing on our experiences using narrative research and portraiture approaches in a design research process, we will critically assess the opportunities and strengths, and limitations and risks of these methods. Being highly qualitative in nature, and resulting in different findings for each participant, it is difficult to validate the method presented in this paper in any traditional sense, since its efficacy cannot objectively be measured. Further, because of the unique dialogue created between researcher and each researched individual, replicating any such study would be 'exceedingly difficult' [37, p.55]. However, we believe the strength of this method goes beyond the success of any individual idea generated, or the level of increased empathy for the user. Its strength extends to the possibilities of studying and designing for topics which are broad and undefined, and which require a great deal of attention to diversity within the topic area and the target group throughout the design process, as we have aimed to illustrate. Moreover, employing a narrative and portraiture approach in the ways we have described allows for retaining the accounts of the individual users throughout ideation phases while generalizing and summarizing afterwards, as opposed to other methods, such as personas, which do so earlier in the process and thus increase the risk of oversimplifying and loosing interesting design opportunities.

Following the guidelines we have provided we believe this method can be applied to a large number of different topics and projects, with 'craft' being just one example. We see the described method as exploratory, both regarding the topic under study and the ideas generated, and as with most methods aiming at ideation, further development of the ideas and evaluations for feasibility are required in later stages of the process. We are further aware that the excerpts of the portrait provided in this paper may seem, to some extent, decontextualized – however, this is due to the limited space for the inclusion of full portraits. We believe that narrative research and portraiture have great potential for providing more contextualization of interview findings, and the 'full picture' about an interviewee, because they combine participant and researcher narratives. They can therefore be a valuable tool in design processes, as long as the researcher takes into account and documents contextual specifics in the portrait, i.e. social, cultural, of the topic or activity that is under study in the interview.

5.1 Opportunities and Strengths

As we have shown, portraiture symbolises a creative means of organising and presenting research findings that can further be used to inform design activities. It forms a departure from often-employed thematic techniques of analysis, which may otherwise undermine the 'holism' of the user. In presenting the data holistically – including verbatim data from the transcript, researcher's observations about the participant and their environment, as well as the use of contextual visual data (photos of the workspaces, materials, and tools) – the researcher may also find an 'overarching vision' [21, p.248]; that is, in viewing the 'whole' as well as the thematic. While more commonly-used methods such as personas are often created after data collection by summarizing and blending multiple users into one or more fictional users – thereby letting the real users fade to the background *before* idea generation – portraiture takes into account the diversity of individual, real people, and retains valuable insights about these people throughout the design process. This can be greatly beneficial in preventing the adoption of superficial and erroneous assumptions [30] early in the process while ideation has yet to begin.

As our case has illustrated, the portraiture method also offers the opportunity, particularly in explorative research where little is known about the topic or design, to generate a range of emergent themes and ideas. For the craft project there was no predefined direction in which design solutions should be sought, which made the data collection and analysis cover a large variety of topics. Where a thematic analysis did not provide enough focus for the idea generation because of the breadth of the study and the large number of potential themes to explore, the portraits acted to provide depth and focus, because rather than thinking about ‘anything’ we could think about the needs of one specific person – a real person – and what could be designed for this person. Afterwards ideas could be extrapolated to a larger target group. Apart from focusing the idea generation, the portraits provided new insights which led to new, out-of-the-box ideas. When looking at designing based on portraits as opposed to designing based on specific design requirements or research questions, the portraits were perceived to be less restrictive and limiting. Quite simply, this is because whilst strict design requirements can act to stifle creativity and idea generation, portraits opened up the design space and allowed for the removal of limiting criteria, which was of vital importance to our brainstorming process.

5.2 Limitations and Risks

A potential limitation of the narrative approach is that not everyone will be able to narrate their lives and experiences. Although we did not encounter this in our study, not all people are equally capable of giving responses in storied form, even when prompted to do so [20]. This may lead to a distressed participant, a frustrated researcher, and a wasted time slot for all concerned. It may therefore be worth having a back-up interview schedule, with more structured questions for those who – as the interview progresses – appear to lack the impulse, inclination, or ability to narrate.

As already noted, narrative research approaches require a change in interview strategy: questions must provide opportunities to elicit storied responses [4]. This introduces the risk that the novice interviewer may experience difficulties, which could result in the data generated not being suitable for the intended analysis or ideation. Bear in mind also that with this approach a single interview can generate many hours of data. Therefore it is important that whilst participants are given time and space to compose their stories, your topic list should be short and focused. .

A narrative approach to research and analysis encompasses a range of unique ethical concerns in addition to those traditionally associated with qualitative research, particularly around anonymity, because of the level of contextual and personal detail [41]. The ways in which the words and meaning of the research participants are represented is also a contentious issue, and researchers must also be wary around claims to ‘give voice’, when the reality is that the main aim of the research is to help the interviewer, not the interviewee [42]. These critiques are counterbalanced by practices of self-awareness, reflexivity, and the explicit identification of the researcher’s own perspective within the portrait [19], as we noted above in Paul’s portrait. Furthermore, the perceived lack of means to assess the ‘validity’ is answerable by reference to the championing of multiple perspectives and ‘situated truths’ rather than absolutes, which characterizes both portraiture and narrative inquiry more broadly [20, p.185, 21]. Ultimately, it remains the task of the responsible, reflexive, and conscientious researcher to see to it that these concerns are met, which may provide a challenge to the creation of portraits based on narratives.

Despite these challenges, narrative research approaches and portraiture have widely been used within the social sciences, and with our treatment of the methods and by providing guidelines to how they may be used for design and HCI, we hope to inspire researchers in these fields to continue to learn from social science practices and adopt these methods where appropriate to benefit their design and research processes.

6 Conclusion

As we have shown in this paper, the holistic incorporation of the portraiture approach to data analysis into idea generation processes can offer some additional benefits compared to existing methods within HCI and design research which utilize narratives and storied data. For us, the portraiture method comes into its own within the realm of design research, since it has the potential for dual functionality within this field. Firstly, it represents a deeply holistic and contextual means for analyzing and disseminating findings, thereby facilitating rich understanding of the users and their design needs. Secondly, in doing so, it can act as an active catalyst for innovative ideation, informing, shaping, and enhancing the subsequent design process. We have illustrated the use of narrative and portraiture methods for design research with an illustrative example from our research on digital crafting, and have provided transparent documentation as to the ways in which we achieved this. We have critically discussed the opportunities and strengths offered by the incorporation of these methods in our own study, as well as addressed its limitations and risks. We hope we have not only indicated the benefits these specific analytical methods hold for HCI and design communities, as part of the wider trend towards narrative-centered research in these disciplines, but also the ways in which – as methods which hold at their core a commitment to maintaining participants' voices and a sense of their various individual needs – these represent key methods in the important endeavor of 'designing for diversity'.

Acknowledgements. This work was supported by Microsoft Research through its PhD Scholarship Programme, and the Economic and Social Research Council. We further thank the participants in the interview study, our colleagues at the University of Surrey and Microsoft Research Cambridge, and in particular our PhD supervisors, David Frohlich, Abigail Sellen, Elise van den Hoven, Roger Tarling, and Karen Bullock.

References

1. Dourish, P.: Implications for design. In: Proc. CHI 2006, pp. 541–550. ACM Press (2006)
2. Bardzell, J., Bardzell, S., DiSalvo, C., Gaver, W., Sengers, P.: The humanities and/in HCI. In: Ext. Abstracts CHI 2012. ACM Press (2012)
3. Bruner, J.: Life as a narrative. Social Research 54(1), 11–32 (1987)
4. Hollway, W., Jefferson, T.: Doing qualitative research differently. SAGE, London (2000)

5. Sinclair Bell, J.: Narrative Inquiry: More than just telling stories. TESOL Quarterly 36(2), 207–213 (2002)
6. Williams, G.: The genesis of chronic illness: narrative re-construction. Sociology of Health and Illness 6(2), 175–200 (1984)
7. Presser, L.: Collecting and analyzing the stories of offenders. Journal of Criminal Justice Education 21(4), 431–446 (2010)
8. Van Maanen, J.: Tales of the field: on writing ethnography, 2nd edn. University of Chicago Press, Bristol. University Presses Marketing, Chicago, Ill (2011)
9. Bhömer, M.T., Helmes, J., O'Hara, K., Hoven, E.V.D.: 4Photos: a collaborative photo sharing experience. In: Proc. NordiCHI 2010. ACM Press (2010)
10. Cao, X., Lindley, S.E., Helmes, J., Sellen, A.: Telling the whole story: anticipation, inspiration and reputation in a field deployment of TellTable. In: Proc. CSCW 2010. ACM Press (2010)
11. Golsteijn, C., den Hoven, E.: Facilitating parent-teenager communication through interactive photo cubes. Personal and Ubiquitous Computing 17(2), 273–286 (2013)
12. Clarke, R., Wright, P., McCarthy, J.: Sharing narrative and experience: digital stories and portraits at a women's centre. In: Ext. Abstract CHI 2012. ACM Press (2012)
13. McCarthy, J., Wright, P.: Technology as experience. MIT Press, Cambridge (2004)
14. Wright, P.C., McCarthy, J.C.: Experience-centered design: designers, users, and communities in dialogue. Synthesis lectures on human-centered informatics. Morgan & Claypool Publishers, San Rafael (2010)
15. Jung, H., Bardzell, S., Blevis, E., Pierce, J., Stolterman, E.: How Deep Is Your Love: Deep Narratives of Ensoulment and Heirloom Status. International Journal of Design 5(1), 59–71 (2011)
16. Blythe, M., Monk, A., Park, J.: Technology biographies: field study techinques for home use product development. In: Ext. Abstracts CHI 2002. ACM Press (2002)
17. Wixon, D., Holtzblatt, K., Knox, S.: Contextual design: an emergent view of system design. In: Proc. CHI 1990. ACM (1990)
18. Millen, D.R.: Rapid ethnography: time deepening strategies for HCI field research. In: Proc. DIS 2000, pp. 280–286. ACM Press (2000)
19. Lawrence-Lightfoot, S.: Reflections on portraiture: a dialogue between art and science. Qualitative Inquiry 11(1), 3–15 (2005)
20. Riessman, C.K.: Narrative methods for the human sciences. Sage Publications, Thousand Oaks (2008)
21. Lawrence-Lightfoot, S., Davis, J.H.: The art and science of portraiture. Jossey-Bass, San Francisco (1997)
22. Chapman, T.K.: Expressions of 'voice' in portraiture. Qualitative Inquiry 11(1), 27–51 (2005)
23. Davis, J.H.: Balancing the whole: Portraiture as methodology, in Qualitative research as psychology: Expanding perspectives in methodology and design. In: Camic, P.M., Rhodes, J.E., Yardley, L. (eds.), pp. 199–217. American Psychological Association, Washington, DC (2003)
24. Wright, P., McCarthy, J.: Empathy and experience in HCI. In: Proc. CHI 2008. ACM Press (2008)
25. Orr, J.E.: Talking about machines: an ethnography of a modern job. Collection on technology and work. ILR Press, Ithaca (1996)
26. Nielsen, L.: From user to character: an investigation into user-descriptions in scenarios. In: Proc. DIS 2002. ACM Press (2002)
27. Cooper, A.: The inmates are running the asylum. Sams, Indianapolis (1999)

28. Blythe, M.: Pastiche scenarios. Interactions 11(5), 51–53 (2004)
29. Djajadiningrat, J.P., Gaver, W.W., Frens, J.W.: Interaction relabelling and extreme characters: methods for exploring aesthetic interactions. In: Proc. DIS 2000. ACM Press (2000)
30. Triantafyllakos, G., Palaigeorgiou, G., Tsoukalas, I.A.: Design alter egos: constructing and employing fictional characters in collaborative design sessions. In: Proc. BCS HCI 2009. British Computer Society (2009)
31. Blomquist, Å, Arvola, M.: Personas in action: ethnography in an interaction design team. In: Proc. NordiCHI 2002. ACM Press (2002)
32. Chang, Y.-N., Lim, Y.-K., Stolterman, E.: Personas: from theory to practices. In: Proc. NordiCHI 2008. ACM Press (2008)
33. Csikszentmihalyi, M.: Creativity: flow and the psychology of discovery and invention. Harper, New York (2010)
34. Gauntlett, D.: Making is connecting: the social meaning of creativity from DIY and knitting to YouTube and Web 2.0. Polity Press, Cambridge (2011)
35. Sennett, R.: The craftsman. Yale University Press, New Haven (2008)
36. Harling Stalker, L.L.: A tale of two narratives: Ontological and epistemological narratives. Narrative Inquiry 19(2), 219–232 (2009)
37. Hackmann, D.G.: Using portraiture in educational leadership research. International Journal of Leadership in Education: Theory and Practice 5(1), 51–60 (2002)
38. Hill, D.A.: The poetry in portraiture: Seeing subjects, hearing voices, and feeling contexts. Qualitative Inquiry 11(1), 95–105 (2005)
39. Elliot, J.: Using narrative in social Research: Qualitative and quantitative approaches. Sage Publications, London (2005)
40. Cotterill, P., Letherby, G.: Weaving Stories: personal auto/biographies in feminist research. Sociology 27(14), 227–237 (1993)
41. Smythe, W.E., Murray, M.J.: Owning the story: Ethical considerations in narrative research. Ethics and Behavior 10(4), 311–336 (2000)
42. Atkinson, P., Delamont, S.: Rescuing narrative from qualitative research. Narrative Inquiry 16(1), 164–172 (2006)

Hoptrees: Branching History Navigation for Hierarchies

Michael Brooks, Jevin D. West, Cecilia R. Aragon, and Carl T. Bergstrom

University of Washington Seattle, USA
{mjbrooks,jevinw,aragon,cbergst}@uw.edu

Abstract. Designing software for exploring hierarchical data sets is challenging because users can easily become lost in large hierarchies. We present a novel interface, the hoptree, to assist users with navigating large hierarchies. The hoptree preserves navigational history and context and allows one-click navigation to recently-visited locations. We describe the design of hoptrees and an implementation that we created for a tree exploration application. We discuss the potential for hoptrees to be used in a wide variety of hierarchy navigation scenarios. Through a controlled experiment, we compared the effectiveness of hoptrees to a breadcrumb navigation interface. Study participants overwhelmingly preferred the hoptree, with improved time-on-task with no difference in error rates.

Keywords: Navigation, tree visualization, hierarchy, breadcrumbs, visual interfaces, usability.

1 Introduction

Numerous visualization techniques have been developed to help users extract useful information from *hierarchical* data sets [1]. In large trees, the huge number of nodes presents a significant challenge. Many interfaces for browsing hierarchical structures show only a subset of the full hierarchy at one time; the user alters the view by refocusing on a new section of the tree, navigating without viewing an overwhelming amount of information at once.

Narrowing the focus solves one problem, but in doing so creates another: focusing on a subset of the tree makes it easy to get lost. Users seeking to return to a previous focus must exert additional cognitive effort to remember how the previous location relates to the current location, and must continue to remember that relationship as they navigate through the tree to the new focal point. Due to the limitations of human working memory, users may need to revisit locations multiple times, especially when comparing several areas in the tree.

Navigation tools such as breadcrumb trails assist users with navigating complex website hierarchies, and browsing history features are included in all major web browsers. However, with either of these tools, if a user follows a non-linear series of steps through the hierarchy, part of the user's browsing history becomes inaccessible. Moreover, these techniques fail to relate the structure of the user's browsing history to the organization of the hierarchy.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 316–333, 2013.

In this paper, we present the hoptree, an interactive graphical interface for exploring hierarchical data that addresses these issues by displaying a clickable, branching history of nodes visited, structured according to their relationships in the hierarchy. The hoptree was designed to reduce cognitive effort by preserving and displaying recently visited paths in the hierarchy. We describe the design of the hoptree, and a prototype implementation that we have released for public use in web-based tree navigation applications. To evaluate the effectiveness of the hoptree, we conducted a controlled usability study comparing the hoptree widget to breadcrumb trails. With the hoptree, users completed comparison tasks more quickly, with fewer mouse clicks, and with greater satisfaction.

2 Related Work

2.1 Browsing Hierarchies

Analysis and understanding of large, hierarchically structured data sets is of critical importance in many fields. The visualization research community has studied this problem, leading to a variety of visualization and exploration techniques. These employ strategies to make large, complex trees easier to interpret, e.g. specialized zoom techniques [2–4] and layout algorithms [5, 6]. This work has led to a variety of innovative designs and strategies for presenting the most important information to users and hiding unneeded information, while simultaneously maintaining a sense of location and context within the tree, such as the SpaceTree [7]. However, most existing techniques are targeted to specific tasks or domains and in large hierarchies, becoming lost continues to be a challenging problem.

2.2 Breadcrumb Navigation

For many classic hierarchy exploration and navigation tasks, such as browsing the web or working with a file system, the sophisticated techniques in the literature have not been widely adopted. The hoptree widget is a simple design that could make tree navigation tasks easier and faster in a variety of general applications.

Breadcrumbs, or breadcrumb trails, are widely employed for helping users navigate on the web. Breadcrumbs usually consist of a linear sequence of links that provide a sense of location and facilitate quick navigation. Three types of breadcrumb trails are distinguished based on whether the trail reflects the application's hierarchy, the user's navigation history, or a set of dynamic attributes [8].

The benefits of breadcrumbs have been studied within the web usability community. A 2003 lab study of navigation within a test website found that only 6% of users' total page clicks were on the breadcrumbs, and did not detect any significant efficiency gains from using the breadcrumb [9], although the breadcrumb trail did lead to more accurate mental models of website structure.

Hull found that demonstrating and explaining the use of breadcrumbs beforehand improved efficiency for search tasks [10]. Look-ahead breadcrumbs, an augmentation of traditional breadcrumbs where clicking on the breadcrumb trail provides a menu with pages reachable from that item [11], were preferred by participants in a lab study, although no significant speed improvements were detected [12].

2.3 History Tracking

History tracking functionalities are available in many software applications, where they help users explore and recover application states. For supporting visual analysis tasks, Heer et al. describe a design space of graphical history tools and provide a review of recent research in this area [13]. History tracking widgets with branching structures have also been investigated for visual analytics tasks [14, 15]. For history tracking in web browsers, the MosaicG graphical history constructed a node-link representation of web browsing history with pages connected according to the order that they were visited by the user [16]. Similarly, PadPrints constructs a graphical hierarchy of pages that the user visits in their web browser [17]. PadPrints was found to reduce the number of page accesses and reduce the time taken to complete tasks requiring revisiting pages. Graphical history mechanisms can facilitate exploration and iterative analysis because they provide the ability to easily return to a prior state.

These history tracking tools present a history of linked *states* or *actions*. In the terminology of Cockburn and Greenberg's analysis of requirements for effective revisitation tools for web browsing, MosaicG and PadPrints structure their page display organization temporally, according to the user's visit history [18]. Our goal in this paper is to help the user understand and navigate a predefined hierarchical information structure. Accordingly, the hoptree primarily structures the graphical history according to the underlying information hierarchy, not according to user behavior.

3 Hoptree Design

We describe the features of the hoptree widget and the navigational problems each feature is designed to solve. We explain the hoptree design in terms of how it differs from breadcrumb trails.

3.1 Preserving Navigation History

Traditional breadcrumb trails display the "path" to the current location in the hierarchy. Hoptrees display not only the path to the current location, but also the paths to previously visited locations. Before the user begins exploring the tree, the hoptree widget shows only the path to the current node (e.g. the tree's root node). With each navigation action, the path to the new location is added to the hoptree's representation. Thus, the hoptree gradually builds up a more complete diagram of the hierarchy that the user is exploring (Fig. 1).

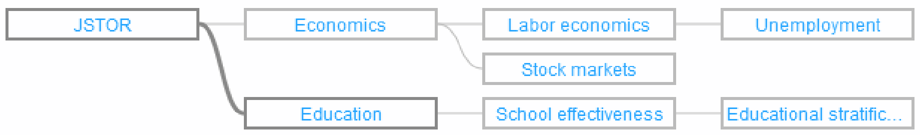


Fig. 1. A hoptree showing nodes that the user recently visited. The user is currently viewing the *Education* node.

Because it shows multiple paths at once, the hoptree helps users construct a mental model of the tree. While a breadcrumb trail may help users remember where they are relative to the root of the tree, the hoptree may also help the user remember how their current location relates to several previous locations.

The hoptree lays out its node-link representation of the hierarchy according to a left-to-right, top-to-bottom algorithm. Within the bounds of the hoptree widget, the root node is positioned on the left. Child nodes are positioned to the right of their parents, with siblings grouped together and aligned vertically. Nodes are connected to one another by curved edges.

3.2 Interactivity

As with breadcrumb trails, the hoptree widget is interactive. The user may click on a node in the hoptree to quickly "hop" to that position in the hierarchy. This increases the speed with which users may navigate the tree. The hoptree is especially effective when the user wishes to compare several tree locations, because it facilitates quick revisiting of recently accessed nodes.

3.3 Automatic Pruning

A crucial aspect of hoptrees is a pruning strategy by which tree locations that are no longer of interest are removed from the display. Without such a pruning strategy, the hoptree would rapidly increase in size and complexity as the user explored new areas of the hierarchy, becoming unwieldy and hard to use.

The hoptree's design assumes that, most of the time, users do not require access to their entire history. The pruning strategy we selected allows the hoptree to display up to three different tree leaves. If the user's next navigation would result in a fourth leaf being added to the hoptree, the update algorithm removes the oldest leaf in the hoptree, along with all of its parents that are not parents of a remaining leaf. Fig. 2 illustrates how a branch is removed after a typical update.

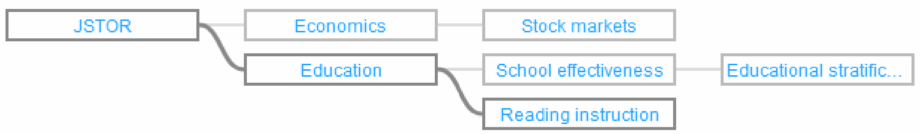


Fig. 2. The same hoptree from **Fig. 1**, after the user has visited *Reading instruction*, causing a layout adjustment and the pruning of the *Labor economics* subtree.

Displaying *three* branches provides a balance between utility and usage of screen real-estate. At least two branches are needed to support simple two-way comparison tasks. Showing three branches also permits comparisons among three locations, or, alternatively, the preservation of a related branch while two other locations are compared. The benefits of displaying more than three branches are not as clear; we leave alternative pruning strategies and branch limits to future work.

All decisions about which tree locations to preserve and which to prune are handled automatically by the hoptree update algorithm, eliminating maintenance work for the user. The tradeoff is that when the heuristic for estimating the relevance of nodes is incorrect, nodes that the user actually planned to revisit may occasionally be pruned from the tree. Thus, it is also be reasonable to investigate ways of giving the user more direct control over the hoptree's pruning strategy.

4 Prototype Implementation

We created an implementation of the hoptree for a specific tree exploration tool, the Gender Browser, an application under ongoing development in our research group. The Gender Browser[1] is implemented in JavaScript for display in HTML5-compliant web browsers. For completeness, we briefly describe the Gender Browser because it provides the context for our experiment. Then, we describe the architecture of our prototype hoptree implementation.

4.1 The Gender Browser

The Gender Browser is a tree exploration application that shows gender patterns in scholarly authorship across academic disciplines. The tree structure comes from automatic hierarchical clustering of a database of published journal articles provided to us by JSTOR[2]. The database used in our study contained approximately 350,000 articles by 480,000 authors, from over 2,000 different journals.

The Gender Browser displays a tree of scientific disciplines, and sub-disciplines based on articles a clustering of the article citation network using the hierarchical map equation algorithm [19]. In the version we used for our experiment, the tree contained 15 top-level disciplines, each of these disciplines containing several sub-disciplines. There were 448 nodes total, with 283 leaves at a depth of about 6 levels. All but 36 (8%) of the disciplines were manually labeled by examining its most-cited papers. We approximated the ratio of female to male authors in each discipline by automatically assigning a gender to every author using an approach similar to [20].

The tree of academic disciplines is displayed using an Icicle Plot visualization built using the JavaScript InfoVis Toolkit (JIT) [21] and jQuery. Each node in the tree is broken into two blocks of different colors, illustrating the percentage of women and men authors in that discipline. When the mouse cursor hovers over a node, a tooltip displays the associated discipline's name and the percentage of female and male authors. Fig. 3 shows the Gender Browser with the hoptree positioned above it.

[1] `http://eigenfactor.org/gender`
[2] `http://www.jstor.org`

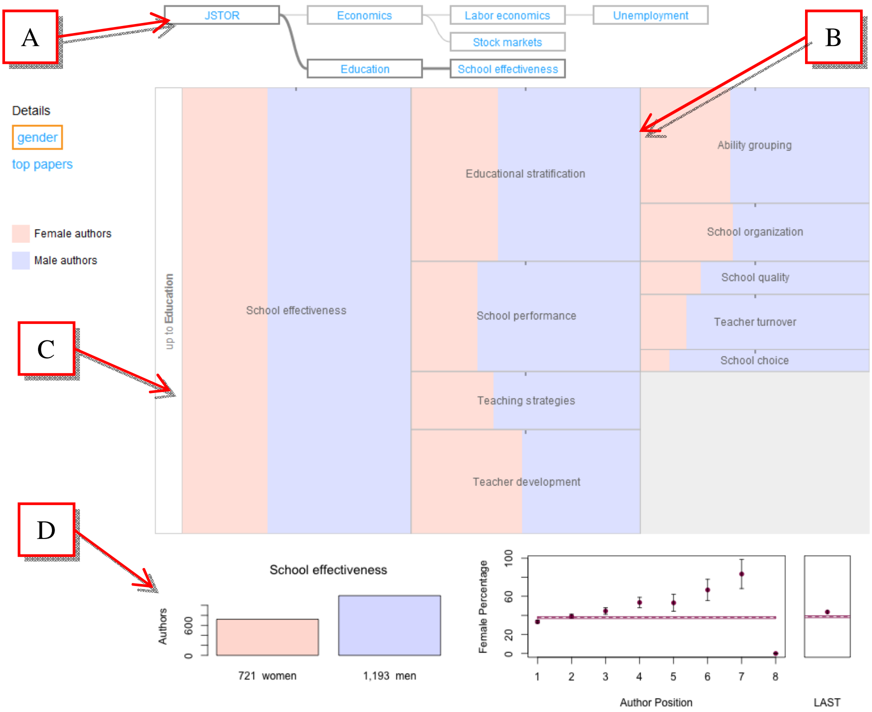


Fig. 3. The Gender Browser. The user has visited several specialties within *Economics*, and is now focused on *School effectiveness*, an area within *Education*. The hoptree (A) is positioned above the main icicle plot (B). On the left, the back bar (C) allows the user to return to *Education*. The detail views at the bottom (D) show the number of women and men authors in *School effectiveness* and the percentage of female authors in different positions on author lists. The red bar indicates the overall percentage of female authors in the field.

Below the icicle plot are two additional displays showing details about the gender breakdown of the currently selected discipline, including the absolute number of men and women authors in the discipline, and the percentage of women authors in the discipline for each position in article author lists. On the left side of the window, the "top papers" button allows the user to swap out the two detail displays with a list of the 10 most cited papers within the current discipline.

When the user clicks on a node in the icicle plot, the clicked node zooms to fill the height of the display. Parent and sibling nodes fade out, and child nodes expand. This allows the user to drill down into the hierarchy to explore finer disciplinary subdivisions. A bar on the left side of the icicle plot shows the name of the discipline that is the parent of the currently selected discipline. Clicking on this "back" bar will jump up one level in the tree with an animated transition.

4.2 Hoptree Implementation

We used the JavaScript InfoVis Toolkit to implement the hoptree prototype for the Gender Browser. The toolkit includes a SpaceTree visualization that was easily extended to display a hoptree. We wrapped the SpaceTree visualization with code to manage updates to the hoptree structure, including node additions and pruning.

Users can navigate by clicking on either the Gender Browser's icicle plot or nodes in the hoptree, and updates are synchronized between the two widgets. We have released our hoptree implementation as an open-source jQuery plugin that could be applied to a variety of web-based hierarchy browsing applications[3]. The hoptree prototype exposes a simple programming interface for integration with hierarchy browsing applications. On navigation events, the only information that must be passed to the hoptree is a string representing the new path.

5 Empirical Evaluation

In order to determine whether hoptrees are helpful for navigating a hierarchy, we compared our hoptree implementation to a breadcrumb trail navigation widget in a controlled experiment.

We developed three versions of the Gender Browser that differed only in the type of navigational support provided. A "plain" version lacked any special navigation support beyond the "back" bar for navigating up one level in the tree. A second version added a breadcrumb trail displayed above the icicle plot. The breadcrumb trail is implemented as a simplification of the hoptree (it displays only one branch at a time), providing a uniform visual appearance. The third version displays a hoptree widget above the icicle plot (Fig. 3).

5.1 Hypotheses

We designed the experiment to test three hypotheses. First, we expected that users would recognize the navigation advantages of the hoptree and breadcrumb interfaces over the plain version. Therefore, we hypothesized that they would express higher preferences for either the hoptree or the breadcrumb, and lower preferences for the plain interface. Because users might prefer familiar breadcrumb trails, we did not expect that the hoptree would necessarily have the highest preference scores overall.

Second, because the breadcrumb and hoptree versions provide instant access to nodes that the user may be interested in, we hypothesized that users would complete tasks faster in the breadcrumb condition than the plain condition, but would be fastest in the hoptree condition. For the same reason, we expected users to perform fewer mouse clicks while completing the tasks with the hoptree, followed by the breadcrumb trail and the plain interface.

Third, we hypothesized that for more complicated comparison tasks, the rate of correct answers would be highest with the hoptree version of the Gender Browser,

[3] `http://github.com/michaelbrooks/hoptree`

and to a lesser extent, the breadcrumb version. If users complete tasks faster using these interfaces, they may be better able to remember the data needed to make comparisons between several tree locations. For the same reason, we hypothesized that users would have higher confidence in their answers when using the hoptree and breadcrumb interfaces.

5.2 Experiment Design

We compared the three versions of the Gender Browser interface (Plain, Breadcrumb, and Hoptree) in a within-subjects experiment where participants completed three trials, one with each version of the Gender Browser. Each trial consisted of a set of eight questions that required comparisons among several different nodes at varying degrees of separation in the tree. We developed three distinct sets of questions that matched each other as closely as possible in difficulty and structure. We describe the questions in greater detail in the Tasks section, below.

To control for ordering, we carefully considered all possible pairings of *Interface* and *Question Set*, all possible orderings of *Interface* and *Question Set*, and orderings of the combined factor *Interface x Question Set*. We assigned participants to interface and question set orderings so as to balance all of these combinations as closely as possible, given the number of participants in our study.

Before beginning the experiment, participants were given a brief tutorial that explained the plain Gender Browser, but did not show or discuss either the breadcrumb trail or the hoptree. The purpose of the tutorial was to reduce the difficulty of learning how to answer questions about the Gender Browser's main icicle plot and detail displays. For example, after completing the tutorial, users would already have an idea of where to look to find the percentage of first female authors in a given discipline. We did not explain or even show the hoptree or breadcrumb trail during the tutorial, in order to avoid unfairly biasing participants towards using these features. Each trial was preceded by an easy practice question so that participants could get used to the task format and the new interface.

We handed participants cards with the questions printed on them one at a time. Participants were asked to read and understand the question on the card, and allowed to ask clarification questions. Once the question was understood, we began a timer and the participant attempted to use the Gender Browser to answer the question. When the participant had an answer ready, we stopped the timer and recorded the time taken to answer question, the number of clicks in different parts of the interface (recorded by the prototype itself), the correctness of the answer, and the participant's confidence in their answer on a 1-to-10 scale. In between each question, the Gender Browser was *not* reset, so that on each question participants could benefit from the history they had already accumulated in the breadcrumb and hoptree widgets.

After finishing each of the three trials, we asked participants to rate their own level of success, the difficulty of the questions, how much work was required to answer the questions, and level of frustration, all on 1-to-10 scales. Once all three trials were complete, we asked participants to choose from among the three different versions of the Gender Browser the one they preferred, which one was easiest to use, which one was the most frustrating, and which one allowed them to complete the tasks fastest. We asked follow-up questions to understand the reasons behind these answers.

5.3 Tasks

In selecting tasks for the experiment, we emulated the activities users might normally engage in while using the Gender Browser. Based on our own usage of the tool and discussions with colleagues, we determined that a plausible usage pattern could include looking up a few specific disciplines of personal interest and comparing the gender patterns among them. For example, a researcher using the tool might first locate her own specialty within the tree. She might then compare the gender frequency in her specialty with some other fields she is familiar with. We based the experiment tasks on this style of exploration, focusing on comparisons among multiple fields.

Using this approach, we developed a set of tasks, including revisiting [7, 22] and comparison tasks [23]. These task types are based on plausible user behavior and are supported by previous literature, a naturalistic evaluation would be required to determine if they are actually the tasks that users would normally perform using the tool.

Because every participant used each version of the Gender Browser (Plain, Breadcrumb, and Hoptree), we needed to develop three sets of questions. We first created one set of questions that we used as a template for the other two. Each question set followed the same structure, but was scoped within a different top-level branch of the Gender Browser's tree, reducing the likelihood of direct interference during the study. The template question set focused on the *Ecology and Evolution* subtree, and began by asking the user to answer a few easier retrieval questions such as "*What was the percentage of women authors in **Small mammal ecology**?*" Questions grew progressively more difficult, involving more comparison between ever more distant tree locations. For example, the last question in this set was "*How much larger is the number of women who published papers in **Migratory birds** (within **Avian reproductive ecology**) than on **Ungulates** (within **Mammalian herbivore ecology**)?*" As illustrated in this example, for fields that would be difficult to locate because of their depth in the tree and unfamiliarity (e.g. *Ungulates*), we included a reference to the usually easier-to-find parent node (e.g. *Mammalian herbivore ecology* in the above example).

After creating the first question set as a template, we analyzed the trajectory through the tree that users would need to follow as they progressed from question to question. For the second and third question sets, we chose two different top-level branches of the Gender Browser's tree (*Molecular and cell biology* and *Education*) where we were able to design questions that would reproduce this trajectory and maintain a similar difficulty level. Because it was impossible to create exactly equivalent question sets, pairings and orderings of question sets and interfaces were changed for each participant.

5.4 Participants

We recruited eighteen people from engineering, biology, and design departments to participate in our experiment. Ages ranged from 19 to 54 (mean of 30), and there were 8 women and 10 men. Six of the participants were undergraduates, while the rest had at least some graduate education.

Because two of the question sets focused on specific academic topics, we asked participants if they had ever studied the three subject areas at a college level or higher: Ecology and evolution, Molecular and cell biology, and Education. Eight of the participants had studied at least one of the biology topics, and six had studied Education. Twelve participants said that they had been authors on academic publications. We did not detect any important differences between these groups in the experiment.

6 Results

We summarize the results of our experiment for each of the measures we collected: user preferences, task time, mouse clicks, correctness, and confidence.

6.1 Preferences

After each trial, participants were asked to report how successful they felt they had been, how demanding the questions were, how much work the trial required, and how frustrating the trial had been. These answers were provided on 10 point scales. Fig. 4 shows the mean scores for each of the three interfaces.

We interpreted the ratings are ordinal data and used non-parametric Kruskal-Wallis tests to check for overall differences in preferences between the three interfaces. We found that interface differences had a significant overall effect on the *amount of work* participants felt they had to do to answer the questions ($\chi^2(2) = 5.3$, $p = 0.02$). Interface differences also had a significant effect on how *successful* participants felt ($\chi^2(2) = 4.76$, $p = 0.028$). There was also a trend in the level of *frustration* ($\chi^2(2) = 3.4$, $p = 0.063$). For these measures, we detected no significant pairwise differences, using Mann-Whitney tests with a Bonferroni correction ($\alpha = 0.0166$).

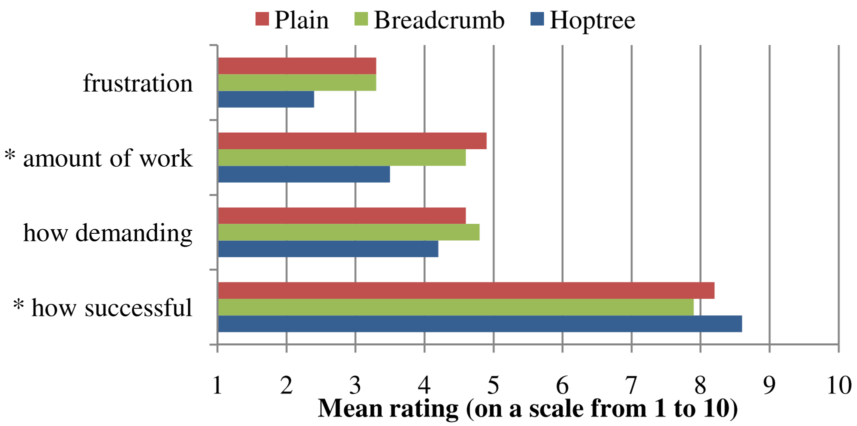


Fig. 4. Mean ratings in four categories, for each three interface. * indicates significance.

In the post-experiment questionnaire where participants were asked to choose the most preferred, easiest to use, fastest, and most frustrating version of the Gender Browser, participants expressed strong and consistent preferences for the Hoptree version. Table 1 summarizes these results. Chi-square tests indicated that the distribution of participants' answers over the three interfaces was significantly different from uniform on all four questions ($p < 0.001$).

Table 1. Number of users who selected each of the three interfaces as most preferred, easiest, fastest, and most frustrating. * indicates significant differences between interfaces.

	Plain	Breadcrumb	Hoptree
* Preferred	0	2	16
* Easiest	0	2	16
* Fastest	1	1	16
* Most Frustrating	13	5	0

These results support our hypothesis that users prefer the Hoptree or the Breadcrumb over the Plain interface. Most users also preferred the Hoptree over the Breadcrumb despite any preexisting familiarity with breadcrumb trail navigation.

6.2 Navigation Clicks

For each participant, we recorded the number of clicks performed in each condition. We separately recorded the number of clicks on the Gender Browser's main icicle plot, the Gender Browser's back button, and the navigational widget (breadcrumb or hoptree, not present in the Plain condition). Overall, participants had the fewest clicks using the Hoptree interface and the most with the Plain interface. Fig. 5 shows mean clicks total and by interface component: icicle plot, back bar, or navigation widget.

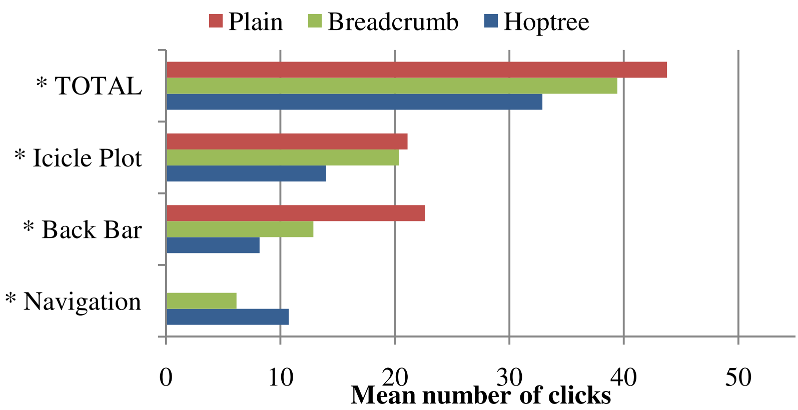


Fig. 5. Mean clicks on each interface. Total clicks, broken into clicks on the icicle plot, back bar, or navigation (hoptree or breadcrumb). * indicates significance.

Because the click count data was not normally distributed, we again used non-parametric Kruskal-Wallis tests to check for overall differences. The difference in total number of clicks (not separated by area) was statistically significant ($\chi^2(2) = 15.8$, $p < 0.001$). There were also significant differences between the three interfaces in the number of clicks on the icicle plot ($\chi^2(2) = 3.4$, $p = 0.063$), back button ($\chi^2(2) = 21.8$, $p < 0.001$), and navigation ($\chi^2(2) = 23.6$, $p < 0.001$).

We compared the click counts between each pair of interfaces using Mann-Whitney tests with a Bonferroni correction ($\alpha = 0.0166$). Between the Hoptree and Breadcrumb interfaces, the difference in total click count was significant ($p < 0.011$) and the difference in icicle plot clicks was significant ($p < 0.001$). Between the Breadcrumb and Plain interfaces, the number of back button clicks was significantly different ($p < 0.001$). Between the Hoptree and Plain interfaces, all differences in click counts were significant ($p < 0.001$).

Some participants barely used the navigational support widgets. Three out of the eighteen participants did not seem to realize that they could use the hoptree widget to navigate during the trial, and did not click on it at all. Three other participants clicked on it only once or twice, at the very end of the trial. All others clicked on the hoptree more than 10 times over the course of the 8 questions. For the Breadcrumb version of the interface, eight participants clicked on the breadcrumb zero or one times, and five of these participants were the same individuals who did not make use of the hoptree. Because the pre-trial tutorial did not include any description of the navigation widgets, these participants may have simply ignored it, focusing their attention only on working with the icicle plot portion of the Gender Browser.

6.3 Task Time

Participants took an average of 327 seconds (± 81) to complete the entire trial (all eight questions) using the Plain interface, 325 seconds (± 71) using the Breadcrumb interface, and 302 seconds (± 91) with the Hoptree. The time taken by individual participants varied widely, contributing to the high standard deviations for these data. Because the questions varied in difficulty, there was also a great deal of variation in time taken for different questions within the trial.

Before analyzing task time, we checked whether the data satisfied the assumptions required for parametric tests. A Shapiro-Wilk test indicated that the *TimeTaken* was not sufficiently normally distributed ($p < 0.001$). We transformed the timing data using a natural log transformation, which produced distributions that were acceptably normal for all three interfaces.

We then analyzed *Log(TimeTaken)* using a mixed-effects model analysis of variance. Like traditional repeated measures ANOVA, mixed model analyses can be used for factorial designs with between- and within-subjects factors. However, this technique is robust with missing data and imbalanced designs, and it models the experimental subject as a random effect because its levels are drawn randomly from a population. These tests retain larger denominator degrees of freedom than traditional ANOVAs, but detecting statistical significance is no easier because wider confidence intervals are used [24, 25]. The fixed-effects we used included *Interface* (Plain, Breadcrumb, Hoptree), *QuestionSet* (1–3), *Question* (1–8), *Trial* (1–3), and several interaction effects; *Participant* was modeled as a random effect.

There was a significant difference in *Log(TimeTaken)* for the three interface variations ($F(2, 283.8)=4.73$, $p < 0.01$). Pairwise comparisons between the three interfaces with a Bonferroni correction for multiple comparisons ($\alpha = 0.0166$) indicated that the difference between the Hoptree interface and the Plain interface was significant ($p < 0.005$). With the significance correction, there was only a marginal difference between the Hoptree and Breadcrumb interfaces ($p < 0.024$). The difference between the Breadcrumb and Plain interfaces was also not significant.

As mentioned previously, some participants did not use the hoptree or breadcrumb interfaces during the experiment. In a post-hoc analysis, we partitioned participants into *users* (13 participants) who clicked on either the breadcrumb or hoptree more than once, and *non-users* (5 participants) who clicked on neither interface more than once. Fig. 6 displays the average time to complete all eight questions with each of the three interfaces, for both *users* and *non-users*. A mixed-effects model analysis of variance of total time taken found a significant difference between interfaces within the group who used the navigation features ($F(2, 235.9)=4.4$, $p < 0.013$), and pairwise comparisons found that the Hoptree interface performed significantly better than both the Breadcrumb ($p < 0.015$) and Plain interfaces ($p < 0.008$). There were no significant differences detected within the non-users group.

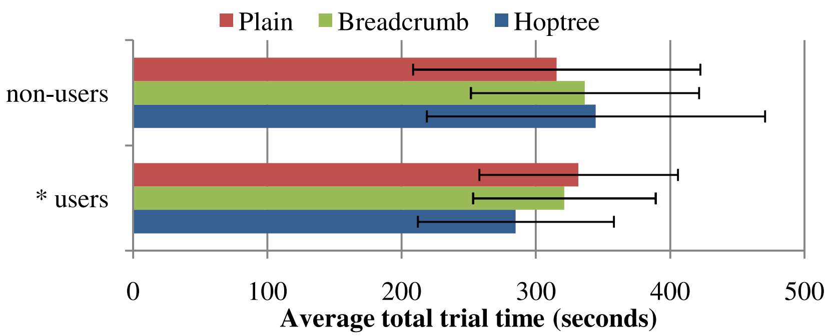


Fig. 6. Average total times by interface, for participants who used navigation features (users) and those who did not (non-users). * indicates significance.

These results support the hypothesis that users could complete the tasks fastest with the hoptree, but suggest that increased awareness of the hoptree widget, through training, design changes, or longer usage time, may be necessary before these benefits can be fully realized.

6.4 Correctness and Confidence

Out of 432 total questions that were asked, only 22 incorrect answers were given. The Breadcrumb interface had 11 incorrect answers, followed by the Plain interface with 7 and the Hoptree with 4. A chi-square test did not find this distribution to be significantly different from uniform. The median confidence score given by participants was 9 out of 10, for all three interface conditions.

These results do not support our hypotheses that there would be fewer incorrect answers and higher confidence on the Hoptree and to some extent the Breadcrumb interfaces. The trend we observed toward a lower error rate on the Hoptree interface bears further study.

6.5 Task Characteristics and Learning Effects

We analyzed the differences in the amount of time taken between different questions and between different question sets to verify whether or not we had succeeded in designing questions with varying levels of difficulty and question sets with comparable levels of difficulty.

The time taken on the different question sets varied slightly. On Set 1, participants took an average of 37 seconds (± 22) to answer each of the questions; for Set 2, the average time was 43 seconds (± 29); and for Set 3, the average total was 39 seconds (± 25). However, the mixed-effects model analysis of variance, discussed previously, did not find a significant main effect of Question Set on *Log(TimeTaken)*.

Within the question sets, the time taken on each question increased gradually starting with a mean of 13 seconds (± 5) on Question 1 and ending with a mean of 56 seconds (± 25) on Question 8. Overall differences in *Log(TimeTaken)* among the eight questions were significant ($F(7,84.7) = 124.8$, $p < 0.001$). These results indicate that the question sets were fairly similar in difficulty, but that the questions within the sets varied in difficulty, as intended.

We also checked for differences between the 1^{st}, 2^{nd}, and 3^{rd} trials that participants completed, to determine the degree to which general familiarity with the interface or process improved task time. Participants did seem to get faster as they became accustomed to the types of questions being asked and the Gender Browser interface. Questions completed in Trial 1 had a mean time of 43 seconds (± 27), in Trial 2 a time of 38 seconds (± 25), and in Trial 3 a time of 37 seconds (± 24). However, the mixed-effects model analysis of variance found only a trend in *Log(TimeTaken)* between the three trials ($p = 0.067$).

7 Discussion

The results indicate that hoptrees allow users to navigate the hierarchy more quickly and in fewer clicks. Users also preferred the hoptree over the other versions tested. We observed several interesting patterns of use that suggest future lines of research.

As noted above, certain participants barely used the navigational widgets during the experiment. Because the tutorial prepared them for only the plain version of the interface, the participants may have simply ignored the addition of the relatively subtle navigation widgets, perceiving it as outside the scope of the task, as it had been explained. It may be possible to improve the design of both the hoptree and breadcrumbs so that users more quickly recognize the affordances and utility of the navigation widgets. Prior research on breadcrumbs found that, in a website navigation task, only a small percentage of clicks were on the breadcrumb trail [9] while another study found that training in how to use the breadcrumb resulted in greatly improved performance [10]. We would like to investigate the effects of design changes, training, or increased familiarity. Slight changes to the design of the hoptree, such as beveled edges or link-like underlining, may improve the discoverability of the tool. We predict that this would result in increased usage, leading to significant efficiency improvements.

Some of the participants who *did* use the hoptree and breadcrumb trail seemed to instantly and naturally understand how it worked. For others, there was a detectable "aha moment" where they realized that they could use the hoptree to complete the tasks. For several participants who had already completed trials with the more tedious plain or breadcrumb interfaces, the moment when the hoptree first branched to show the history of their previous explorations was accompanied by smiles or appreciative exclamations such as "I like this thing right here" (pointing to the hoptree). When one participant began using the Breadcrumb interface after completing the Hoptree trial, he remarked "It's amazing how much that thing helped." The post-experiment ranking results indicate that even for those participants who did not realize how the hoptree worked until it was too late to take advantage of it still appreciated its advantages.

Although our results suggest that the hoptree led to a significant decrease in time taken overall, we observed some cases where participants actually seemed to take longer with the hoptree than they would have with the breadcrumb or plain interfaces. This phenomenon sometimes occurred during the questions which required comparison between multiple tree locations. Specifically, with the plain or breadcrumb interfaces, participants answering a comparison question would often follow the following procedure: (1) visit the first location in the tree; (2) find and memorize the piece of information the question asked about; (3) find the second location; (4) find the required piece of information at that location; (5) perform the required comparison; (6) report the answer.

On the other hand, when using the hoptree, there was a fast and easy way for participants to hop back and forth between locations. Many participants took advantage of this, and did not memorize the information as they would have with the plain or breadcrumb interfaces. Instead they began by visiting all of the locations the question asked about, sometimes without even looking for the information at each location. Visiting the locations makes each of the locations available for revisiting through the hoptree, so participants would then quickly revisit each location through the hoptree widget to retrieve the required information. They sometimes visited all of the locations more than once to double check their answers.

The popularity of this unexpected "measure twice" strategy, enabled by the hoptree, may have reduced the size of the speed improvement we observed. At the same time, the ability to quickly check comparisons, instead of having to commit multiple pieces of data to memory, is itself an important advantage. We believe that the ease with which users could check their work with the hoptree contributes to the higher preference scores in our experiment.

Given this phenomenon, we would have expected to see higher rates of correct answers and higher confidence ratings on the hoptree interface than on the breadcrumb and plain interfaces. However, there were very few incorrect responses overall and differences were not significant. From the data we collected, confidence levels also did not seem to be affected in any noticeable way by the different interfaces. It is possible that in a larger experiment, or with more difficult questions, differences would become apparent. The questions in our experiment all had right-or-wrong answers, and most participants worked until they were sure they had the correct answer. As a result, most participants seemed uncertain how to rate their confidence, and in most cases they chose the same confidence level for nearly all of the questions.

Several participants who used the hoptree widget more extensively commented on some aspects of the tool. Two participants mentioned that they were annoyed by the animations between tree locations that took place in the Gender Browser's icicle plot. While not a direct feature of the hoptree, it seems that once the hoptree created the potential for instant traversal between locations, the Gender Browser's relatively slow animated transitions became annoying.

Some participants also commented on the hoptree's pruning strategy. While several users said that the pruning strategy seemed appropriate and useful, a few people wanted either greater control over what the hoptree chose to preserve, or a more extensive history. The strategy we selected is designed to minimize the amount of maintenance work for users, at the risk of occasionally pruning nodes that the user wants to revisit. Depending on the application domain and the types of browsing activities that users are engaging in with the hierarchy, it might be preferable to allow a greater level of control. For example, allowing users to "pin" certain branches to the hoptree, preventing them from being pruned, could be a useful optional feature.

8 Conclusions and Future Work

We have introduced the hoptree, a novel visual interface for more quickly and easily navigating hierarchies, such as tree visualizations, web sites, and file systems. We have explained the design of the hoptree and our prototype implementation, which we have published as an open source jQuery plugin[4]. We compared the hoptree to a breadcrumb navigation widget within the context of the Gender Browser. Our results demonstrate that the hoptree has significant speed and efficiency advantages, and that users prefer the hoptree to breadcrumb trails for tasks involving comparisons.

For this lab experiment we created tasks involving targeted information seeking and comparisons, but users interacting with a large hierarchy naturally might engage in more open-ended, exploratory activities. The path a user would naturally take through a hierarchy such as the Gender Browser would focus on personally meaningful information, while the information we asked participants to find may have felt arbitrary. Future work should investigate the impact of hoptrees and navigation tools in open-ended scenarios. Do users explore the tree more deeply when the hoptree is present? Do they explore a larger number of disciplines? Do they spend more time exploring the tree?

---

[4] `http://github.com/michaelbrooks/hoptree`

Future research should also study hoptrees in greater detail to extract more general principles that could guide the design of hierarchy navigation tools in the future. For example, an eye-tracking study could more precisely investigate the effects of hoptrees and other navigation tools on cognitive load and lostness [26] during hierarchy exploration. Comparison of history tracking tools that are structured according to the information hierarchy, like hoptrees and breadcrumbs, against designs that are structured by the users' visit path as in [17], may also yield new insight.

Acknowledgments. We thank our participants for their time and patience. We also thank Nicolas Belmonte, the creator of the JavaScript InfoVis Toolkit, for his invaluable assistance developing the hoptree prototype and the Gender Browser. This work was supported in part by NSF grant SBE-0915005 and by a generous gift from JSTOR.

References

1. Schulz, H.-J.: Treevis. net: A Tree Visualization Reference. IEEE CG & A 31, 11–15 (2011)
2. Bartram, L., Ho, A., Dill, J., Henigman, F.: The continuous zoom: A constrained fisheye technique for viewing and navigating large information spaces. In: Proc. UIST 1995, pp. 207–215. ACM (1995)
3. Blanch, R., Lecolinet, E.: Browsing zoomable treemaps: structure-aware multi-scale navigation techniques. IEEE TVCG 13, 1248–1253 (2007)
4. Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., Roseman, M.: Navigating hierarchically clustered networks through fisheye and full-zoom methods. ACM ToCHI 3, 162–188 (1996)
5. Lamping, J., Rao, R., Pirolli, P.: A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In: Proc. CHI 1995, pp. 401–408. ACM (1995)
6. Heer, J., Card, S.K.: DOITrees revisited: scalable, space-constrained visualization of hierarchical data. In: Proc. AVI 2004, pp. 421–424. ACM (2004)
7. Plaisant, C., Grosjean, J., Bederson, B.B.: SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation. In: Proc. INFOVIS 2002, pp. 57–64. IEEE (2002)
8. Instone, K.: Location, path and attribute breadcrumbs. 2002 Information Architecture Summit, pp. 16–17 (2002)
9. Rogers, B.L., Chaparro, B.: Breadcrumb navigation: Further investigation of usage. Usability News 5 (2003)
10. Hull, S.: Influence of training and exposure on the usage of breadcrumb navigation. Usability News 6 (2004)
11. Teng, H.: Location breadcrumbs for navigation: An exploratory study (2005)
12. Blustein, J., Ahmed, I., Instone, K.: An evaluation of look-ahead breadcrumbs for the WWW. In: Proc. HYPERTEXT 2005, pp. 202–204. ACM, New York (2005)
13. Heer, J., Mackinlay, J., Stolte, C., Agrawala, M.: Graphical histories for visualization: Supporting analysis, communication, and evaluation. IEEE TVCG 14, 1189–1196 (2008)
14. Shrinivasan, Y.B., Van Wijk, J.J.: Supporting the analytical reasoning process in information visualization. In: Proc. CHI 2008, pp. 1237–1246. ACM (2008)

15. Kreuseler, M., Nocke, T., Schumann, H.: A History Mechanism for Visual Data Mining. In: Proc. INFOVIS 2004, pp. 49–56. IEEE (2004)
16. Ayers, E.Z., Stasko, J.T.: Using graphic history in browsing the World Wide Web. In: Uchikawa, Y., Furuhashi, T. (eds.) WWW 1995. LNCS, vol. 1152, pp. 1–8. Springer, Heidelberg (1996)
17. Hightower, R.R., Ring, L.T., Helfman, J.I., Bederson, B.B., Hollan, J.D.: Graphical multiscale Web histories: a study of padprints. In: Proc. HYPERTEXT 1998. ACM (1998)
18. Cockburn, A., Greenberg, S.: Issues of page representation and organisation in web browser's revisitation tools. In: Proc. OZCHI 1999 (1999)
19. Rosvall, M., Bergstrom, C.T.: Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS ONE 6, e18209 (2011)
20. Kaye, J.: "Jofish": Some statistical analyses of CHI. In: Proc. CHI EA 2009. ACM, New York (2009)
21. Belmonte, N.G.: The JavaScript InfoVis Toolkit, `http://thejit.org/`
22. Song, H., Kim, B., Lee, B., Seo, J.: A comparative evaluation on tree visualization methods for hierarchical structures with large fan-outs. In: Proc. CHI 2010, p. 223. ACM Press, New York (2010)
23. Stasko, J., Catrambone, R., Guzdial, M., Mcdonald, K.: An evaluation of space-filling information visualizations for depicting hierarchical structures. International Journal of Human-Computer Studies 53, 663–694 (2000)
24. Wolfinger, R., Tobias, R., Sall, J.: Mixed models: a future direction. In: Proc. SAS UGI 1991, pp. 1380–1388 (1991)
25. Schuster, C., Von Eye, A.: The Relationship of ANOVA Models with Random Effects and Repeated Measurement Designs. Journal of Adolescent Research 16, 205–220 (2001)
26. Otter, M., Johnson, H.: Lost in hyperspace: metrics and mental models. Interacting with Computers 13, 1–40 (2000)

User-Centric vs. System-Centric Evaluation of Recommender Systems

Paolo Cremonesi[1], Franca Garzotto[1], and Roberto Turrin[2]

[1] Politecnico di Milano, Milano, Italy
[2] ContentWise, Milano, Italy
{paolo.cremonesi,franca.garzotto}@polimi.it,
roberto.turrin@contentwise.tv

Abstract. Recommender Systems (RSs) aim at helping users search large amounts of contents and identify more effectively the items (products or services) that are likely to be more useful or attractive. The quality of a RS can be defined from two perspectives: *system-centric,* in which quality measures (e.g., precision, recall) are evaluated using vast datasets of preferences and opinions on items *previously* collected from users that are *not* interacting with the RS under study; *user-centric*, in which user measures are collected from users interacting with the RS under study. Prior research in e-commerce has provided some empirical evidence that system-centric and user-centric quality methods may lead to inconsistent results, e.g., RSs that were "best" according to system-centric measures were not the top ones according to user-centric measures. The paper investigates if a similar mismatch also exists in the domain of *e-tourism*. We discuss two studies that have adopted a system-centric approach using data from *210000 users,* and a user-centric approach involving 240 users interacting with an *online hotel booking* service. In both studies, we considered four RSs that employ an implicit user preference elicitation technique and different baseline and state-of-the-art recommendation algorithms. In these four experimental conditions, we compared system-centric quality measures against user-centric evaluation results. System-centric quality measures were *consistent* with user-centric measures, in contrast with past studies in e-commerce. This pinpoints that the relationship between the two kinds of metrics may depend on the business sector, is more complex that we may expect, and is a challenging issues that deserves further research.

Keywords: Recommender systems, E-tourism, Evaluation, Decision Making.

1 Introduction

Recommender Systems (RSs) aim at helping users search large amounts of digital contents and identify more effectively the items that are likely to be more useful or attractive. For consumers overwhelmed by excessively wide offer of products or services, recommendations reduce information overload, facilitate the discovery of what they need or are interested to, help them to make choices among a vast set of

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 334–351, 2013.

alternatives, and potentially improve their decision making process. From a provider's perspective, RSs are regarded as a means to improve users' satisfaction and ultimately increase business.

Most recommender systems operate by predicting the opinion (i.e., the numerical *rating*) that a user would give to an item (such as a movie, or a hotel), using a statistical model built from the characteristics of the item (content-based approaches) or the opinions of a community of users (collaborative-based approaches).

Some research has explored the effectiveness of RSs as decision support tools in the e-tourism domain, and has investigated how they influence users' decision making processes and outcomes [7,24,31,33,13,28]. Empirical evidence suggests that RSs improve user's decision making and their influence depends on a variety of factors which are related to the quality of the recommender system.

RS quality can be defined either in terms of system-oriented metrics, which are evaluated algorithmically (e.g., precision, recall), or with user-centric experiments. [8,12,28].

– In *user-centric* evaluation, users interact with a running recommender system and receive recommendations. Measures are collected by asking the user (e.g., through interviews or surveys), observing her behavior during use, or automatically recording interactions and then subjecting system logs to various analyses (e.g., click through, conversion rate).
– With *system-centric* methods, the recommender system is evaluated against a *prebuilt* ground truth dataset of opinions. Users do not interact with the system under test but the evaluation, in terms of *accuracy*, is based on the comparison between the opinion of users on items as estimated by the recommender system and the judgments *previously* collected from real users on the same items.

Although the user-centric approach is the only one able to truly measure the user's satisfaction on recommendations and the quality of the decision making process, conducting empirical tests involving real users is difficult, expensive, and resource demanding. On the contrary, system-centric evaluation has the advantage to be immediate, economical and easy to perform on several domains and with multiple algorithms.

Recently, many researchers have argued that the system-centric evaluation of RSs in e-commerce applications does not always correlate with how the users perceive the value of recommendations [2,5,6,19,22,27]. This may happen because system-centric evaluation cannot reliably measure non-accuracy metrics such as novelty – the extension to which recommendations are perceived as new – which more reflects the user and business dimensions. These works suggest that RS effectiveness in e-commerce applications should not be evaluated simply in terms of system-oriented accuracy but user-centric metrics should be adopted as well.

These contrasting results between system-centric and user-centric evaluation of RSs do not necessarily hold for e-tourism applications, because of the peculiar nature of the touristic product [11,25,26,30]:

- Touristic products lack the feature of "try-before-buy" or "return in case the quality is below expectance". Online tourist service purchasing involves a certain amount of risk taking.
- A priori comprehensive assessment of the quality of the touristic product is impossible: tourists must leave their daily environment to use it.
- The touristic product has to do with an overall emotional experience.
- In many circumstances, novelty is a weak quality attribute of touristic products. Tourists can "reuse" and buy the same product again and again if they consider the experience emotionally satisfying.

Because of these differences that might impact on users' decision making, the online selling of touristic services cannot be considered as a special case of e-commerce, and the quality characteristic of this process might differ significantly in the two domains.

This paper explores the influence of recommendations on decision making in the wide application arena of online tourism services, specifically considering hotel booking. Our research is grounded on a specific case study – the online reservations service provided by Venere.com, a subsidiary company of the Expedia group, one of the worldwide leaders in the hotel booking market, featuring more than 120,000 hotels, bed and breakfasts and vacation rentals in 30,000 destinations worldwide. Our joint work with Expedia addresses the following research question:

> *Do the algorithms which perform best in terms of system-centric quality generate recommendations that provide the best effects on decision making?*

We focus our research on the effects of recommendations on decision making in relationship to a specific design factor – the recommendation *algorithm* used. We aim at exploring the *differences* between users who use an online booking system *without recommendations* and those who use the same booking system *extended with personalized recommendations* generated by *different algorithms*.

Our research investigates the effects of recommender algorithms from both a user-centric ("subjective") point of view and a system-centric ("objective") perspective. To explore our general research question and to evaluate if system-oriented metrics are able to correctly capture the quality of the decision making process from a user perspective, we carried on two wide and articulated empirical studies:

1. a *system-centric evaluation* to measure the objective quality in terms of *accuracy* (recall and fallout); this involved 210,000 simulated users, characterized by absence or presence of personalized recommendations, the latter being generated by three different algorithms (collaborative, content-centric, and hybrid);
2. a set of *user-centric experiments* involving 240 users and measuring different decision making attributes in *four experimental conditions,* characterized by the same four recommenders adopted in system centric evaluation.

The comparison of the evaluation outcomes shows that system-centric and user-centric metrics lead to consistent results, in contrast with past studies in e-commerce [5,6], and suggests that in the online hotel booking domain system-centric accuracy measures are good predictors of the beneficial effect of personalized recommendations on user's decision making. Our findings pinpoints that the relationship between the system-centric and user-centric metrics may depend on the business sector, is more complex that we may expect, and is a challenging issues that deserves further research.

2 Related Work

2.1 Recommender Systems in e-Tourism

The potential benefits of RSs in e-tourism have motivated some domain-specific researches. Ricci et al. in [24] present NutKing, an online system that helps the user to construct a travel plan by recommending attractive travel products or by proposing complete itineraries. The system collects information about personal and travel characteristics and provides hybrid recommendations. NutKing searches for user-centric similar items and later ranks them based on a content-centric similarity between items and user's requirements. Levi et al. in [18] describes a recommender system for online hotel booking. The system adopts a recommendation technique symmetric to technique described in [24] and adopts sentiment-analysis to estimate user's rating from their reviews. Zanker et al. [33] present an interactive travel assistant, designed for an Austrian spa-resort, where preference and requirement elicitation is explicitly performed using a sequence of question/answer forms. Delgado et al. in [7] describe the application of a collaborative attribute-centric recommender system to the Ski-Europe.com web site, specialized in winter ski vacations. Recommendations are produce by taking into account both implicit and explicit user feedbacks. Implicit feedback is inferred whenever a user prints, bookmarks, or purchases an item (positive feedback) or does nothing after viewing an item (negative feedback).

2.2 Evaluation of Recommender Systems

Several studies have investigated how to measure the effectiveness of recommenders. A systematic review of *system-centric* evaluation techniques is reported by Herlocker et al. in [12]. More recently, some researchers [3,19,20,22,27] have argued that RS effectiveness should not be evaluated simply in terms of system-centric metrics and have investigated *user-centric* evaluation methods, which focus on the human/computer interaction process (or User eXperience, UX) [18,22,29].

Swearingen and Sinha [27] were among the first studies to point out that subjective quality of a RS depends on factors that go beyond the quality of the algorithm itself. Without diminishing the importance of the recommendation algorithm, these authors claim that RS effectiveness should not be evaluated simply in terms of system-centric accuracy metrics. Other design aspects, ignored by these metrics, should be measured, and in particular those related to the acceptance of the recommender system and of its recommendations.

Along the same vein, other researchers have investigated the so called user-centric methods, which focus on how user characteristics are elicited and recommended items are presented, compared, or explained. They explore "subjective" quality of RSs and attempt to correlate it to different UX factors. They highlight that, from a user's perspective, an effective recommender system should inspire credibility and trust towards the system [22] and it should point users towards new, not-yet-experienced items [18].

Due to the intrinsic difficulty of performing user studies in the RS domain, empirical results in this field are tentative and preliminary. Celma and Herrera [4] report an experiment that studied how users judged novel recommendations provided by a CF and a CBF algorithm in the music recommendation context. Ziegler et al. [34] and Zhang et al. [32] propose diversity as a quality attribute: recommender algorithms should seek to provide optimal coverage of the entire range of user's interests. This work is an example of a combined use of automatic and user-centric quality assessment techniques. Pu et al. [22] developed a framework called ResQue, which defines a wide set of user-centric quality metrics to evaluate the perceived qualities of RSs and to predict users' behavioral intentions as a result of these qualities.

Table 1. Experimental conditions used in the two studies

Study	Type	Independent variables (algorithms)	Dependent variables			Users
1	System-centric simulation	HotelAvg PureSVD DirectContent Interleave	Accuracy (recall and fallout)			210,000 (simulated)
2	User-centric experiment		Choice satisfaction	Satisfaction	subjective	240 (total)
			Choice risk	Trust		
			Perceived time	Effort		
			Elapsed time		objective	
			Extent of hotel search			
			Menu interactions	Efficacy		

3 The Design of the Studies

The research question presented in the Introduction has been explored with *two studies* – a system-centric simulation and a user-centric experiment – summarized in Table 1. In both the studies, the effects of recommendations have been explored under 4 different experimental conditions defined by one manipulated variable: the *recommendation algorithm*. Our study considers *one* non-personalized algorithm and *three* personalized RSs representatives of three different classes of algorithms: *collaborative*, *content* and *hybrid*.

– **HotelAvg** is a non-personalized algorithm and presents hotels in decreasing order of *average user rating* [15]. This is the default ranking option adopted in our study when the user does not receive personalized recommendations. The same ranking strategy is adopted by most online hotel booking systems such as TripAdvisor, Expedia, and Venere.

- **PureSVD** is a collaborative algorithm based on matrix-factorization; previous research shows that its accuracy is one of the best in the movie domain [6].
- **DirectContent** recommends hotels whose content is similar to the content of hotels the user has rated [18]. Content analysis takes into account the 481 features (e.g., category, price-range, facilities), the free text of the hotel description, and the free text of the hotel reviews. *DirectContent* is a simplified version of the LSA algorithm described in [1].
- **Interleave** is a hybrid algorithm that generates a list of recommended hotels alternating the results from *PureSVD* and *DirectContent*. Interleave has been proposed in [3] with the name "mixed hybridization" and, although trivial in its formulation, has been shown to improve diversity of recommendations.

3.1 Study 1: System-Centric Evaluation

The first study analyzes the accuracy of recommendations as a function of the recommender algorithm. For the evaluation, Venere.com made us available a catalog of more than 3,000 hotels and 72,000 related users' reviews. Each accommodation is provided with a set of 481 features concerning, among the others: accommodation type (e.g., residence, hotel, hostel, B&B) and service level (number of stars), location (country, region, city, and city area), booking methods, average single-room price, amenities (e.g., spa), and added values (e.g., in-room dining). User's reviews associated to each accommodation consist of numeric ratings and free-text.

We have enriched the dataset with additional reviews extracted from the TripAdvisor.com web site using a web crawling tool. Table 2 reports the detailed statistics of the dataset used in our experiments.[1]

Table 2. Dataset used in the two studies

	Total (Venere+TripAdvisor)	**Venere**	**TripAdvisor (crawled)**
Hotels	3,100	3,100	–
Users (reviewers)	210,000	72,000	138,000
Reviews and ratings	246,000	81,000	165,000
Hotel features	481	481	–

Dependent Variables

System-centric quality can be measured by using either accuracy metrics (e.g., precision, recall and fallout) or error metrics (e.g., RMSE and MAE). The hybrid algorithm tested in this study cannot be evaluated with error metrics since it does not compute actual ratings [15]. Hence we have considered only accuracy metrics. In particular we

[1] The dataset is available by contacting the authors.

focused our attention on *recall* (the conditional probability of suggesting a hotel given it is relevant for the user) and *fallout* (the conditional probability of suggesting a hotel given it is irrelevant for the user) as the *dependent* variables. A good algorithm should have large recall (i.e., it should be able to recommend hotels of interest to the user) and low fall-out (i.e., it should avoid to recommend hotels of no interest to the user).

The methodology adopted to measure the two variables is the same used in many works on recommender system – e.g., [6,16]. Ratings in the dataset were randomly split into two subsets: *training* set (90% of the ratings) and *test* set (10% of the ratings). In order to measure recall, we first trained the algorithms using the ratings in the training set. Then, for each user in the test set and for each hotel in the test set that was rated 10-stars or 9-stars by the user (we assumed these hotels to be relevant for the users) we followed these steps:

- We randomly selected 1,000 additional hotels that were not rated by the user. We assumed that the user was not interested in most of them.
- We predicted the ratings for the relevant hotel and for the additional 1,000 hotels.
- We formed a recommendation list by picking the *N* hotels with the largest predicted ratings (top-N recommendation).

For each recommendation we have a *hit* (e.g., a successful recommendation) if the relevant hotel is in the list. Therefore, the overall recall was computed by counting the number of hits (i.e., the number of successful recommendations) over the total number of recommendations

$$\text{recall}(N) = \frac{\text{number of times the relevant hotel is in the list}}{\text{number of recommendations}}$$

A similar approach was used to measure fallout, with the only difference being that we selected non-relevant hotels – defined as the hotels rated lower than 2 out of 10 stars. The fallout was computed as

$$\text{fallout}(N) = \frac{\text{number of times the non–relevant hotel is in the list}}{\text{number of recommendations}}$$

Recall and fallout range from 0% to 100%. An ideal algorithm should be able to recommend all of the interesting hotels (i.e., recall equals to 100%) and to discard all of the uninteresting hotels (i.e., fallout equals to 0%).

3.2 Study 2: User-Centric Evaluation

The second study analyzes opinions and behavior of 240 users interacting with an online hotel booking service.

Dependent Variables

We model the *effects* on decision making that can be associated to the introduction of personalized recommendations using some subjective attributes resulting from the user's perception and judgment of the decision activity (subjective variables) as well as some objectively measurable attributes of the decision processes (objective variables). The variables, listed in Table 1, have been defined according to the model described in [30]:

- **Choice satisfaction:** the subjective evaluation of the reserved hotel in terms of quality/value for the user;
- **Choice risk:** the user's perception of uncertainty and potentially adverse consequences of booking the chosen hotel, measured in terms of the perceived degree of mismatching between the characteristics of the chosen hotel emerging from the use of the system and the real characteristic of the accommodation;
- **Perceived time:** the user's judgment on the length of the decision making process;
- **Elapsed time:** the time taken for the user to search for hotel information and make a reservation decision;
- **Extent of hotel search:** the number of hotels that have been searched, for which detailed information has been acquired;
- **Number of sorting changes:** the number of times the user *changed the ordering* of hotels in the list view: the *default* ordering is (i) descending order of average user ratings, when there are no recommendations, and (ii) descending order of estimated user relevance, when there are recommendations. Ordering change is a measure of the *efficacy* of RSs in situations where the conversion rate, i.e., the percentage of recommended items that are actually purchased by users [33] cannot be assessed. This happens, for example, when a system does not present a *separate* list of recommended items, but recommendations are rendered by sorting items in descending order of relevance as estimated by the recommender algorithm, which represents the "de facto" recommendation list.

A *questionnaire* has been used to measure *choice satisfaction*, *choice risk* and *perceive time*. Measures of *effort* and *efficacy* (*execution time, extent of hotel search, number of sorting changes)* have been obtained by analyzing the interaction data collected from the system logs.

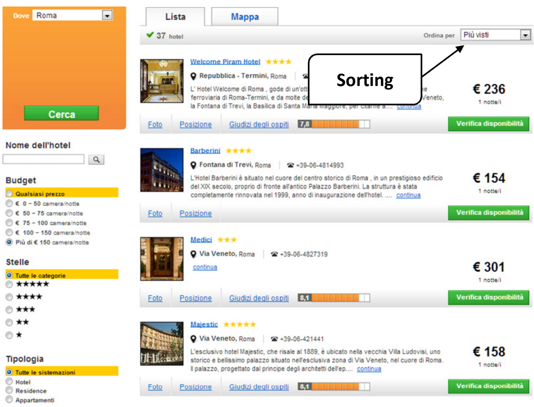


Fig. 1. Screenshot of the PoliVenus page showing a list of hotels

Instruments

For the purpose of our study, we have developed PoliVenus, a web-centric testing framework for the hotel booking field, which can be easily configured to facilitate the execution of controlled empirical studies in e-tourism services (Figure 1). PoliVenus implements the same layout as Venere.com online portal and simulates *all* its functionality with the exception of payment functions. PoliVenus contains the same catalog of hotels and reviews used in the first study. The Polivenus framework is based on a modular architecture and can be easily customized to different datasets and types of recommendation algorithms. Its current library of algorithms comprises 20 algorithms, developed in cooperation with ContentWise[2]. PoliVenus can operate in two different configurations:

- **Baseline configuration:** it corresponds to the existing Venere.com portal. Users can filter the hotel catalogue according to hotel characteristics (e.g., budget range, stars, accommodation type, city area), and retrieve non-personalized results. They can sort the list by popularity, service level, price, or user average rating (the default sorting option). The default sorting option corresponds to the HotelAvg non-personalized recommender algorithm.
- **Personalized configuration:** it is almost identical to the baseline configuration, the only difference being the personalized recommendations. When watching a list of hotels previously filtered according to accommodation characteristics, the user is offered an additional option to sort hotels on the basis of the personalized recommendations (this is the default sorting option).

The user profile required by the algorithms to generate recommendations is based on the user's current interaction with the system (*implicit elicitation*). This choice is motivated by three specific reasons:

(i) we want to support users who have no rating history or who are not interested in logging into the system;
(ii) we are interested in exploring a smooth integration of personalized recommendations in existing online booking systems; to enable explicit elicitation would require the introduction of an intrusive add-on;
(iii) according to a large number of works, the lower effort of implicit elicitation (as compared to explicit elicitation) is related to higher perceived effectiveness of recommendations [9,10,14,23].

The implicit elicitation mechanism adopted in PoliVenus is the following: whenever a user interacts with an object on the interface, the system assigns a score to the hotel related to that object (e.g. link, button, map, picure, etc.). With all these *signals*, PoliVenus builds the user profile for the current user session: the user profile contains implicit hotel ratings, where each rating is computed as the maximum of all the signals generated for that hotel. The user profile is continuously updated with every new signal and the list of recommended hotels is updated accordingly.

Following the definition of Ricci and Del Missier in [24] we considered short-term and long-term signals. In our implementation the short-term signal is the last

[2] www.contentwise.tv

interaction of the user with the system. All previous interactions with the system are considered long-term signals. In order to give more importance to the most recent interaction, we have employed an exponential decay function to the ratings. Whenever a new signal enters in the user profile, all previous ratings are divided by a dumping factor. The magnitude of the damping factor controls the decay rate. In our experiments we have used a damping factor of 2.

Participants and Procedure

Our main research audience is represented by users aged between 20 and 40 who have some familiarity with the use of the web and had never used Venere.com before the study (to control for the potentially confounding factor of biases or misconceptions derived from previous uses of the system). The total number of recruited subjects who completed the task and filled the questionnaire by the deadline was *240*. They were equally distributed in the four experimental conditions. We recruited participants from current students and ex-alumni from the School of Engineering and the School of Industrial Design of our university. They were contacted by e-mail, using university mailing lists. The invitation included the description of the activities to be performed and the reward for taking part in the study. Users were not aware of the true goal of the experiment.

To encourage participation, and to induce participants to play for real and to take decisions as they would actually do when planning a vacation, we used a lottery incentive [21]. Participants had the chance of winning a *prize*, offered to a randomly selected person who completed the assigned decision making task and filled the final questionnaire by a given deadline. The prize consisted of a coupon of the value of 150€ to be used to stay in the hotel fictitiously reserved using PoliVenus.

All participants were given the following instructions: "*Imagine that you are planning a vacation in Rome and are looking for an accommodation during Christmas season; choose a hotel and make a reservation; dates and accommodation characteristics (stars, room type, services, and location) are at your discretion. After confirming the reservation (simulated), please complete the final questionnaire*".

Table 3. Excerpt from the questionnaire

Question	Range	Dependent Variable
How much are you satisfied with your final choice?	1:not too much 5:very much	Choice Satisfaction
The time required to book the hotel is:	1:short 5:overmuch	Perceived Time
How much will the characteristics of the reserved hotel correspond to those of the real accommodation?	1:not much, 5:very much	Choice Risk

After accessing PoliVenus, reading the instructions, and agreeing on the study conditions (lottery participation and privacy rules), each participant was automatically moved to the homepage of the PoliVenus hotel reservation system and randomly assigned to one of the four experimental conditions. After committing the reservation, the user was directed to the questionnaire page containing 11 questions, a subset of which is reported in Table 3.

4 Results

4.1 Study 1: System-Centric Evaluation

Figure 2 presents a plot of recall versus fallout, where each point on the curve corresponds to a value of the number *N* of recommended hotels. A perfect recommendation algorithm will generate a curve that goes straight to the upper left corner of the figure, until 100% of relevant hotels and 0% of non-relevant hotels have been recommended.

The figure shows that the hybrid *Interleaved* algorithm is the most accurate. The accuracy of the content and collaborative algorithms (*DirectContent* and *pureSVD*) is not significantly better than the accuracy of the non-personalized baseline *HotelAvg*.

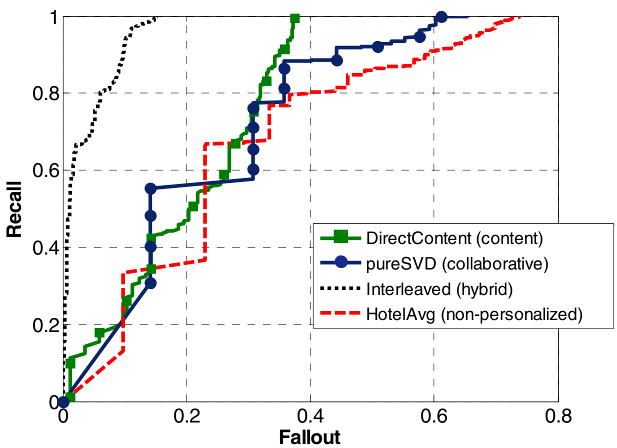


Fig. 2. System-centric evaluation: Recall vs.Fallout

4.2 Study 2: User-Centric Evaluation

We first polished the collected data by removing the ones referring to subjects who showed apparent evidences of gaming with the testing system (e.g., those who interacted with the system for less than 2 minutes) or left too many questions unanswered. In the end, we considered the data referring to *229* participants. They were almost equally distributed in the four experimental conditions, each one involving a number of subjects between 54 and 58.

Subjective Metrics

We used ANOVA to test the effects of the algorithm on the dependent variables. The tests suggest that the algorithm has a significant impact ($p<0.05$) on all of the subjective variables (Choice Satisfaction, Perceived Risk, Perceived Time). We ran multiple pair-wise comparison post-hoc tests using Tukey's method on subjective variables. The results are shown in Figures 3–5, where the mean is represented by a circle and the 95% confidence interval as a line.

The results show that the adoption of the hybrid *Interleaved* algorithm significantly increases user satisfaction and decreases both perceived risk and effort with respect to the non-personalized *HotelAvg* configuration. On the contrary, content and collaborative algorithms (*DirectContent* and *pureSVD*) do not affect user satisfaction and perceived effort with respect to the baseline scenario, although *DirectContent* reduces the perceived risk.

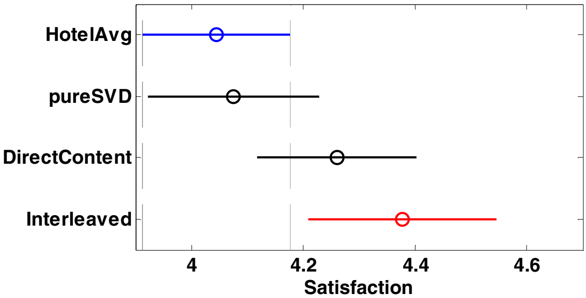


Fig. 3. Choice Satisfaction: how much are you satisfied with your final choice? (1: not too much – 5: very much)

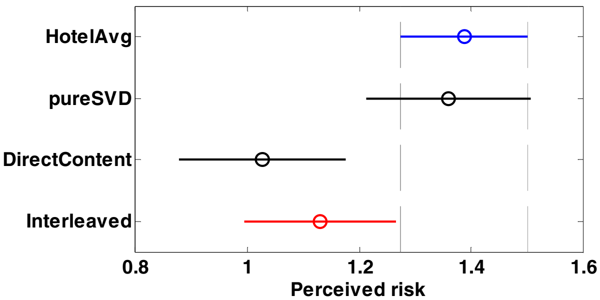


Fig. 4. Choice Risk: how much the characteristics of the reserved hotel will correspond to those of the real accommodation? (1: not much – 5: very much)

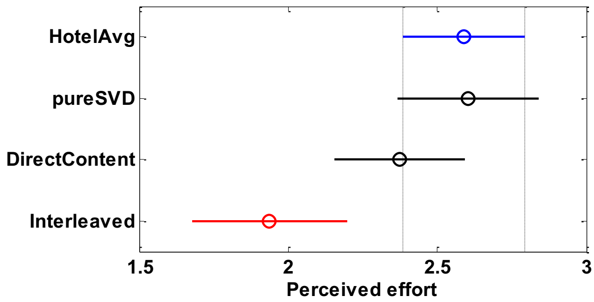


Fig. 5. Perceived Effort: the time required to book the hotel is … (1: short – 5: overmuch)

Objective Metrics

Behavioral data emerging from the tracking of users' interaction provides some interesting results. PoliVenus does not us enable to directly measure the efficacy of the RS using conversion rate metrics, because it does not explicitly present a separate list of recommended items. The hotel list view, where hotels are sorted in descending order of relevance as estimated by the recommender algorithm, is the actual recommendation list. As mentioned in the previous section, we have estimated the *efficacy* of the RS by measuring how many users changed the default sorting options (recommended hotels first) by setting other parameters (e.g., price, popularity, stars). The results show that only *37%* of the users with personalized recommendations changed the sorting of hotels compared to *54%* of the users in the baseline scenario (e.g., with non-personalized recommendations).

Surprisingly, the value of the two objective effort variables (*elapsed time* and *number of explored hotels*) is not affected by the recommender systems. Regardless the presence of personalized recommendations, the *number of explored hotels* is between 5 and 8 for almost all of the users. In other words, customers wish to compare in details at least 5 and no more than 8 alternative choices before committing to a final decision. Moreover, the time required to complete the task is between 10 and 22 minutes for 95% of the users and is not affected by the presence of personalized recommendations.

5 Discussion and Conclusions

This section discusses the validity of our empirical study, analyses the findings in relationship to our research question, and provides some possible interpretation of results, also in light of the current state of the art.

5.1 Validity of Our Study

The internal validity of our study is supported by the accuracy of our research design and by the quality of study execution. We have carefully implemented a mechanism to randomly assign participants to the different experimental conditions. We have adopted a lottery incentive to improve the accuracy of the task's execution and offered a shared motivation to all participants. Obviously, the individuals' intrinsic characteristics and actual behavior always bring to an experiment a myriad of factors that can be hardly controlled. In terms of external validity, the results of our study are limited to those participants and conditions used in our study. The applicability of our results might not be confined to the specific online booking system used. Most services available in the market provide a user experience very similar to Venere.com, in terms of filtering criteria and information/navigation structures, and it is likely that replications of our study on other systems may lead to results consistent with our findings. Finally, the high overall number of testers (240) and the relatively high number of subjects involved in each experimental condition (between 54 and 58) may allow us to generalize our results to a wider population of users aged 20-40.

In the "with-RS" experimental conditions, there might be a potentially confounding factor which might affect the results. The list of presented hotels for a user in the "with-RS" condition changes after each interaction, as the user profile is updated with each new "signal". On the contrary, the user in the "without-RS" condition sees no updates to the list of presented hotels, unless she explicitly changes the search criteria. This phenomenon could explain why users not receiving personalized recommendations take more initiative in changing the sorting of hotels in the list. However, as the different "with-RS" conditions provided exactly the same results, we may argue that the effect of recommender systems on the users' behavior is genuine and not affected by the different "dynamics" of the list of hotels "with" and "without recommendations". To further validate this assumption, in future tests it may be worth considering changing the list of hotels in the menu entry of presented hotels after each signal also for the no-RS subjects (in some random re-sort).

5.2 Research Question

Overall, our findings *answer our research question*: system-centric evaluation can be used to compare the effects of different recommender algorithms on the decision-making process in online hotel booking. A finer grained analysis of the statistically relevant relationships among all the different variables offers a much more articulated picture of the results, which are apparently in contrast with previous findings and suggest a number of interesting considerations for e-tourism applications:

- system-centric accuracy metrics (e.g., recall and fallout) are a good approximation of the quality perceived by the use;
- personalized recommendations do not reduce the decision-making effort; moreover, there is a mismatch between objective and perceived effort.

System-centric vs. User-centric Evaluation

The comparison between Figures 2 and 3 shows a strong consistency between system-centric accuracy (e.g., recall and fallout) and the users' perceived quality. We can claim that, in the e-tourism domain, system-centric quality attributes are good predictors of how the users perceive the quality of a recommender algorithm. This result is apparently in contrast with previous works [6,19,22,27] which state that user satisfaction is not correlated with accuracy of algorithms.

A possible interpretation of this result is to consider that, in traditional e-commerce applications, non-accuracy attributes such as novelty have an important role in shaping the user satisfaction. This is not necessarily the case with touristic services as they have a number of peculiar aspects that call for a rethinking of the decision making process normally assumed in traditional e-commerce and affect which RSs attributes impact on the user final choices [11,25,26,30]:

- The touristic product is complex and emotional. Booking hotel rooms requires a considerable decision effort, as the user might lack background knowledge of the characteristic of the possible accommodations. A priori comprehensive assessment of hotel accommodations is impossible. Touristic products lack the feature of "try-before-buy" or "return-if-not-satisfied". In case the quality of the accommodation is below expectance, the whole travel experience will be negatively affected and the memory of it will be persistent over time. The opportunity to fill that particular period of the user life with a positive experience will be lost.
- The touristic product is volatile. Hotel rooms are limited in number and, especially in high-seasons, the best-value accommodations are sold-out months in advance. Even optimizers (i.e., users who would choose the "best" possible hotel) could be urged to take sub-optimal decisions.

These aspects imply that (i) online hotel booking involves a certain amount of risk taking, and (ii) the user needs to minimize such risk by searching for an accommodation perfectly matching his/her largely unexpressed requirements. Accuracy, e.g., the perfect matching between user needs and solutions proposed, may become the most important attribute of recommendations as it reduces the perceived risk.

Decision Effort

There is *no significant variation among personalized algorithms with* respect *to objective effort*: none of them statistically differ from the baseline in terms of execution time and extent of product search, i.e., number of explored hotel pages. Our results show a *mismatch between satisfaction* and *effort*: users exposed to hybrid and content recommendations perceived the decision activity process as more satisfying than those without personalized recommendations, although they spent the same time in the process. Moreover, there is a *mismatch between objective and perceived effort*: users exposed to hybrid recommendations perceived the decision activity process as shorter (Figure 5), and more trustworthy (Figure 4), than the others, although they spent the same time and explored the same number of hotels.

This result is partially in line with some previous studies which explored subjective vs. objective effort under different conditions of RS design [24]. Works on preferences elicitation pinpoint for example that more effort-intensive sign up activities do not necessarily imply higher perceived effort [27], as if the effort perception were mitigated by the benefits of a more satisfying decision process. Other works hypothesize that, thanks to RSs, users spend less time in searching for items and more time in the more satisfactory activity of exploring information related to the choice processes [24].

5.3 Conclusions

In spite of some limitations, our work represents a contribution to the research and practice in RS design and evaluation, for the specific domain of online booking and from a more general perspective. Our research differs from previous work in this domain for a number of aspects:

- We smoothly *extend* a *true* online booking system (Venere.com) with personalized recommendations without creating major modifications to the "standard" interaction flow and overall user experience. Our user – both in the control group and in the treatment groups – is always in control of all information and functionality of conventional online booking services. In contrast, most previous works either are based on prototype RSs or create ad hoc user experiences for the subjects exposed to recommendations.
- We compare *three different personalized algorithms* against the baseline scenario (non-personalized) and against each other. Previous works limit their analysis to a single recommendation algorithm evaluated against a non personalized baseline. In addition, we consider recommenders involving *implicit* elicitation, while most of existing studies in the e-tourism domain address recommenders with (more intrusive) explicit elicitation
- Our results on the comparison between system-centric and user-centric evaluation are totally new for the e-tourism domain. They show that the peculiarity of the touristic products may exploit different aspects of recommender systems, and suggest the possibility to adopt system-centric evaluation techniques as a good approximation of the user experience.

In relationship to other studies in e-tourism and in other domains, the *research design* of our empirical study is per se a strength of our work, for a number of reasons: the large number of variables that have been measured, the sophisticated technological instrument used (the PoliVenus framework), the vast size of the involved subjects (240), and the lottery based incentive mechanisms adopted to motivate users and commit them to realistic and sound task execution.

Overall, our findings extend our understanding of the potential of introducing recommendations to improve decision making processes. At the same time, our results help us to identify potential weakness of current state of the art approaches in RSs and can orient future investigations in the field.

References

1. Bambini, R., Cremonesi, P., Turrin, R.: A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 299–331. Springer, US (2011)
2. Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding choice overload in recommender systems. In: Proc. of the Fourth ACM Conference on Recommender Systems (RecSys 2010), pp. 63–70. ACM, New York (2010)
3. Burke, R.: Hybrid web recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
4. Celma, O., Herrera, P.: A new approach to evaluating novel recommendations. In: Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys 2008), pp. 179–186. ACM, New York (2008)
5. Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A., Turrin, R.: Looking for "good" recommendations: a comparative evaluation of recommender systems. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 152–168. Springer, Heidelberg (2011)
6. Cremonesi, P., Garzotto, F., Turrin, R.: Investigating the Persuasion Potential of Recommender Systems from a Quality Perspective: an Empirical Study. ACM Transactions on Interactive Intelligent Systems 2(2), 41 pages (2012)
7. Delgado, J., Davidson, R.: Knowledge bases and user profiling in travel and hospitality recommender systems (2002)
8. Deshpande, M., Karypis, G.: Item-centric top-N recommendation algorithms. ACM Trans. Inf. Syst. 22(1), 143–177 (2004)
9. Drenner, S., Sen, S., Terveen, L.: Crafting the initial user experience to achieve community goals. In: Proc. of RecSys 2008, pp. 187–194. ACM, New York (2008)
10. Golbandi, N., Koren, Y., Lempel, R.: On bootstrapping recommender systems. In: Proc. of the 19th ACM Int. Conf. on Information and Knowledge Management, pp. 1805–1808. ACM, New York (2010)
11. Hennig-Thurau, T., Gwinner, K.P., Walsh, G., Gremler, D.D.: Electronic Word-of Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? Journal of Interactive Marketing 18(1), 38–52 (2004)
12. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22(1), 53 (2004)
13. Jannach, D., Zanker, M., Jessenitschnig, M., Seidler, O.: Developing a conversational travel advisor with advisor suite. In: Sigala, M., Mich, L., Murphy, J. (eds.) Information and Communication Technologies in Tourism, pp. 43–52. Springer, Vienna (2007)
14. Jones, N., Pu, P.: User technology adoption issues in recommender systems. In: Proc. of Networking and Electronic Commerce Research Conf. (NAEC 2007), pp. 379–394 (2007)
15. Konstan, J.A., Riedl, J.: Recommender systems: from algorithms to user experience. User Model. User-Adapt. Interact. 22, 101–123 (2012)
16. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008), pp. 426–434. ACM, New York (2008)
17. Levi, A., Mokryn, O., Diot, C., Taft, N.: Finding a needle in a haystack of reviews: cold start context-centric hotel recommender system. In: Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys 2012), pp. 115–122. ACM, New York (2012)

18. Lops, P., De Gemmis, M., Semeraro, G.: Content-centric recommender systems: State of the art and trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 73–105. Springer, US (2011)
19. Mahmood, T., Ricci, F., Venturini, A., Höpken, W.: Adaptive recommender systems for travel planning. In: Proceedings of the International Conference on Information Technology and Travel & Tourism, ENTER, pp. 1–11. Springer (2008)
20. Mcnee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: Extended Abstracts on Human Factors in Computing Systems (CHI 2006), pp. 1097–1101. ACM, New York (2006)
21. Porter, S.R., Whitcomb, M.E.: The Impact of Lottery Incentives on Survey Response Rates. Research in Higher Education 44(4), 389–407 (2003)
22. Pu, P., Chen, L., Hu, R.: A User-Centric Evaluation Framework for Recommender Systems. In: Proc. of RecSys 2011, pp. 157–164. ACM, New York (2011)
23. Rashid, A.M., Karypis, G., Riedl, J.: Learning preferences of new users in recommender systems: an information theoretic approach. SIGKDD Explorer Newsletter 10, 90–100 (2008)
24. Ricci, F., Missier, F.: Supporting travel decision making through personalized recommendation. In: Karat, C.-M., Blom, J., Karat, J. (eds.) Designing Personalized User Experiences in eCommerce. Human–Computer Interaction Series, vol. 5, pp. 231–251. Springer, Netherlands (2004)
25. Ricci, F., Wietsma, R.: Product Reviews in Travel Decision Making. In: Proc. of the International Conference on Information and Communication Technologies in Tourism 2006, Lausanne, Switzerland, pp. 296–307. Springer Verlag (2006)
26. Senecal, S., Nantel, J.: The Influence of Online Product Recommendations on Consumers' Online Choices. Journal of Retailing 80(2), 159–169 (2004)
27. Swearingen, K., Sinha, R.: Beyond algorithms: An hci perspective on recommender systems. In: Proceedings of the ACM SIGIR Workshop on Recommender Systems (2001)
28. Takács, G., Pilászy, I., Németh, B., Tikk, D.: Scalable Collaborative Filtering Approaches for Large Recommender Systems. J. Mach. Learn. Res. 10, 623–656 (2009)
29. Tintarev, N.: Explanations of recommendations. In: Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys 2007), pp. 203–206. ACM, New York (2007)
30. Werthner, H., Ricci, F.: E-commerce and tourism. Commun. ACM 47(12), 101–105 (2004)
31. Xie, M., Lakshmanan, L.V.S., Wood, P.T.: Comprec-trip: A composite recommendation system for travel planning. In: 2011 IEEE 27th International Conference on Data Engineering (ICDE), pp. 1352–1355 (April 2011)
32. Zhang, Y., Callan, J., Minka, T.: Novelty and redundancy detection in adaptive filtering. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002), pp. 81–88. ACM, New York (2002)
33. Zanker, M., Fuchs, M.W., Mario Tuta, H., Müller, N.: Evaluating recommender systems in tourism - a case study from austria. In: Proceedings of the International Conference on Information Technology and Travel & Tourism, ENTER, pp. 24–34. Springer (2008)
34. Ziegler, C.N., McNee, S.M., Konstan, J.A.: Lausen. G., Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web (WWW 2005), pp. 22–32. ACM, New York (2005)

Video Navigation with a Personal Viewing History

Abir Al-Hajri, Gregor Miller, Sidney Fels, and Matthew Fong

Human Communication Technologies Laboratory,
University of British Columbia, Vancouver, BC, Canada
{abira,gregor,ssfels,mfong}@ece.ubc.ca

Abstract. We describe a new video interface based on a recorded personal navigation history which provides simple mechanisms to quickly find and watch previously viewed intervals, highlight segments of video the user found interesting and support other video tasks such as crowd-sourced video popularity measures and consumer-level video editing. Our novel history interface lets users find previously viewed intervals more quickly and provides a more enjoyable video navigation experience, as demonstrated by the study we performed. The user study tasked participants with viewing a pre-defined history of a subset of the video and answering questions about the video content: 83.9% of questions (average) were answered correctly using the personal navigation history, while 65.5% were answered using the state-of-art method; they took significantly less time to answer a question using our method. The full video navigation interface received an 82% average QUIS rating. The results show that our history interface can be an effective part of video players and browsers.

Keywords: Video Navigation, Navigation History, Video Summarization.

1 Introduction

The dramatic increase in the quantity of video now available, either in the online form of services such as YouTube™, Vimeo™ or the more personal form of home video, provides new challenges such as finding specific videos (or intervals within video), authoring new videos from existing content (e.g. video summarization or home movie editing), sharing video playlists (or even intervals) of online video and annotating dynamic content. The increased volume and shrinking average duration of online video has led to new use cases: users are generally free to skip ahead to find intervals which interest them; users may re-watch part of the video they enjoyed; users may employ temporal video links to provide instant access to a specific time within a video, skipping irrelevant sections; playlists and temporal links provide opportunities for customization and sharing of video consumption among users. In short, users can and want to view personally interesting intervals easily and without viewing the entire video. These use cases are supported within current interfaces, however there are no existing mechanisms to support new use cases which arise as a natural consequence, such as navigation of previously viewed intervals. Additionally, users often re-watch segments of video as part of the contemporary browsing and navigation experience

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 352–369, 2013.

[24, 25, 28], and may wish to share these intervals with others. A historical record of video viewing would allow users to quickly find popular intervals, comment on and share them.

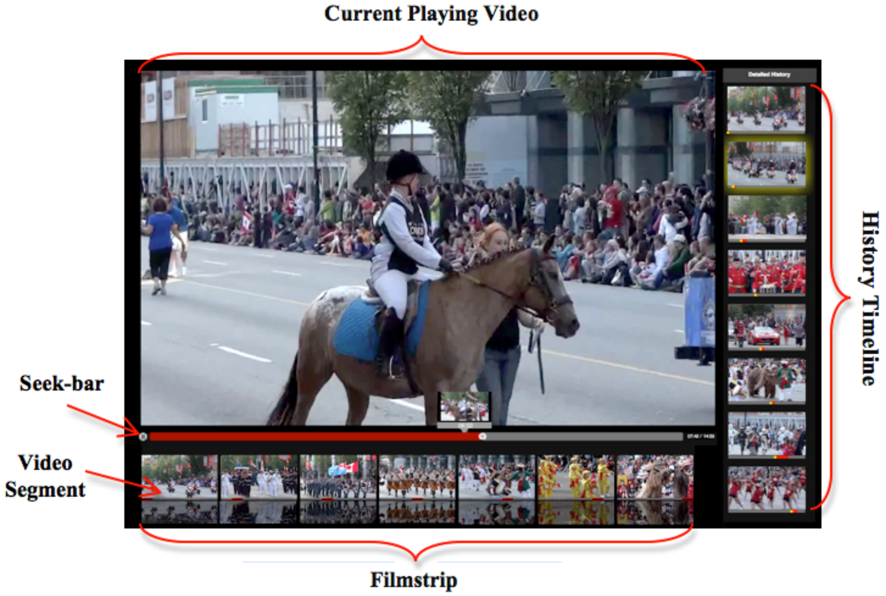


Fig. 1. Our video navigation interface: the majority of space is devoted to the currently playing video (top left) with a seek-bar preview; below is a horizontal array of Video Segments arranged by video-time (the Filmstrip), and a vertical array of Video Segments to the right (the History Timeline) ordered top-down by user-time i.e. the order the intervals were viewed. These components described in more details in section 4.

We introduce our novel video interface, which captures the user's viewing history (analogous to a web browser history) and allows them to reuse it as a navigation tool, through a user-time visualization with supporting interactions. The history supports actions such as faster access to previously viewed intervals, visualization of popular intervals within the video based on a user's temporal view count, simpler authoring via selection of viewed intervals and provides the user with the ability to create a personalized map of the video (e.g. deciding on their own separation of the video into chapters, or highlighting interesting intervals). This type of history interface is not currently supported in commercial products, thus, we are studying interfaces for managing personal viewing history to evaluate effectiveness for inclusion in future video interfaces.

Our research focuses on the utility of video history, examining use cases and establishing which are significantly affected given a user-based timeline of video viewing. The interface we designed (shown in Figure 1) includes a filmstrip visualization below the main video player representing the entire video, and a history to the right of the player with a sequence of intervals ordered top-down by user-time. We discovered through pilot studies that the filmstrip and history are equivalent in navigation speed to find a previously viewed event in the case where the entire video is viewed, although the history interface was preferred since users were able to customize which intervals were visualized. Our study shows that when a subset of the video is viewed, the history significantly outperforms the state-of-art, and 75% of participants preferred the history in this use case. The study found that our history-based authoring interface was useful, easy to use and enjoyable, and participants were keen to see this functionality in YouTube.

We begin with a discussion of the related work, followed by a description of the video navigation interface (including all widgets and interactions used). The evaluation of the use of a history and the interfaces we designed to support it is presented, looking at users' impressions of using a personal video history and the power of using history for fast video navigation.

2 Motivation

Consider the following scenario with Robert and Elisa:

> *Robert received a YouTube link from his friend Tom about his family vacation to Rome last summer. Robert and his wife Elisa watched the video along with other videos about Italy to enjoy a sense of Tom's holiday. Robert enjoyed some segments of the videos and wanted to share them with his kids when they came home. When they arrive, he browses his history view and jumps to the good bits to show them. After seeing how nicely the Italy videos looked, Elisa suggests creating a video of their last vacation to Vancouver using various videos and photos they captured. Robert had watched these videos many times and he was excited about including particular events from the 12 hours of video he had obsessively taken during their 4-day holiday. He goes to his history view and grabs the shots of the parade and fire works of Canada day, the suspension bridge and swimming at Lynn Canyon Park, along with another 10 segments that he had watched at various times with Elisa and the kids before. He then drags them into the authoring window and creates a new video that he shares with his friends. .*

This scenario emphasizes the major contributions of our work: (1) a novel video navigation interface for video viewing that captures and visualizes a user's personal history; (2) an experiment that confirms using personal video navigation history is faster in finding previously seen events within videos in contrast to the state-of-art method, filmstrip; (3) a simple video mash-up tool that can be used to summarize videos by using intervals from history or simply watching the corresponding intervals, in comparison to other video editing software where users need to add a video clip, watch it, and then trim it.

3 Related Work

Retrieving and reusing information from the past is a common activity for users [6]. In the domain of web browsers, researchers have introduced and developed different tools that keep records of user's browsing experiences for later use. However, keeping and reusing navigation experiences in a video space has not received similar attention. Web browsing history is useful for various domains, thus, we expect our research on effective mechanisms for video watching history management will also benefit various fields where video plays an increasingly important role e.g. education, entertainment, etc. History of user actions has been widely investigated for multiple purposes in different domains including web [2, 19, 15], workflow [12, 21], tutorial generation [1] and information visualizations [14, 13]. These are based on mostly static media and so do not directly apply to video: unlike static media, two timelines exist when capturing user history: video timeline and user (or history) timeline. Recording a user's actions, therefore, requires more sophisticated representations, visualization layouts and interaction techniques. In our interface to eliminate the confusion between different timelines, we used a vertical layout to indicate a user's timeline while a horizontal layout indicates a video timeline consistent with usual video players.

The relatively high interest in the social web, sharing and the use of videos online has motivated researchers to investigate the use of video navigation history. Users unintentionally leave footprints during their video-browsing process, which can add value to the content. Yu et al. [28] and Syeda et al. [25] investigated how users' video browsing footprints can be utilized in video summarization and to facilitate future viewers. They used video browsing footprints to rank different scenes within a video, which can later be used to extract clips when generating video summaries or previews. However, these systems [25, 28] have not given the users any control over their personal history during the viewing process (i.e. users cannot access their history). They recorded the users' actions, which were later used by the researchers to generate previews or evaluate the ranking of interesting parts within videos. In our interface we offer the users the ability to utilize their personal history for navigation and authoring as well as summarization.

Video navigation history can play an important role in user-based information retrieval from videos on the web. Shamma et al. [24] and Yew et al. [27] have proposed a shift from content-based techniques to user-based analysis because it provides a more promising basis for indexing media content in ways that satisfy user needs. Leftheriotis et al. [18] proposed a video indexing tool on the web based on user interactions to rank different videos. They generated a thumbnail to represent each video based on the frame that has been mostly viewed by users. Gkonela et al. [10] also indicated that this simple user heuristic while navigating videos can be used as effectively to detect video-events as when content-based techniques are applied. In our interface, we applied a flexible thumbnail visualization of each history segment: initially, each segment is visualized using a thumbnail that indicate the beginning of that segment which then changes while the user navigates this segment.

Navigating a video space or even long videos can be demanding and time consuming. There have been many interaction techniques proposed in the literature to address this problem in order to quickly navigate and search video content; here, we will only mention a few. Typical navigation tools used in most video systems are the VCR-like controls (play, pause, seek, fast forward, and rewind). DVD systems provided users with a higher level of video access using a random access to video segments based on scene indices and chapter boundaries via a table of contents index. Some researchers

[26, 23, 5, 8] allowed video navigation using their abstraction thumbnails or key-frames (e.g. filmstrip), where clicking on a thumbnail directly positions the video at a particular point in time corresponding to that thumbnail. Others [3, 9] proposed new playback mechanisms to help users rapidly skim through uninteresting content while slowing down to watch interesting parts of the video. Since the video content changes as time passes, the separation between the video content and the controls used to manipulate the video introduces a target acquisition problem. Thus, some researchers have focused in introducing new techniques to manipulate the video timeline rather than using the traditional VCR controls. They allow users to navigate the video by directly manipulating the video content along its natural movement path rather than using the timeline slider [17, 16, 7, 11]. Pongnumkul et al. [22] developed a new navigation tool that uses a map-based storyboard where different locations of video segments are directly visualized on the map.

Video navigation history provides a potential interaction technique that can be used to facilitate fast navigation and event search. According to Shamma et al. [24], the more affective a scene is, the more the corresponding interval of the video is viewed or consumed. Thus, the cumulative seamless user interaction history could be leveraged for the benefit of future viewers (i.e. social navigation). For example in YouTube, when enough user data is available, user behavior will exhibit similar patterns which can then be used by new viewers to quickly navigate to the interesting part of that video. Yu et al. [28] used low-level feature extraction along with user's footprints (i.e. view count) to rank each scene and offer users other scenes that have some correlation with what they are watching. This provides users with a quick navigation method to similar scenes. Mertens et al. [20] visualized users' footprints on the video timeline, which allowed users to quickly navigate to the corresponding scenes in the video. In our interface, we used VCR like controls along with thumbnails of the previously viewed video segments that allow simple navigation to these intervals. Our study revealed that using this method speeds up searching for previously seen events within the video.

4 Video Navigation Interface

Given the volume of available video data, methods to effectively navigate this space are needed to provide users with more enjoyable video viewing experiences. The principle goal of our interface is to provide efficient access to previously viewed video: in particular, intervals within video. We accomplish this by recording a continuous video history for the user as they consume video content. This paper considers only the history of a single video as a starting point, and demonstrates that further research is warranted for multiple-video history based on our success.

Internally, the user's video watching history is represented by a very simple data structure organized by linear user time. Whenever a video begins playback the video ID and current user time are recorded; these are noted again when the video stops playing, to produce interval start and end timestamps. An accumulated view count is maintained for every instant of time in the video: the structure stores intervals of view count derived from the intersection of all viewed intervals.

We designed our interface in a way that allows users to play, view and navigate videos similar to any video player as we envision our history management to be used to augment it. Since our interface records the user's navigation history, we propose new components that offer the users the flexibility of using this history for fast navigation, summarization (or authoring) and sharing. Thus, our interface is based on five separate components: a Player which provides the user with a familiar video setting, a Filmstrip which provides a navigation tool for the entire video content, a History Timeline which visualizes the personal viewing history, a Video Mash-up and a Preview Window which help users creating and previewing their own edit list for video summaries or previews. The main interface components are: the Player and the History Timeline.

Fig. 2. The video player component contains play/pause button, video seek-bar, and a pop-up thumbnail shown when the pointer is over the time bar

Player. A familiar video player (similar to YouTube, QuickTime, etc.), shown in Figure 2, contains: play/pause, applied via a button (bottom left) or by clicking directly on the video; seeking via playhead (white circle on timeline) or mouse click on the timeline (red/grey bar); a frame preview triggered by the mouse over the timeline (including when seeking interactively i.e. with mouse button held down over time slider); and a current playhead time label (white text, bottom right). This component is used for direct control of the selected video. To watch a specific video from disk, users click on the "Open" button and choose their video file.

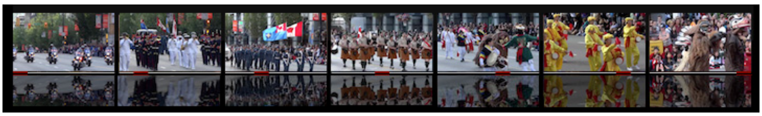


Fig. 3. The Filmstrip component visualizes the entire video into *n* equal length segments

Filmstrip. The Filmstrip component, shown in Figure 3, is the state-of-art video navigation tool. It provides a visualization aid to different parts of a video for faster navigation and supports access to the entire video content. The Filmstrip simply consists of a fixed number of Video Segments (*n*) from the playing video. The entire video is divided into *n* equal length intervals, where each interval is represented by a small seek-able Video Segment described below. These intervals are created systematically based on the length of the video and the number of segments to be visualized.

User Viewing History Timeline (History Timeline). This is the central component for visualizing the personal viewing history of the user. We use a filmstrip metaphor for visualization: the History Timeline consists of thumbnails (similar to interactive storyboards) as shown in Figure 1. These thumbnails represent Video Segments that the user watched. We represent this as a Video Segment component as discussed in the next section. Every time the user seeks to a new temporal location, a new Video Segment is added to the end of the user's history list and represents the interval a user watches in the main video player. The History Timeline is a scrollable box, to provide navigation of the entire viewing history and allows the user to find a specific interval.

The video history is similar conceptually to webpage bookmarks that are used to refer back to visited pages, however, it is more complicated as each Video Segment in the history view is a video interval and also includes a representation of the time the user watched it. Thus representing both of these quantities requires additional mechanisms for visualization and interaction. Moreover, there may be confusion with the currently viewed timeline and with the history views since there exist two timelines: when/what the user watched and the video's timeline. We address this by visualizing the history timeline vertically, while the video timeline remains horizontal, consistent with user's mental model of video. The history is unique in that it records in piecewise linear user-time, not video time (we only display the history of intervals watched and do not represent the complete video from where it came from in the History view.). Each Video Segment is a small interactive video widget described next.

Video Segment. Each interval a user watches is represented as a small video widget that contains the viewed interval only, visualized by the starting frame of that interval. The location of the interval within the entire video is visualized as a red portion underneath the widget to help users spatially contextualize the temporal location of intervals within the complete video where it came from. The video widget, shown in Figure 1, is seek-able and playable to allow users to easily search within the intervals and minimize the time needed to search for a previously viewed scene. On mouse over, a play/pause overlay is displayed over the widget. Clicking on the overlay plays the corresponding video interval within this small widget only. (Note, while technically, this is a rewatching of video, we do not add this activity in the History View.) Moving the cursor over the bottom third portion of the widget pops up a zoom-in visualization of the interval timeline, which can be used to seek within the video interval using mouse motion as a cue. The seeking point is visualized by a yellow line and the thumbnail updates to reflect the current seek position. To play a Video Segment in the main video player, users can drag the entire widget (or the seek location in the widget) to the main video player: dragging the entire widget plays the corresponding interval from the beginning, while dragging the seek bar location plays from that specific time.

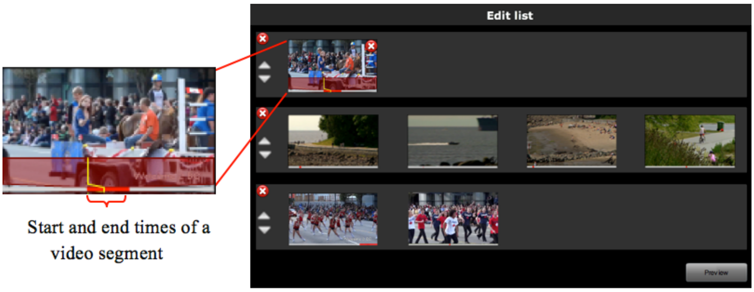


Fig. 4. The Video Mash-up component illustrates a user edit list which consists of a user defined Video Segments with the options of delete, reorder, play (on click), navigate (on cursor move shown as a yellow line), and modify segment's start and end times.

Video Mash-Up. The video mash-up is used for short video composition such as summaries, previews, or trailers (analogous to video editing software e.g. iMovie, Movie Maker, etc). We included this component in our prototype to evaluate one of the use cases we idenitified that takes advantage of potential strength of having personal viewing history. The video mash-up, shown in Figure 4, represents a video edit list that shows the videos, and intervals within each video, that a user may combine into a summary. In summary creation mode, when the user subsequently plays a video (either through opening a new video, a segment from history or seeking within the current video) a new record is added to the edit list. If the video already exists in the edit list, a new interval is added (unless there is an existing identical interval). Otherwise if the video does not exist, a new video record with the interval is added to the edit list. Thus, adding segments or intervals to the edit list is simply done by watching these intervals, in comparison to other video editing software where users need to add a video clip, watch it, and then trim it. Using this component makes video editing simpler for novice users as we discovered upon evaluation.

The video mash-up component allows the user to preview a specific interval by using the play button for that interval. Users can also modify their edit list by directly updating interval start or end times, delete an interval, delete an entire video record, or reorder video records. Previewing the entire summary is provided by clicking on the "Preview" button.

Summary Preview. After creating a video mash-up, the user can preview it using the summary preview mode, where a new video player window pops up, and the rest of the interface is disabled and darkened. The summary preview takes the user's edit list as input and plays the intervals in order using this content. It allows users to replay, modify (back to the main interface to modify the edit list), or export it if they wish to save it - the edit list is exported as an XML file that can be shared or modified later.

Personal Navigation Footprints. The recorded video history offers additional benefits, such as within-video history and crowd-sourced user histories (neither of these were included in our study). The personal history or footprints offers users a visualization of the view count of every point of time in the video: the blue bar in Figure 5 becomes brighter for the most-viewed intervals, and each one can be selected to play. It provides the user with information on the most-viewed intervals of each video (a temporally tracked frequency). As mentioned previously, a view count is maintained for all points of time for each viewed video. As the user views a video the view count will dynamically change, thus the count is continuously updated and the information propagated backwards through the history. The view counts are then summed and normalized within each video.

Fig. 5. A Video Timeline is combined with navigation footprints visualization, where the more an interval within a video is watched the brighter that region becomes. Blue: a single user history (Personal Navigation Footprints), Red: a combined multiple users histories (Crowd-Sourced Navigation Footprints).

Crowd-Sourced Navigation Footprints. The personal history from multiple users are combined to form the most popular intervals for a video (i.e. social navigation) similar to these are shown in the red bar in Figure 5. This provides users who are watching a video for the first time with a mechanism to watch what others have found to be most interesting. This is an efficient navigation technique, predicated on trusting the crowd to be "correct" on interesting video content.

5 Evaluation

An extensive set of pilot studies were performed (12 participants total) to investigate a user's preferred scenario for history, and to triangulate the use cases in which a history is beneficial for video viewing. Throughout this triangulation process it became more apparent to the authors how complex and intricate the video-viewing task can be. For tasks such as seeking to a specific time or finding a particular event from a previously watched video, using a history was found to be as good as using a filmstrip methodology. Under more modern viewing patterns (e.g. nonlinear viewing behavior) where users view only parts of the video (e.g. trailers, summaries, playlists or direct temporal links) using a history to find events and seek to them was found to be more efficient.

Using this result from the pilot studies, a full comparative user study was performed to evaluate the design and performance of our interface and to demonstrate the utility of personal video navigation histories. We developed an evaluation protocol which satisfies the use cases previously defined (without biasing our design or the control) and provides the user with sufficient viewing history while keeping the experiment relatively short. For fair comparison we ensured our interface mimics currently adopted approaches with logical extensions.

Using our protocol, we investigated whether visualizing and using a video navigation history would make searching for previously seen events more efficient. We conducted a comparative between-subjects user study, comparing the performance of users employing a personal navigation history against those with the state-of-the-art navigation method (Filmstrip) to find events within a previously seen video (i.e. finding answers to questions similar to [28, 10]).

Both methods have similar layouts and functionality, differing on the process of segment creation described in section 4. In Filmstrip the intervals are created systematically, while in the History Timeline segments are constructed based on a personal navigation history; each participant tried both methods, on different videos. In Filmstrip tasks the participants were presented with the main Video Player and a horizontal Filmstrip components only. They only used these components and their features to find the answer to each question. In History Timeline tasks the participants were presented with the main Video Player and a vertical History Timeline components.

Our experiment was divided into two phases: phase 1 was conducted to compare the two methods (Filmstrip and History), and phase 2 was performed to qualitatively analyze the entire proposed interface, described in section 4. In phase 1, participants experimented 3 components (Video Player, Filmstrip, and History Timeline) of the interface described in section 4, while in the second phase they were exposed to all components and features.

5.1 Apparatus

The experiment application was developed in Flash CS4 and ActionScript 3.0. The experiment ran on an Intel dual-processor dual-core 3 GHz Mac Pro desktop with 8GB RAM and equipped with a 24" Dell LCD monitor with a resolution of 1920 x 1200 pixels at a refresh rate 0f 60Hz. A Microsoft optical wheel mouse was used as the input pointing device with default settings and the Adobe Air environment was set to 1920 x 900 pixels while running the Flash program.

5.2 Participants

Twelve paid university students (different from those in the pilot studies), 6 female and 6 male, participated in the experiment. Participants ranged in age from 20 to 30, all were experienced computer users and have either normal or corrected-to-normal vision. Participants reported watching videos either on a daily basis (10 participants) or 3-5 times a week (2 participants). Each participant worked on the task individually.

This project targets heavy video users, and according to Purcel[1], video is a growing trend especially amongst those aged 18 to 29. Thus, students are a primary target audience as they are representative of the current/near-future heavy users of video interfaces. We used students across the University to capture this demographic.

5.3 Design

To evaluate the efficiency of using navigation history as an aid for efficient search within a previously seen video, we gave participants a pre-defined history, which they

[1] http://www.pewinternet.org/ /media//Files/Reports/2010/PIP-The-State-of-Online-Video.

had to watch and subsequently answer questions based on the content. We followed this procedure rather than allowing participants to create their own history: this decision was made based on the pilot studies where participants found creating a history based on an unknown list of questions was confusing. Some participants said, "What interests me might not be what you are looking for". Thus, they tried to create many history segments at each point they thought there was potential for a question. The result was a long list of short segments which required significant scrolling when searching. In short, the uncertainty of what should be included on the list led to a large number of segments and created some confusion that affected the performance while using the method.

In order to tackle this problem, the participants could be allowed to create their own history based on their interest after which they can be asked questions that exist within these segments. However, the variation between participants interest would make the comparison difficult. Thus, instead of asking participants to create their own history we decided to give them all the same history from which to answer the list of questions (analogous to watching a playlist created by a friend or a video summary). All participants watched the same clips from the video and were asked questions from these clips only. Thus, we ensured they had experienced the same clips and eliminated the personalization factor within the history. Our method was compared with the state-of-the-art method (Filmstrip). The History Timeline contained the segments of the predefined history only while Filmstrip contained seven evenly divided intervals over the entire video.

In this experiment, we used five different short videos (V1-5). V1 was used to explain the interface's components and features and demonstrate how to use them. The other four videos were used for the actual experiment where each participant used a single method for each video. Each participant experienced each method in a different video to eliminate the learning effect. The number of questions per video were: V1: 20; V2: 38; V3: 25; V4: 38; and V5: 32.

The participants were divided into two groups, A and B. Both groups experienced the videos in the same sequence but with different method sequences. They both started the familiarity phase where they tried a hybrid of the two methods to answer example questions and to become familiar with both methods. Group A was given Filmstrip for V2 and V4 and History Timeline for V3 and V5, while group B used the opposite. Thus, the results were compared between the two groups for each video.

To quantitatively compare the results of the two methods, we measured five variables for each video: (1) The percentage of questions answered for each video within a specified time range. The total time given is 15 seconds multiplied by the total number of questions (chosen based on the pilot studies); (2) The time needed to answer each question - this is the time measured between two subsequent correct submission clicks; (3) The number of wrong submissions (errors) for each question, which is measured by counting the number of submission clicks that did not result in a correct answer; (4) The number of seeks within a video for each question; and (5) The number of previews performed in a video for each question, which is measured by counting the number of seeks and playbacks within the small Video Segments. These counts were tracked to analyze the user's behaviour and to investigate the difference between the two methods.

The participants were also asked to fill out questionnaires and give their comments on the entire interface, and suggestions for improvement. This was conducted to qualitatively compare the two methods and to provide an indication of the importance of correcting each particular aspect of the proposed interface. Participants rated different interface features and tools using a 7-point Likert scale. The experiment lasted approximately two hours per participant.

Fig. 6. The Experiment Interface[2] illustrates the familiarity phase

5.4 Procedure

Participants were given a description of the procedures to be employed in the study, informed of the goals and purposes of the study, and informed consent was requested to participate in the study. Two methods were evaluated in the study: Filmstrip and History Timeline.

The researcher started by showing the interface (shown in Figure 6), describing its components, explaining the functionality of the tools within the interface and how to answer a question in order to complete each task. Participants were then asked to try the same interface with the first video and task. They were given the freedom to play with the interface and ask any questions. This familiarity stage was intended to allow the participant to try all the interface components and become familiar with the available functionality. The participants started watching the video provided without knowing the questions they will be asked.

For each video, the task started by playing the segments from the pre-defined history in the main video player. Participants were asked to watch the playback and pay attention to the video content in order to answer questions later. Once all segments from the history were viewed, the video paused and a list of questions (compiled by researchers) was displayed to the left of the main video player, as shown in Figure 6.

[2] The video images are for illustration purpose only as different stimulus was used in the experiment (i.e. Toy Story 3, Alma, Geri's Game, and For the Birds.).

Participants were asked to read these questions before clicking on the “Start” button, when the timer starts and they could begin providing answers. After clicking on the “Start” button, the method’s corresponding components were only shown and participants were advised to answer as many questions as they can within the given time by watching the clip which contains the answer.

The questions were randomly ordered (i.e. they were not temporally ordered according to their occurrence in the video). Each participant had the freedom to choose the order they would like to follow to answer these questions (e.g. temporally, linearly, difficulty, memorability, etc.). The participants started answering these questions by clicking on a “Start” button after reading the list of questions. In the familiarity phase, the participant could use either the Filmstrip or History Timeline to navigate the video in an attempt to find an answer. To answer a question the participant needed to re-watch the clip that contains the answer in the main video player. They navigated Video Segments in the provided component (Filmstrip or History Timeline) to find a clip or a video frame that contains the answer. Once they found the frame, they dragged the Video Segment of a specific video frame using the seek thumb of the Video Segment to the main video player to start playing that clip. After watching the interval that contains an answer for a single question they clicked on a submit button beside the corresponding question to proceed to the next question.

Each submitted clip was automatically evaluated: it was considered correct if it contains at least one frame from the ground truth answer. If the answer was correct, the question was faded out from the list and a message was displayed to participants (e.g. “Great. You have 2:14 to answer the rest. Hurry up.”) to encourage them and maintain time pressure. However, if the answer was wrong, a message is shown which asks the participant to try again “Sorry. Try again”. Participants repeated the same actions till they answered all questions or the time elapsed. A new list of questions was displayed and the participants continued applying the same procedure to answer these questions. The questions for each video were divided into two lists to avoid overloading participants with so many questions. A short break was provided between question sets.

After completing the second list of questions, the familiarity stage ended and the participants were advanced to phase 1 of the experiment where they watched a new video in similar settings to the watching phase of the familiarity stage. Upon the completion of the watching phase, a new list of questions was given based on the content of this video. In this stage the user was provided with only one method (Filmstrip or History Timeline) to be used in finding the answers. Once all questions were answered or the time elapsed, the participants proceeded to the next video where they repeated the previous stage but with a different video and using a different method than the one used for the previous video. Participants continued tasks until they completed all four videos.

During the experiment, participants were asked to fill out three questionnaires. These questionnaires were designed using the standard Questionnaire for User Interface Satisfaction (QUIS version 6.0 [4]), modified to reflect the functionality and tools applied and the usability of our interface. Some sections were removed (e.g. Terminology and system information, and System capabilities), some questions were modified, and others were added. The first questionnaire was after V4 in which participants evaluated the method they used for that video. The second questionnaire was after V5 where participants rated the usability of the second method they used. After completing the second questionnaire, participants were advanced to the second phase of the experiment. They were given the full interface, described in section 4, where they have the freedom to create their own history and explore use cases which might be applicable to them. Every performed seek within the video or any watched interval in the main video player was recorded and visualized as a Video Segment in the History Timeline. This helped the participants to understand how history is created for later usage. Once they were satisfied playing with the interface, they were asked to fill out the last QUIS questionnaire, which evaluates the entire interface.

Table 1. Methods Descriptives Using t-test for Equality of Means

	Filmstrip		History Timeline		
	M	SD	M	SD	t-test
Mean % of answered questions	65.53	14.69	83.93	14.87	4.53**
Average time per question	24.11	5.85	18.52	3.85	4.13**
Average no. of seeks	1.59	0.45	1.50	0.48	ns
Average no. of previews	14.37	5.14	12.74	4.91	ns
Average no. of errors	0.15	0.09	0.14	0.11	ns

$^{**}p < .01$.

Note. M = Mean. SD = Standard Deviation. ns = not significant. Average time per question is measured in seconds.

5.5 Results and Discussions

The study showed positive and promising results on the features of the interface. Most participants commented that they enjoyed their time using the interface and they can imagine seeing its features applied, especially in social networking websites (e.g. YouTube). They were impressed with the ease of learning and using the features within the interface.

T-test analysis results, illustrated in Table 1, showed that our novel method (History Timeline) was significantly faster than the state-of-the-art (Filmstrip) which allowed participants to answer more questions within the same time. However, both methods demonestrated similar behaviour in terms of average number of seeks, previews and errors.

Each participant answered two identical questionnaires, one for each method, to compare the overall reaction, the time it takes to learn and general impressions. Each questionnaire contains thirty-seven 7-point Likert scale questions. The data collected for the two methods were analyzed and compared in terms of these 37 questions. The results showed a significant difference taking the overall average rating of each method (Filmstrip ($M = 5.05$, $SD = 1.35$), History Timeline ($M = 5.55$, $SD = 1.06$), $t(886)= 2.08$, $p < 0.04$). The History Timeline showed no significant difference from the Filmstrip in all questions except for the question “Learning to use the method features" where (1: Difficult, 7:Easy) History Timeline (*M* = 5.4, *SD* = 1.07), and Filmstrip (*M* = 4.7, *SD* = 1.16), $t(22) = 2.27$, $p < 0.04$. This also coincided with participants preference where 9 participants out of 12 stated that the History Timeline was faster and easier in finding the answers for the questions. Participants reported that History Timeline was faster because it is based on a personal mental context map created for the video. It was easier for them to refer back to the corresponding segments when needed which is not the case for Filmstrip. Since Filmstrip segments was created systematically, participants needed at least to navigate one segment. If the answer was within that segment they submitted it, otherwise they needed to navigate the preceding or subsequent segment. This was also demonstrated by the participants' quantitative results where they were able to answer more questions using History Timeline. All questions for both methods were rated above 5 out of 7 except for four questions: "Satisfaction with the method", "Learning to use the method features", "Number of steps per task", and "Using the method is effortless".

After acquiring some familiarity with the overall interface in phase 2, participants were asked to complete a third questionnaire to rate the entire interface, described in section 4. Responses were compiled for each of the 12 participants in the study, along with any written comments that the participants had. The overall mean rating of all sections of the QUIS was 5.73, on a 7-point scale, and all questions were rated above 5. For the "Overall Reaction" section, "Satisfaction" overall rating was the only factor that was rated significantly lower than the mean response (*M* = 5.2, *SD* = 0.57), indicating that this area may need additional consideration. The "Ease of use" overall rating (*M* = 6, *SD* = 0.53) was significantly more than the mean response, indicating that users can easily learn and use the interface. From these "Overall Reactions" to the interface, we can conclude that users found our interface easy to use, helpful, useful, and flexible. In the "General Impressions" section, the overall ratings for "Screens are aesthetically pleasing", "Screen designs and layout are attractive", "Interface is impressive", "Interface can do a great deal", and "Interface is fun to use", were rated significantly better than the mean response. This indicates that users found our interface aesthetically pleasing and fun to use. The remaining items of this section were not significantly different from the mean response.

In addition to the QUIS item findings, participants were also asked to list three most negative and positive aspects of the interface. Some participants reported that "I found it weird to have a vertical list of video pieces", "Not being able to delete segments from history", and "Not being able to favorite some segments from history". The comment about the vertical scrolling coincided with their responses to the location of the component where the mean rating for the layout of History Timeline was significantly lower than that of the Filmstrip. We designed the component in this layout to eliminate the confusion between the History Timeline and the video timeline; this may need additional investigation. Participants found the interface helpful and impressive in being able to dynamically create points or bookmarks, which would allow them to skip to a favorite clip. Most participants stated that the interface is fun to use once you get the hang of it.

In order to explore whether users foresee some potentials for the interface we asked them "Where and how do they think video navigation history can be useful".

Participants provided us with valuable responses that gave us some insights of how this interface can be further modified and tested. They foresee that it would be useful in educational environment as well for home usage. There were some overwhelming comments made by the participants about the interface. One participant commented “I definitely see how this would be really helpful for long videos because I will not have to waste my time watching the whole video again to get to the important stuff. I could directly use my previous history to navigate to these intervals”. Others said, “I would love to see this implemented within social websites. I could see how I would use it and definitely I will have more fun”; “I need to have this. Could we have it in YouTube?”; and finally “I really do like this interface and I would love to have it to create wonderful clips from my home videos”.

Based on the participants’ valuable comments and suggestions, we believe using personal navigation history is helpful in navigating a video space. This could work for different applications such as highlights or summary of videos, a movie using home videos, sharing interesting clips, quickly navigating and skimming previously watched videos, and finally watching new videos’ interesting parts using crowd-sourced viewing histories. Receiving these positive feedbacks from the participants and how they welcomed the idea of using their personal navigation history motivate us to investigate how our interface would work if integrated with social websites (e.g. YouTube, Vimeo) and explore how this could change the way users consume and navigate video.

6 Conclusion and Future Work

The work we have presented describes a new way to navigate and view video media using a personal video history. We introduced a novel interface for video navigation through the creation and subsequent use of a user’s personal viewing history. The interface is based on observations of new use cases involving re-watching of within-video intervals and the increasingly temporal nature of video navigation. We performed a study that found positive results and highly positive affect and comments from participants. A comparative study based on a use case of fast navigation found significant results in favor of our method. To our knowledge, this is the first paper on using history navigation through video browsing, and as such, not enough work has been done in this area to evaluate the effectiveness on a larger scale. So we intend to investigate scalability, filtering mechanisms, interactions required when using a multiple-video history to allow users with large history to accurately navigate through them. Once these mechanisms have been tested, we aim to deploy a field study to check the validity and the scalability of history in action.

References

1. Bergman, L., Castelli, V., Lau, T., Oblinger, D.: Docwizards: a system for authoring follow-me documentation wizards. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, UIST 2005, pp. 191–200. ACM, New York (2005)

2. Card, S.K., Pirolli, P., Van Der Wege, M., Morrison, J.B., Reeder, R.W., Schraedley, P.K., Boshart, J.: Information scent as a driver of web behavior graphs: results of a protocol analysis method for web usability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2001, pp. 498–505. ACM, New York (2001)
3. Cheng, K.-Y., Luo, S.-J., Chen, B.-Y., Chu, H.-H.: Smartplayer: user-centric video fast-forwarding. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 789–798. ACM, New York (2009)
4. Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of SIGCHI 1988, ACM/SIGCHI, pp. 213–218 (1988)
5. Christel, M., Moraveji, N.: Finding the right shots: assessing usability and performance of a digital video library interface. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, MULTIMEDIA 2004, pp. 732–739. ACM, New York (2004)
6. Cockburn, A., Mckenzie, B.: What do web users do? an empirical analysis of web use. International Journal of Human-Computer Studies 54, 903–922 (2001)
7. Dragicevic, P., Ramos, G., Bibliowitcz, J., Nowrouzezahrai, D., Balakrishnan, R., Singh, K.: Video browsing by direct manipulation. In: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 237–246. ACM, New York (2008)
8. Drucker, S.M., Glatzer, A., De Mar, S., Wong, C.: Smartskip: consumer level browsing and skipping of digital video content. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, CHI 2002, pp. 219–226. ACM, New York (2002)
9. Foote, J., Boreczhy, J., Girgensohn, A., Wilcox, L.: An intelligent media browser using automatic multimodal analysis. In: Proceedings of the Sixth ACM International Conference on Multimedia, MULTIMEDIA 1998, pp. 375–380. ACM, New York (1998)
10. Gkonela, C., Chorianopoulos, K.: Videoskip: event detection in social web videos with an implicit user heuristic. Multimedia Tools and Applications, 1–14 (2012), 10.1007/s11042-012-1016-1
11. Goldman, D.B., Gonterman, C., Curless, B., Salesin, D., Seitz, S.M.: Video object annotation, navigation, and composition. In: Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST 2008, pp. 3–12. ACM, New York (2008)
12. Grossman, T., Matejka, J., Fitzmaurice, G.: Chronicle: capture, exploration, and playback of document workflow histories. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, pp. 143–152. ACM, New York (2010)
13. Guimbretiere,´ F., Dixon, M., Hinckley, K.: Experiscope: an analysis tool for interaction data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 1333–1342. ACM, New York (2007)
14. Heer, J., Mackinlay, J., Stolte, C., Agrawala, M.: Graphical histories for visualization: Supporting analysis, communication, and evaluation. IEEE Transactions on Visualization and Computer Graphics 14(6), 1189–1196 (2008)
15. Hupp, D., Miller, R.C.: Smart bookmarks: automatic retroactive macro recording on the web. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST 2007, pp. 81–90. ACM, New York (2007)

16. Karrer, T., Weiss, M., Lee, E., Borchers, J.: Dragon: a direct manipulation interface for frame-accurate in-scene video navigation. In: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 247–250. ACM, New York (2008)
17. Kimber, D., Dunnigan, T., Girgensohn, A., Shipman, F., Turner, T., Yang, T.: Trailblazing: Video playback control by direct object manipulation. In: 2007 IEEE International Conference on Multimedia and Expo, pp. 1015–1018 (July 2007)
18. Leftheriotis, I., Gkonela, C., Chorianopoulos, K.: Efficient video indexing on the web: A system that leverages user interactions with a video player. In: Alvarez, F., Costa, C. (eds.) UCMEDIA 2010. LNICST, vol. 60, pp. 123–131. Springer, Heidelberg (2012)
19. Li, I., Nichols, J., Lau, T., Drews, C., Cypher, A.: Here's what I did: sharing and reusing web activity with actionshot. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 723–732. ACM, New York (2010)
20. Mertens, R., Farzan, R., Brusilovsky, P.: Social navigation in web lectures. In: Proceedings of the Seventeenth Conference on Hypertext and Hypermedia, HYPERTEXT 2006, pp. 41–44. ACM, New York (2006)
21. Nakamura, T., Igarashi, T.: An application-independent system for visualizing user operation history. In: Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST 2008, pp. 23–32. ACM, New York (2008)
22. Pongnumkul, S., Wang, J., Cohen, M.: Creating map-based storyboards for browsing tour videos. In: Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST 2008, pp. 13–22. ACM, New York (2008)
23. Sawhney, N., Balcom, D., Smith, I.: Authoring and navigating video in space and time. IEEE Multimedia 4(4), 30–39 (1997)
24. Shamma, D.A., Shaw, R., Shafton, P.L., Liu, Y.: Watch what I watch: using community activity to understand content. In: Proceedings of the International Workshop on Multimedia Information Retrieval, MIR 2007, pp. 275–284. ACM, New York (2007)
25. Syeda-Mahmood, T., Ponceleon, D.: Learning video browsing behavior and its application in the generation of video previews. In: Proceedings of the Ninth ACM International Conference on Multimedia, MULTIMEDIA 2001, pp. 119–128. ACM, New York (2001)
26. Yeung, M.M., Yeo, B.-L.: Video visualization for compact presentation and fast browsing of pictorial content. IEEE Transactions on Circuits and Systems for Video Technology 7(5), 771–785 (1997)
27. Yew, J., Shamma, D.A., Churchill, E.F.: Knowing funny: genre perception and categorization in social video sharing. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, pp. 297–306. ACM, New York (2011)
28. Yu, B., Ma, W.-Y., Nahrstedt, K., Zhang, H.-J.: Video summarization based on user log enhanced link analysis. In: Proceedings of the Eleventh ACM International Conference on Multimedia, MULTIMEDIA 2003, pp. 382–391. ACM, New York (2003)

A New Approach to Walking in Place

Luís Bruno, João Pereira, and Joaquim Jorge

INESC-ID, Rua Alves Redol, 9 1000-029 Lisboa, Portugal
lbruno@ipbeja.pt, {jap,jaj}@inesc-id.pt

Abstract. Walking in Place (WIP) is an important locomotion technique used in virtual environments. This paper proposes a new approach to WIP, called Speed-Amplitude-Supported Walking-in-Place (SAS-WIP), which allows people, when walking along linear paths, to control their virtual speed based on footstep amplitude and speed metrics. We argue that our approach allows users to better control the virtual distance covered by the footsteps, achieve higher average speeds and experience less fatigue than when using state-of-the-art methods based on footstep frequency, called GUD-WIP.

An in-depth user evaluation with twenty participants compared our approach to GUD-WIP on common travel tasks over a range of short, medium and long distances. We measured task performance using four distinct criteria: effectiveness, precision, efficiency and speed. The results show that SAS-WIP is both more efficient and faster than GUD-WIP when walking long distances while being more effective and precise over short distances. When asked their opinion via a post-test questionnaire, participants preferred SAS-WIP to GUD-WIP and reported experiencing less fatigue, having more fun and having a greater level of control when using our approach.

Keywords: Walking in place, virtual locomotion, virtual speed control, performance, motor control.

1 Introduction

Virtual locomotion is the most common task performed by users in virtual environments (VEs). Although in many cases virtual locomotion is not the primary user objective [12], the motion supports numerous other tasks, such as manipulating objects, collecting information [1] and surveying virtual environments.

Navigation control can rely on either indirect interaction [17] using physical controls (buttons or joystick) or direct interaction based on gestures expressed by the motion of body parts. Physical locomotion [12] is a metaphor that groups different locomotion techniques, such as real walking and Walking in Place (WIP). WIP is a popular technique because it enables people to control virtual locomotion, thus mimicking real walking movements [6]. Unlike unrestricted real walking, the WIP technique allows users to operate in small areas of interaction. This technique also frees the subject's hands, making them available to perform other interaction tasks. One usage scenario is the architectural review task where reviewer moves through buildings or outside spaces using a WIP technique and points and writes notes using hands.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 370–387, 2013.

Different WIP systems have been devised with different inputs, outputs, control laws of the virtual displacement and simulated movements [6]. Different technologies have been used to detect steps in place: magnetic trackers [2,4,5], force sensors placed on shoe insoles [5], optical camera trackers [8], Wiimote Nintendo™ accelerometers [14] and Wii Balance Boards [13]. Several different body segment motions are tracked to generate virtual output, including the head [4], knees [5,10] and shins [2,8]. These evolutions in user input have enabled the improvement of footstep latency times (starting and stopping travel) [2,10] and have assured the continuity and smoothness of the movement between and within steps [2,8]. In addition, different types of virtual locomotion control laws are based on neural networks [2], knee pattern recognition [5], signal processing [2] and biomechanical state machines [8]. In reference to viewpoint movement, the Slater WIP system [4] updates a predefined virtual distance only when a footstep is detected. This type of displacement motion is discrete and confusing for users. Later systems [2,8] updated the viewpoint in each frame and attempted to simulate a sinusoidal forward velocity curve.

GUD-WIP [8] is a state-of-the-art WIP method that estimates walking speed based on the subject's height and mainly on the subject's step frequency. These metrics emanate from the gaiting biomechanics literature [15], which indicates that are positively correlated with footstep horizontal length [16].

However the physical movements of gaiting in place are slightly different. The feet movements are predominantly vertical and the time to perform footsteps can be different to the horizontal footsteps. To control speed using the step frequency users should manage the time variable. Due to different biomechanics features the minimum time to perform vertical footsteps can be lower than to the horizontal footsteps. So, in the WIP method user can achieve unexpected higher speed values which cause difficulties to control forward walking.

Other important metric derived from WIP biomechanics features is the footstep amplitude. This metric describes the footstep vertical length and is defined as the distance between the foot off (and foot strike) positions and the maximum foot height achieved. With this metric user manages the vertical distance covered by the foot to control virtual speed. We argue that user manages better this distance metric than the time metric to control virtual speed. So, our hypothesis states that is more affordable and perceptible to user control the virtual speed using the amplitude than the frequency metric in common walking tasks.

We developed an exploratory study to understand how subjects establish a relationship between step vertical length and three simulated virtual rates of speed (i.e., slow, moderate and fast). Based on the quantitative results, we propose a new approach to walking in place, called SAS-WIP (Speed-Amplitude-Supported Walking-in-Place), which models the forward linear virtual speed as a function of two kinematic metrics: footstep amplitude and speed. Our approach follows the main goals of the recent WIP systems and adds a simple and perceptible technique for transforming step amplitude as a primary factor to generate virtual speed.

Few studies have attempted to compare different WIP approaches and other interfaces [6]. Furthermore, many evaluations resort to subjective questionnaires, from which results are difficult to reproduce. To test our hypothesis we developed a comparative study of our SAS-WIP approach and GUD-WIP [8]. We measured the task performance of twenty participants in travel and stopping precision tasks through three different distances (i.e., short, medium and long). We assessed the user performance according different criteria: effectiveness, precision, speed and efficiency.

We argue that SAS-WIP outperforms GUD-WIP with the following features:

- Provides to users a better forecast of the virtual distance covered by a single step.
- Allows for higher average speeds to be achieved, especially when covering long distances.
- Requires less effort from users, resulting in less fatigue, especially when covering long distances.

The results show that the SAS-WIP interface exhibited greater effectiveness and precision than the GUD-WIP interface in stopping precision tasks for short distances. Additionally, SAS-WIP was both more efficient and faster than GUD-WIP for long distances.

The remainder of this paper is organized as follows. Related work is presented on the next section. Next, we describe the SAS-WIP approach, followed by the experiment description, the results and the discussion of the main outcomes. In the conclusion section, we introduce the guidelines and summarize our work.

2 Related Work

Virtual locomotion is a common task involved in controlling virtual environments. This motion can either be the primary task or a secondary goal that allows people to perform specific tasks, such as manipulating objects or collecting information [1]. Different interaction techniques have been developed to support locomotion in VEs, and these techniques are typically used in complex navigation, including changes in direction and speed [6].

The WIP technique, which is considered a physical locomotion interface [12], was designed to improve the sensation of walking while providing efficient navigation in VEs [4]. This technique allows users to use small areas of interaction in the real world, unlike real walking, while freeing their hands for other interaction tasks. To this end, the users' bodies are tracked and analyzed to simulate virtual locomotion in the VE. The movements of different body parts are detected to steer virtual displacement via distinct approaches.

Slater developed the first WIP system for controlling virtual locomotion in VEs [4]. This approach used a neural network to recognize patterns of head movement that correspond to walking-in-place actions. The starting and stopping step detection was not extremely precise, which confused users and caused them to overshoot their target stopping location. Moreover, the forward viewpoint movement is updated only one time for each step by one predefined length. User testing with this system revealed that WIP can increase presence compared to classical joystick-based interaction [7], which is similar to the conclusion reached by Razzaque et al. [9]. In that study, it was suggested that WIP might also increase cybersickness [7].

The Gaiter system [5] for military simulators is dependent on leg movement, and WIP is treated as a gesture that indicates that the user intends to take a virtual step. Gaiter uses the direction and the extent of knee movement to compute an implicit horizontal displacement of a body in the VE. However, to recognize a virtual step, the system must wait until the knee reaches the point of maximum extent, causing a half-step latency [2].

Another WIP system developed by Yan et al. [10] used the speed at which the knee lifts in stepping in place actions to determine the virtual locomotion speed. The authors related different variables of the real locomotion (forward velocity/leg lifting speed, leg lifting speed/step frequency) and the stepping in place (leg frequency/leg lifting speed). The authors report that this system is more responsive than Gaiter system without explicit proof.

The output speed of the LLCM-WIP [2] system is dependent on a continuous signal of the foot heel position. This approach accomplishes the following goals: low starting and stopping latency, smooth locomotion between steps, continuous control of locomotion speed within each step and the incorporation of real-world turning and short-distance maneuvering into virtual locomotion [8]. The output speed curve varies considerably when the position differentiation is 0 when maximum foot height is achieved and between steps.

The state of the art GUD-WIP model creates output speeds that better match those occurring during real walking and that better respond to variations in step frequency, including realistic starting and stopping control of virtual speed using the observed footstep frequency [8]. The virtual speed of this implementation shows considerably less within-step fluctuation than LLCM-WIP. The virtual speed is dependent on a walking equation from the biomechanics literature, which relates to step frequency and subject height and relies on in-place step events. The user-study analysis presented that real walking and GUD-WIP each provide a consistent step frequency to resulting speed.

Although Wendt et al. hypothesized that consistency improves the user experience, which leads to improved usability [8], we are not certain that users correctly interpret step frequency to forecast the virtual distance covered by a single step in precision targeting tasks (short distances). For certain types of applications, such as traveling long distances in VEs at high speed, step frequency usage could be inefficient and tiring. The stopping latency of the GUD WIP may also be unacceptable for certain applications, as mentioned by the authors.

The system of Wendt et al. was not evaluated by participants during VE travel tasks for comparison with other WIP systems or other navigation techniques to understand the task performance addressed and the user experience.

In this paper, we present a participant study to evaluate our SAS-WIP approach in comparison with GUD-WIP. The participants executed travel and target stopping tasks, and the performance is measured based on quantitative metrics. The user experience is articulated with a subjective questionnaire. In this manner, it is be possible to understand the usability provided by each system related to those criteria.

3 Our Approach

To test our hypothesis we developed an exploratory study to understand how subjects establish a relationship between step vertical length and three simulated virtual rates of speed (i.e., slow, moderate and fast). The quantitative results showed a correlation between the two variables; higher footsteps correspond to higher simulated speed. Another feature emerged during this analysis: foot speed. Subjects also related the simulated virtual speed to different levels of foot swing speed. This relationship also proved to be consistent.

Based on these results we develop the SAS-WIP model which primarily takes advantage of the foot height achieved during in-place steps and is complementary to the foot vertical speed. Our system addresses requirements similar to LLCM-WIP [2] and GUD-WIP [8] as smooth between-step locomotion speed, continuous within-step speed control, real-world turning and maneuvering, low starting latency, low stopping latency, in-place step events detection and biomechanics-inspired state machine.

The next sections describe the development process, the features and the model algorithm of the SAS-WIP technique.

3.1 Exploratory Study

In the early stages of our research, we conducted informal observations of people walking in place. We asked subjects to perform sequences of in-place steps with the intention to represent different virtual speed levels (i.e., slow, moderate and fast). We only requested that participants relate foot height during steps to each virtual rate of speed.

Most subjects exhibited common behavior patterns. When asked to simulate slow locomotion, subjects raised their feet to the lowest maximum height at slow foot speed. For moderate locomotion, the footsteps were raised higher and faster than for slow walking. Maximum height and foot speed were observed when we instructed participants to walk briskly. These observations allowed us to identify a subject-consistent pattern in the use of foot height to control virtual speed, which suggests that people perceive that this metric controls different levels of speed. Another metric in subject behavior was also revealed: vertical foot speed. To achieve faster virtual speed, the subjects raised the feet faster. Therefore, we considered it important to study this feature in greater detail.

We conducted a formal exploratory study to quantitatively confirm the results of those observations. We aimed to confirm that subjects consistently relate footstep amplitude and speed with the intended virtual speed. Five participants, four male and one female, aged 22 to 45, took part in the user study and were recruited from our local university campus. None of the participants had motor or physical impairments. The experiment was conducted in a laboratory equipped with an optical tracker system [11] based on infra-red cameras, providing 6DOF tracking data at 100 Hz. To record the foot height and speed, one reflective marker was mounted on each heel. The heel position data for each foot is received as a continuous signal and allows for the calculation of the foot speed by vertical position differentiation.

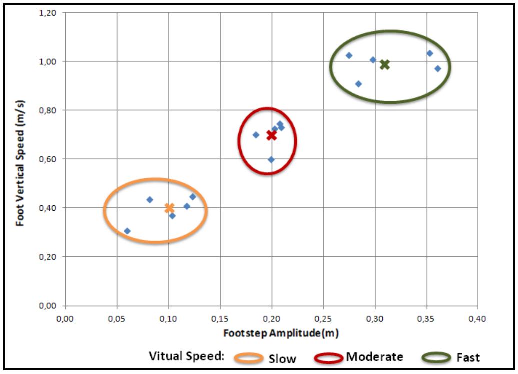


Fig. 1. The footstep amplitude and vertical clustered for the five participants

All participants were asked to walk in place inside a 3x3 meter square. At the beginning of the experiment, the participants were informed of the purpose of the experiment and were asked to perform three series of ten in-place steps for each level of simulated virtual locomotion speed: slow, moderate and fast. The position and timestamp data gathered allowed for the analysis of descriptive footstep amplitude and speed statistics and for the determination of the consistency between those metrics and the virtual speed levels perceived by the participants. As all the samples presented a normal distribution then we used an ANOVA to assess differences between the three speed levels on the two metrics within a significance of 0.05.

A one-way repeated-measures ANOVA revealed significant differences among the three speed levels ($F2 = 65.071$, $p<0.001$) as a function of footstep amplitude sample. A Bonferroni post-hoc test revealed significant differences between all speed levels: slow and moderate ($p=0.009$), slow and fast ($p=0.001$), and moderate and fast ($p=0.023$).

These results suggest that participants expected that virtual speed changes consistently with the footstep amplitude. Fig. 1 shows that amplitude values are consistent with each simulated rate of speed.

We use mean speed to characterize foot motion during one step. This calculus is based on the vertical distance covered by the foot in the descendent and ascendant phases and the respective time periods as defined in Equation (1). The vertical distance value is the double of the amplitude value, *ta i*s the ascendant time and *td* is the descendant time. The amplitude is the distance between the foot-off (and foot strike) detected position and the maximum foot height position.

$$Vel = 2*amplitude/(ta+td) \quad (1)$$

A one-way repeated-measures ANOVA revealed significant differences among the three speed levels ($F_2 = 154.981$, $p<0.001$) as a function of footstep mean speed. A Bonferroni post-hoc test revealed differences between all cadences: slow and moderate ($p=0.003$), slow and fast ($p<0.001$), and moderate and fast ($p=0.008$).

These results show that participants naturally and consistently adapted their footstep speed to match each requested simulated speed. Fig. 1 confirms these results by demonstrating the variation of the vertical mean speed values of the five participants for each of the three simulated speed levels.

Table 1. Descriptive stats of kinematic variables

Virtual Speed	Amplitude (m)				Mean Foot Speed (m/s)			
	Min.	Mean	St. Dev.	Max	Min.	Mean	St. Dev.	Max
Slow	0.05	0.10	0.03	0.14	0.24	0.40	0.08	0.59
Moderate	0.08	0.20	0.03	0.30	0.38	0.70	0.09	0.84
Fast	0.22	0.31	0.06	0.46	0.68	0.99	0.09	1.13

Table 1 shows the descriptive statistics of the footstep amplitude and mean speed metrics in greater detail. One can see that the amplitude values related to slow, moderate and fast speeds are grouped at approximately 0.10 m, 0.20 m and 0.31 m, respectively. Regarding the mean foot speed, their values are grouped at approximately 0.40 m/s, 0.70 m/s and 0.99 m/s for the same speed levels, respectively.

Based on these results, we propose to compute the linear virtual speed in our SAS-WIP metaphor using the observed kinematics variables and their value ranges. The next section describes the primary decisions concerning SAS-WIP.

3.2 SAS-WIP Model Parameters

Our approach uses a biomechanics state machine with three states: ascending, descending and foot support. The main events detected are foot-off, foot max height and foot strike. Fig. 2 shows the foot states and the conditions that trigger each event. The start event (foot off) occurs when the foot vertical position (Pvert) exceeds 0.035 m (P1) and the vertical velocity (Vvert) exceeds 1.0 m/s (V1). The transition from ascending to descending state occurs when the maximum height is achieved (i.e., Vvert is less than 0.05 ms (V2)). The foot strike occurs when Pvert is lower than 0.035 m.

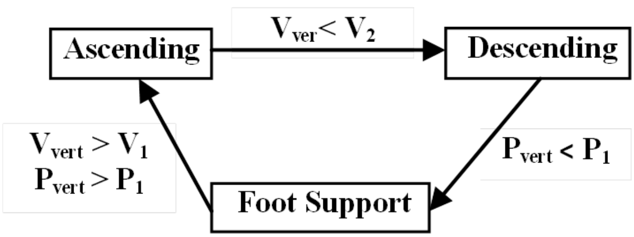


Fig. 2. The state transition diagram for an in-place step for a single foot

In the WIP starting phase, the virtual speed of the first step is updated at three moments, when the following events occur: foot off, max height and foot strike. Therefore, when the foot off of the first step is detected, the system assigns a value of 0.24 m/s to the initial virtual speed, which is the minimum value corresponding to the slow simulated speed achieved in the exploratory speed (see Table 1).

In rhythmic phase, the speed is updated at two moments for each step cycle during the following events: max height and foot strike. The first event occurs at the end of the ascending phase, and the second event occurs at the end of the descending phase. Only for these two moments, the speed is updated based on Equation (2), allowing for the maintenance of steady speed within and between steps.

Equation (2) to compute the virtual speed is supported by the kinematics results of the exploratory study. We separate the calculus into two conditions: (i) subjects that intend to achieve a slow speed or (ii) subjects that want faster speeds (moderate or fast).

To achieve slow virtual speed, the footstep amplitude should be lower than 0.13 m. This value corresponds to the sum of the foot mean amplitude (0.10 m) and the standard deviation (0.03 m) of the slow simulated speed (see Table 1). We observed that users control slow speeds better using only their footstep mean speed.

To achieve faster virtual speed, the main factor is footstep amplitude, which should be higher than the slow speed threshold (0.13 m). In this case, the amplitude is divided by that threshold to obtain a linear scale factor of the virtual speed. This decision was made according to our concept that users expect that virtual speed changes consistently with the step amplitude. To calculate the virtual speed, the amplitude scale is multiplied by the foot speed. This metric assures consistency in the updating of the virtual speed and allows users to allocate or not allocate more walking effort or energy.

$$\textbf{Speed} = \begin{cases} \text{footSpeed}, & \text{if } Amplitude \leq 0.13 \\ \text{footSpeed} * \dfrac{\text{Amplitude}}{0.13}, & \text{if } Amplitude > 0.13 \end{cases} \tag{2}$$

Accurate stopping is important to users to better predict how to reach the desired locations. Two types of problems exist: false positives (the system stops without user intention) or stops long after the user orders a stop. In the stopping movement phase, we attempted to avoid these problems in the SAS-WIP system. The motion stops when one timeout occurs at the double support (when the feet are supported on the ground) state as described by Equation (3). This value is computed as a function of the mean foot speed of the last descending phase before the double support. This formula was computed from data gathered from the exploratory study. We investigated how long the double support was sustained after different footstep speed levels (from slowest to fastest) occurred.

$$\text{Time} = (-0.2 * \text{foot\_speed} + 0.41) * 1000 \tag{3}$$

We identified a trend in the foot speed variation from the slowest to the fastest motions before the user stop actions. When the foot speed is higher, the timeout is shorter. We opted to stop the virtual movement immediately after the timeout is achieved. GUD-WIP slows and smoothly stops the virtual speed. Our method decreases the stopping latency relatively to GUD-WIP, preventing unintentional motion stoppage.

3.3 Speed Features

Fig. 3 (right image) shows the virtual speed curve generated by the SAS-WIP algorithm from two independent series of in-place footsteps (blue and red lines). The first three steps are slows, the following three steps are moderate and the three final steps are fast. The "Virtual Speed" axis discriminates the output speed generated by SAS-WIP, and the "Position" axis represents the foot height achieved by each foot (left and right).

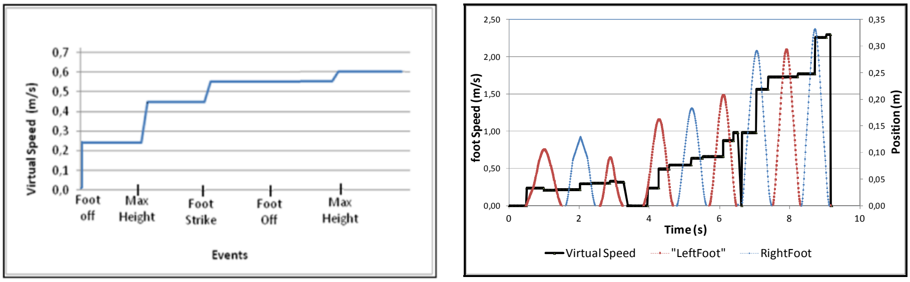


Fig. 3. Virtual speed is updated at the starting and rhythmic phases (left image); virtual speed curve generated by SAS-WIP (right image)

In Fig. 3 (left image), it is possible to identify several features, such as the assignment of the initial speed (0.24 m/s) at foot off of the first step as well as the speed updates at the max height and foot strike events. Regarding the rhythmic phase of the slow steps, the curve shows that the output speed is steady because the foot speeds are similar. Regarding the moderate steps, the effect of amplitude is evident. The virtual speed curve grows consistently with footstep amplitude. The speed control is continuous and responsive to amplitude changes between- and within-step.

The stopping latency is exhibited in the last step of each sequence. There is an evident delay between the end of the last step and the virtual stopping. However, this timeout is detected immediately and is assigned a null speed in the virtual movement.

4 Experiment

To assess the performance afforded by SAS-WIP, we developed a participant study to compare our approach with GUD-WIP.

4.1 Tasks

The participants were asked to execute two different tasks: (i) travel through nine paths of different distances and (ii) stop in front of nine prescribed target locations. The sequence and the distances of the nine paths traveled are as follows: P1 (6.0 m), P2 (1.0 m), P3 (12.0 m), P4 (0.5 m), P5 (24.0 m), P6 (1.5 m), P7 (3.0 m), P8 (48.0 m) and P9 (2.0 m). These different paths covered different distances: short (0.5 m, 1.0 m, 1.5 m, 2.0 m and 3.0 m), medium (6.0 m and 12.0 m) and long (24.0 m and 48.0 m). The short distances are suited to test the precision and the effectiveness provided by the interfaces in the stopping task, and the medium and long distances can check the efficiency and the speed offered to participants by the locomotion interfaces.

In the stopping task, the participants attempt to stop as close as possible in front of a target without overshooting it. This task occurred at the end of the travel task. Each of the nine paths is bound by rectangular targets (length: 1 m, width: 0.1 m) positioned on the sidewalk floor (length: 100 m), as exhibited in Fig. 4. The participants control the viewpoint in this experiment by using the first person model and perceive the proximity to the target by measuring the distance between that object position and the bottom line of the viewport. Therefore, the participants should stop the viewpoint motion when the rectangle is as near as possible to the bottom line of the screen.

The task performance offered to participants by the interfaces is determined by the following criteria: effectiveness, precision, mean speed and efficiency.

4.2 Participants

Twenty participants, 15 male and five female, ranging in age from 20 to 33 years (mean=24.5 years) took part in the user study and were recruited from our local university campus. The majority of the participants had played videogames that use virtual navigation, and no one had motor or physical impairments. No participant had previously tested physical locomotion interfaces like the WIP systems.

4.3 Procedure

At the beginning of the experiment, the participants were told the purpose of the study, and the script was explained. A pre-test questionnaire was administered to gather demographic and navigation skills data for each participant.

Before testing each interface, participants practiced for at least 10 minutes in an environment similar to the formal tests, performing travel and stopping tasks. When ready, the participants began each interface trial, traveling all the nine paths and making stops before each target. When a system error occurred or a participant sensed a particular difficulty, the trial was repeated. Regarding the travel task, the participants were asked to complete each assigned path as fast as they could. At the end of each interface, the trial subjects completed a post-test questionnaire to respond to questions about their experience. Each participant experimented with both interfaces in an order determined using a Latin Square approach to avoid bias associated with the sequence of the trials.

After all the participant trials were complete, a post-experiment questionnaire was administered in order to allow the participants to compare and rank the interfaces based on their subjective preferences.

4.4 Apparatus

This experiment was developed in the same laboratory where the exploratory study occurred. The output of the virtual environment was displayed on a large-scale screen (4x2.25 square meters) as shown in Fig. 4 (bottom picture). The participants are positioned approximately 2.0 m from the screen, within a 3x3 meter square area. The foot motion capture was supported by an optical tracker system [11], which determines the heel position for each foot.


Fig. 4. The 3D virtual environment and the experiment apparatus (from left to right)

4.5 Ensure Equity with GUD-WIP

To ensure a fair comparison between our approach and GUD-WIP, we implemented several adjustments to this system. Because the GUD-WIP stopping latency is slow, we adapted the SAS-WIP stopping mechanism, which described by Equation (3). Therefore, the effectiveness and precision of the two interfaces were only affected by the interface features provided to the participants.

To compare the speed provided by each interface, it is necessary to assure that their value ranges are similar. Theoretically, the two WIP interfaces allow similar speed levels for the slowest and fastest cases, as exhibited in Table 2.

Table 2. Comparison of 3 representative speed levels generated by SAS-WIP and GUD-WIP

GUD-WIP		SAS-WIP		
Freq. (hz)	**Virtual Speed(m/s)**	**Ampl.(m)**	**Foot Speed (m/s)**	**Virtual Speed (m/s)**
2.50	2.54	0.31	0.99	2.36
1.75	1.24	0.20	0.70	1.07
1.00	0.41	0.10	0.40	0.40

Table 2 compares three representative speed levels for each interface, corresponding to slow, medium and fast values. In this comparison, the GUD-WIP virtual speed is computed with a constant height of 1.72m. The variable values of SAS-WIP are obtained from the mean values of Table 1 for the three speed levels. The GUD-WIP frequency values are based on reference values from the literature [15] that are representative of the range speed. The comparative virtual speed values suggest that the two interfaces can provide similar speed levels. Therefore, the speed achieved by the participants should depend on the affordance provided by each interface.

4.6 Design and Data Analysis

We used a within-subjects design where each participant tested all the three interfaces. The nine paths were grouped in three different distance classes: short, medium and long, as explained previously. Therefore, the analyzed data consisted of 20 participants x 3 distance classes x 3 interfaces, which equals 180 paths/targets.

The quantitative results are dependent on the following task metrics: target overshoots frequency, final distance to target, mean speed and number of steps. The data samples of the overshoot frequency and the final distance to target did not present a normal distribution, so we used the non-parametric Wilcoxon test to determine the pairwise differences among the interfaces.

Regarding the "mean speed" and "number of steps" metrics, a two-way repeated-measures ANOVA (interfaces x paths) was applied. In this case, the Greenhouse-Geisser's sphericity corrections were applied whenever Mauchly's test of sphericity showed a significant effect. In both metrics, the paired t-test was used to determine the pairwise differences among the two interfaces.

The subjective questionnaire results were analyzed using the Friedman non-parametric test on the different six criteria (i.e., Fun, Ease of use, Fatigue, Precision, Naturalness and Global Appreciation) and the Wilcoxon post-hoc test.

The level of significance used for all of the hypothesis tests was α=0.05.

5 Results and Discussion

The quantitative data gathered from the trials were statically analyzed to understand the comparative performance provided by the two interfaces. In this section, we explain the effects of the interfaces on the metrics, which are the overshoot frequency, final distance to target, mean speed and the number of steps.

Table 3. Mean and interval confidence values of the metrics: A-Frequency Overshoots, B- Distance to Target (m), C-Mean Speed (m/s) and D-Number of Steps

	SAS-WIP						GUD-WIP					
	Short		Med.		Long		Short		Med.		Long	
	μ	CI	μ	CI	μ	CI	μ	CI	μ	CI	μ	CI
A	0.58	0.25	0.48	0.21	0.48	0.25	0.80	0.35	0.65	0.28	0.60	0.26
B	0.05	0.02	0.08	0.03	0.12	0.12	0.14	0.14	0.14	0.14	0.14	0.14
C	0.50	0.06	1.21	0.23	2.64	0.39	0.50	0.07	0.88	0.11	1.93	0.23
D	4.63	0.62	11.6	1.87	19.3	1.87	3.73	0.49	14.2	1.53	37.0	4.75

In this section, results obtained from the subjective questionnaire post-test are also described.

5.1 Effectiveness

The target overshoot data were recorded during the stopping tasks. One overshoot represented an error by the participant. The performance effectiveness criterion is related to the overshoot frequency. Fewer overshoots indicates greater effectiveness.

Table 3 shows the mean overshoot frequency for the two interfaces grouped by distance classes. The SAS-WIP interface allows, on average, a lower overshoot frequency than GUD-WIP for the three distance classes.

Relative to short distances, the test showed significant differences between SAS-WIP and GUD-WIP (p=0.004). For medium and long distances, significant differences between SAS-WIP and GUD-WIP (p=0.080, p=0.856, respectively) were not found. We contend that short distances are suited to assess the effectiveness provided by the interfaces because users have less time and space to forecast the virtual distance covered by each footstep, in contrast with longer distances. We argue that smaller distances allow for better understanding of which interface provides the best effectiveness control. These arguments and the results suggest that SAS-WIP provides more effectiveness to users than GUD-WIP. In contrast, for medium and long distances, SAS-WIP and GUD-WIP assured similar effectiveness in those tasks. In these cases, the distances do not affect the interface results because participants had greater time and space to predict and adjust the distances covered to stop at the desired point.

5.2 Precision

During the stopping tasks, the participants were instructed to stop as near as possible to the target without overshooting the target. From the trials, the distance between the stop position and the target position was recorded for all targets that were not exceeded. The exceeded cases are analyzed on frequency overshoot metric. This metric was named “final distance to the target” and determines the precision level that an interface provides to participants. A smaller final distance to the target indicates greater precision.

Table 3 shows the mean final distance to the target provided by the interfaces grouped by the three distance classes. On average, the SAS-WIP interface provided the lowest final distance to the target. Regarding short and medium distances, the test showed significant differences between GUD-WIP and SAS-WIP (p=0.002, p=0.015) in contrast to the long distances, where differences between the two interfaces were not revealed (p=0.466).

Similarly to effectiveness, we consider that smaller distances allow for better understanding of which interface provides the finest action control and consequently the best precision. From these arguments and from the short distance results, we can conclude that SAS-WIP provides more precision to users than GUD-WIP in stopping tasks. This last conclusion was complemented by the medium distance results. The long distance precision results were consistent with the long distance effectiveness results.

5.3 Mean Speed

The time and the distance traveled across the different paths were recorded to calculate the mean speed for each path. Greater mean speed indicates faster participant travel. This metric provides how fast the user can walk using each interface.

Table 3 shows the averages of the mean speed provided by the interfaces for the three distance classes. For medium and long distances, on average, SAS-WIP is faster than GUD-WIP, although for short distances, the two WIP interfaces are similar.

The two-way repeated-measures ANOVA revealed the significant effect of the interface ($F_{1.0,16.0}$ =8.067, p=0.012) and of the distance ($F_{2.0,32.0}$=216.722, p<0.001) on the mean speed. Interactions between interfaces and distance classes were found ($F_{2.0,32}$ =7.600, p=0.002).

For short distances, the paired t-test did not reveal differences between the two interfaces regarding the mean speed (t_{17} =-0.148, p=0.884). Concerning medium and long distances, the paired t-test revealed significant differences between the interfaces on the mean speed (t_{17} =3.104, p=0.006; t_{16} =2.805, p=0.013, respectively).

For medium and long distances, the SAS-WIP approach induces participants to travel faster than GUD-WIP, but for short distances, the two WIP interfaces are similar. The speed in performing paths is more important and critical to longer distances. In this manner, the results show that SAS-WIP provides to participants a higher average speed than GUD-WIP. This conclusion suggests that participants more consistently map the virtual speed values to the footstep amplitude than to the footstep frequency.

The tests allow for the identification of several errors committed by participants when they want to achieve faster speeds with GUD-WIP. When the step frequency is high (e.g., 20 Hz) because the footstep period is very short (e.g., 50 ms), the speed achieved is “supersonic” (e.g., 169.91 m/s). This problem occurs when participants increase the step cadence, shortening times between footsteps to travel faster (sometimes the viewpoint moves out of the VE scenario).

5.4 Efficiency

No references were found in the gaiting literature about what is the footstep factor which causes more energy expenditure: frequency or amplitude. So, we chose the number of steps as the metric that indicates the level of effort allocated to this task because is the common element for the two interfaces. So in our concept, this metric provides a measure of efficiency provided by the WIP interfaces. Fewer steps to travel the paths indicate greater efficiency provided by the interface. During the travel task across the different distances, the number of footsteps performed was recorded.

Table 3 shows the mean number of footsteps provided by the interfaces grouped by distance classes. For the medium and long distances, on average, the GUD-WIP interface induced a greater number of steps than did SAS-WIP. For short distances, GUD-WIP provided a lower number of steps than did SAS-WIP.

The two-way repeated-measures ANOVA revealed the significant effect of the interface ($F_{1.0,14.0}$=38.522, p<0.001) and the distance ($F_{2.0,28.0}$ =160.112, p<0.001) on the number of footsteps performed. Interactions between interfaces and distance classes were identified ($F_{1.103,\ 15.446}$ =32.855, p<0.001).

For the two WIP interfaces, regardless of the distances, the effect of the interface was significant ($F_{1.0}$ =38.522, p<0.001). The post-hoc test revealed significant differences between SAS-WIP and GUD-WIP (p<0.001). In contrast, accounting for the three distance classes, the distance had a significant effect ($F_{1.502}$ =160.112, p<0.001) on the number of steps, as we expected. The post-hoc test revealed significant differences between SAS-WIP and GUD-WIP (p<0.001). For short distances, the paired t-test revealed significant differences between the two interfaces (t_{14} =3.450, p=0.004) regarding the number of steps. Concerning medium distances, the paired t-test revealed a significant difference between the two interfaces (t_{14} =-2.419, p=0.030) regarding the number of steps. For long distances, the paired t-test revealed a significant difference between the interfaces (t_{14} =-6.095, p<0.001).

For short distances, GUD-WIP is more efficient than the SAS-WIP approach. When the user intends to move slower with GUD-WIP, the footsteps should be slower so the user performs fewer footsteps during the displacement. In contrast, with SAS-WIP, the user should raise the foot to a short height to move slower, so the footsteps are faster, allowing a greater number of steps.

However, for medium and long distances, where the efficiency is a more important performance criterion, SAS-WIP requires fewer footsteps than GUD-WIP. The results suggest that SAS-WIP requires less effort from participants and causes less fatigue than GUD-WIP, especially when covering long distances, as we expected.

5.5 User Subjective Perception

After the participants finished the experiment, they completed a questionnaire in which they had to rank the interfaces (SAS-WIP and GUD-WIP) by their preference in the context of six subjective criteria: (a) Fun, (b) Easiness of Use, (c) Fatigue, (d) Precision of Control, (e) Naturalness and (f) Global Appreciation.

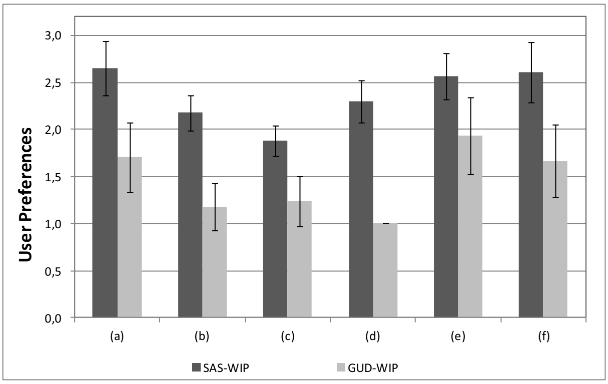


Fig. 5. The results of the subjective questionnaire for each criterion (a) – Fun, (b) – Ease of Use, (c) – Fatigue, (d) – Precision, (e) – Naturalness and (f) – Global Appreciation

To rank the interfaces for each criterion, the participants provided scores from 1 (worst preference) to 3 (best preference). Fig. 5 shows the preference results for each subjective criterion. In the fatigue case, grade 3 indicates that the interface provides the least fatigue.

We found significant differences between the SAS-WIP and GUD-WIP for the following criterion: Fun ($p=0.012$), Ease of Use ($p<0.001$), Fatigue ($p=0.002$), Precision ($p<0.001$) and Global Appreciation ($p=0.02$). Only for Naturalness, we didn't found significant difference between the two WIP interfaces.

The questionnaire results showed that SAS-WIP was ranked as the most appreciated interface (global appreciation) and most fun. Many participants commented during the experiment that they were greatly satisfied with the controllability provided by SAS-WIP and that they found it fun to move in the VE induced by its physical foot movements. SAS-WIP is perceived as less tiring than GUD-WIP.

The last result is consistent with the efficiency and the mean speed results achieved in this paper, which suggest that SAS-WIP induces less effort in participants and less fatigue than GUD-WIP. Several participants explicitly commented that SAS-WIP is less tiring than GUD-WIP, particularly for long distances.

6 Conclusions and Future Work

This paper introduced SAS-WIP (Speed-Amplitude-Supported Walking-in-Place), which is a new approach to WIP that allows for the control of linear virtual locomotion based on footstep amplitude and speed. We identified the features and kinematic variables that best describe the specific foot motions of the SAS-WIP approach via an exploratory study with five participants.

This paper provides a comparative study between two WIP alternatives concepts dependent on different metrics. The study explores common locomotion tasks, such as travel between targets and stopping before objects, and allows for the measurement of the performance provided and the user experience offered.

Table 4. Comparing the performance criteria provided by the interfaces ("~": Not Different, ">": Better than, "<": Worse than)

	Short	**Medium**	**Long**
Effectiveness	SAS > GUD	SAS ~ GUD	SAS ~ GUD
Precision	SAS > GUD	SAS > GUD	SAS ~ GUD
Speed	~ GUD-WIP	SAS > GUD	SAS > GUD
Efficiency	SAS < GUD	SAS > GUD	SAS > GUD

We argued that users can better forecast the virtual distance covered by a single step dependent on the footstep amplitude and speed than on the footstep frequency. The effectiveness and precision performance results obtained in our experiment for the stooping task suggest the confirmation of this statement (see Table 4). Even for medium and long distances, the speed control of one step is more perceptible for subjects (see Table 4).

Although theoretically, the speed equation dependent on the footstep frequency allows for greater speed, the results suggest that the approach dependent primarily on the footstep amplitude achieved a greater mean speed, particularly for medium and long distances.

Supported by the performance results and participant opinions, the effort allocated by users to cover medium and long distances using SAS-WIP seems to be smaller than that provided by those using GUD-WIP. This result suggests that SAS-WIP is more efficient, assuring less fatigue and possibly less energy expenditure.

The SAS-WIP system seems to be a promising WIP method, assuring controllable and comfortable interaction and good user perception, and should be the object of improvements and evaluations with other tasks and environments. However, our WIP approach has one open question related to the effect of the subject height in the virtual speed computed. One participant expressed this concern because of her small stature. We plan to develop a detailed study to understand this effect and how this personalized metric can be implemented in the motion control law.

A challenging task to measure the controllability provided by SAS-WIP is the pursuit of moving objects. We plan to develop a study for measuring the level of proximity to objects provided by our approach in different locomotion contexts.

In the future, we plan to evaluate the effect of the proprioception afforded by our approach on the perception of spatial dimensions in estimating distances in virtual environments. Additionally, we intend to add locomotion direction control (steering) dependent entirely on foot motions to SAS-WIP.

Acknowledgements. This work was supported by FCT (INESC-ID multiannual funding) under the project PEst-OE/EEI/LA0021/2011 and by the project CEDAR PTDC/EIA-EIA/116070/2009.

References

1. Chittaro, L., Venkataraman, S.: Navigation aids for multi-floor virtual buildings: A comparative evaluation of two approaches. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology 2006, pp. 227–235. ACM, Limassol (2006)
2. Feasel, J., Whitton, M., Wendt, J.: LLCM-WIP: Low-latency, continuous-motion walking-in-place. In: Proceedings of the 2008 IEEE Symposium on 3D User Interfaces, pp. 97–104. IEEE Computer Society, Reno (2008)
3. Razzaque, S., Swapp, D., Slater, M., Whitton, M., Steed, A.: Redirected walking in place. In: Proceedings of the Workshop on Virtual Environments 2002, pp. 123–130. Eurographics Association, Barcelona (2002)
4. Slater, M., Usoh, M., Steed, A.: Taking steps: The influence of a walking technique on presence in virtual reality. ACM Transactions on Computer-Human Interaction 2, 201–219 (1995)
5. Templeman, J., Denbrook, P., Sibert, L.: Virtual locomotion: Walking in place through virtual environments. Presence, Massachusetts Institute of Technology 8, 598–617 (1999)

6. Terziman, L., Marchal, M., Emily, M., Multon, F.: Shake-your-head: Revisiting walking-in-place for desktop virtual reality. In: Proceedings of ACM VRST 2010, pp. 27–34. ACM, Hong Kong (2010)
7. Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A., Slater, M., Brooks, F.: Walking > walking-in-place > flying. In: Proceedings of ACM SIGGRAPH 1999, pp. 359–364. ACM Press/Addison-Wesley (1999)
8. Wendt, J., Whitton, M., Brooks, F.: GUD-WIP: Gait-understanding-driven walking-in-place. In: Proceedings of IEEE Virtual Reality 2010, pp. 51–58. IEEE Computer Society, Waltham (2010)
9. Whitton, M., Cohn, J., Feasel, J., Zimmons, P., Razzaque, S., Poulton, S., McLeod, B., Brooks, F.: Comparing VE locomotion interfaces. In: Proceedings of EEE Virtual Reality 2005, pp. 123–130 (2005)
10. Yan, L., Allison, R., Rushton, S.: New simple virtual walking method – walking on the spot. In: Proceedings of 8th Annual Immersive Projection Technology (IPT) Symposium Electronic (2004)
11. Optitrack (September 2012), `http://www.naturalpoint.com/optitrack/`
12. Bowman, D., Kruijff, E., LaViola, J., Poupyrev, I.: 3D User Interfaces – Theory and Practice, 1st edn. Addison-Wesley (2005)
13. Williams, B., Bailey, S., Narasimham, G., Li, M., Bodenheimer, B.: Evaluation of Walking in Place on a Wii Balance Board to Explore a Virtual Environment. ACM Transactions on Applied Perception, vol 8, 19:1–19:14 (2011)
14. Shiratori, T., Hodgins, J.: Accelerometer-based User Interfaces for the Control of a Physically Simulated Character. ACM Transactions on Graphics 27(5), 123:1–123:9 (2008)
15. Dean, G.: An analysis of the energy expenditure in level and grade walking. Ergonomics 8, 31–47 (1965)
16. Inman, V., Ralston, H., Todd, F.: Human walking. Edwin Mellen Pr. (1981)
17. Mine, M.: Virtual Environment Interaction Techniques. UNC Chapel Hill CS Dept. (1995)

Disambiguation Canvas: A Precise Selection Technique for Virtual Environments

Henrique G. Debarba[1,2], Jerônimo G. Grandi[1], Anderson Maciel[1], Luciana Nedel[1], and Ronan Boulic[2]

[1] Instituto de Informática, Universidade Federal do Rio Grande do Sul (UFRGS)
[2] École Polytechnique Fédérale de Lausanne (EPFL)
{jggrandi,amaciel,nedel}@inf.ufrgs.br,
{henrique.galvandebarba,ronan.boulic}@epfl.ch

Abstract. We present the disambiguation canvas, a technique developed for easy, accurate and fast selection of small objects and objects inside cluttered virtual environments. Disambiguation canvas rely on selection by progressive refinement, it uses a mobile device and consists of two steps. During the first, the user defines a subset of objects by means of the orientation sensors of the device and a volume casting pointing technique. The subsequent step consists of the disambiguation of the desired target among the previously defined subset of objects, and is accomplished using the mobile device touchscreen. By relying on the touchscreen for the last step, the user can disambiguate among hundreds of objects at once. User tests show that our technique performs faster than ray-casting for targets with approximately 0.53 degrees of angular size, and is also much more accurate for all the tested target sizes.

Keywords: Selection techniques, 3D interaction, usability evaluation, progressive refinement.

1 Introduction

Selection is one of the four fundamental forms of interaction in a virtual world [1, 2]. It is the ability of the user to specify objects in the virtual environment for subsequent actions [3]. The literature is rich in immediate selection techniques; however, this class of technique is exposed to problems of accuracy, ambiguity and complexity [4]. Many applications rely more on correctness of selection than on time of selection, but ordinary selection techniques in use tend to favor performance over accuracy. We intend to provide a precise yet fast alternative for selection in virtual environments. Therefore, we rely on selection by progressive refinement, as proposed by Kopper et al. [5].

Selection by progressive refinement proposes the breakdown of a selection task into "effortless" subtasks. It aims to avoid the attention and precision usually required by traditional selection techniques, so-called immediate selection techniques [5, 6]. However, there is an inevitable tradeoff between immediate and progressive refinement selection techniques. To complete a selection, the latter requires a process which

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 388–405, 2013.

usually consists of more than one quick subtask, generally resulting in higher accuracy and longer selection time. On the other hand, immediate selection techniques consist in performing the selection in only one step, being generally faster but less accurate.

We propose the *disambiguation canvas*, which is a technique for quick disambiguation of selection. We use the observed high precision of control provided by the touchscreen [7, 8] to allow the disambiguation of the desired object among a subset of hundreds of other objects in only one step of refinement. Previous progressive refinement techniques do not scale as well as ours. Available techniques that disambiguate only in one step are limited to a small subset of objects, while those that refine among large subsets require multiple steps of disambiguation.

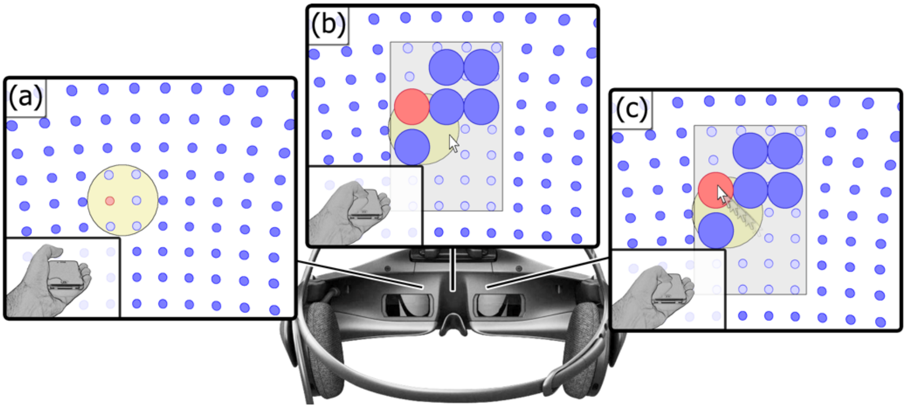


Fig. 1. Disambiguation canvas walkthrough: (a) the user points to the region where the desired object is located; (b) starting a touch rearrange the subset of objects pre-selected by the volume casting technique over a selection canvas; (c) the canvas has an absolute mapping to the mobile device touchscreen, the user slides his thumb in order to point out the desired object. Disambiguation canvas was designed to be compatible with immersive displays, such as the depicted head-mounted display. In this figure, the hand inserts illustrate the user gestures for each step, and are not displayed by our technique.

The *disambiguation canvas* is based on a two steps process. In the first step, the user employs a volume casting technique to point in the direction of the desired target object (see Figure 1a). When the target object is inside or intersecting the volume of selection, the user may start a touch on the mobile device touchscreen to enter the second step of the selection. A rectangle aligned parallel with the image plane – or with the mid orientation of the two image planes when stereoscopic rendering is in use – appears in front of the user; all the subset of objects pre-selected by the volume moves in an animation to form a matrix inside this rectangle (Figure 1b). The rectangle has a 1:1 mapping with the mobile device touchscreen, sliding the thumb on the touchscreen allows the superposition of the desired object by the arrow (Figure 1c). Selection is performed by a *take-off* gesture. If the user wants to leave the disambiguation phase without selecting any object, they simply perform a *take-off* gesture with the arrow over an empty region of the rectangle.

The remainder of this paper is organized as follows. In order to provide the reader with a context, Section 2 summarizes the main selection techniques that consider progressive refinement; Section 3 presents the proposed technique design decisions, prototype hardware technology and software implementation. In Section 4 we present the evaluation of the *disambiguation canvas*, comparing it with the *ray-casting* and *SQUAD* [5], and present the analysis of results. A deeper discussion covering qualitative results as well as suggestions for design changes are presented in Section 5. Finally, in Section 6 we highlight our findings and suggest future developments.

2 Related Work

The design space of selection by *progressive refinement* was presented by Kopper et al. [5]. It is defined as an approach to progressively reduce the group of selectable objects and hence reduce required precision of pointing. From the current literature, we could identify three major groups of progressive refinement selection techniques: *menu disambiguation*, *zoom* and *persistence of pointing*. These are discussed below, as well as some techniques that do not fit these main categories.

Menu disambiguation: generally uses a volume of selection on the initial phase to reduce the effort of pointing into the desired object. Objects that fall inside or intersect that volume are then presented as a subset of objects using some sort of menu for disambiguation.

Dang et al. presented the *transparent sphere* and *transparent cylinder* [9]. On the transparent sphere technique, a positional cursor similar to the *virtual hand* metaphor [2] is used to place a spherical volume of selection in space. Objects inside or intersecting this sphere have their names shown on a menu. Disambiguation is performed by selecting the desired object name. On the other hand, the transparent cylinder uses a *ray-casting* based approach where a cylindrical volume is attached along the cast ray in order to define the subset of objects. Transparent sphere and cylinder present only the name of the target for disambiguation. Thus their original design is unsuitable for a series of applications.

Grossman and Balakrishnan [10] proposed the *flower ray* for interaction with a volumetric display. The *flower ray* uses *ray-casting*, and disambiguates using a marking menu. When entering the menu disambiguation step, intersected objects animate towards the user viewport and spread as a marking menu. However, this technique still requires precision of pointing as it relies on ray-casting for both phases, and would have problems disambiguating among a large subset of objects.

When proposing the taxonomy for the progressive refinement selection techniques, Kopper et al. [5] also presented the SQUAD technique (sphere-casting refined by QUAD-menu). SQUAD consists of defining a subset of objects through their intersection with a sphere volume, and further refining the subset through QUAD menus until only one object remains. As SQUAD relies on several steps of disambiguation, we believe the major drawback of this approach is that the visual search is repeated in each step. If the desired object is similar to others, the visual search can be even more time consuming than the pointing task itself. This question was not addressed in the original study.

Zoom: Bacim et al. [6] propose two techniques for progressive refinement based on zoom, *discrete zoom* and *continuous zoom*. In the *discrete zoom*, the user defines a quadrant of the screen they want to see in more detail, the frustum changes so that the specific quadrant covers all the *field of view* (FOV). In the *continuous zoom* technique, the zoom happens continuously towards the pointing direction. Zoom based techniques have the advantage of showing the objects in their original context. However, manually controlled zoom tends to be more time consuming. Indeed, the evaluation presented by Bacim et al. showed lower performance when compared to the SQUAD technique.

Score accumulation: although not originally classified as progressive refinement selection techniques [5], we advocate that score accumulation techniques present the expected behavior described by the authors. These generally rely on the consistency of pointing, where objects that were targeted for a larger duration accumulate higher scores, becoming more likely to be the intended target of a selection.

Haan et al. [4] use an approach similar to *lightspot* (cone-casting) [11] for the *intenselect* technique. However, on *intenselect* an alternative disambiguation function is applied and expanded to the dimension of time. Objects that fall inside the cast cone accumulate scores over time. Their scores increase proportionally to their proximity to the center of the cone and its casting origin. If an object stops intersecting the cone, its score is gradually lowered. Visual feedback of pointing is given by a bended ray connecting the origin of the ray to the object with the higher score.

Grossman and Balakrishnan have implemented and evaluated the *smart ray* – technique first proposed by Steed [3] – on a volumetric display [10]. All objects intersected by the ray accumulate scores; to disambiguate, the user moves the origin and direction of the ray so that it always intersects the intended object. As long as the user succeeds maintaining the ray over the desired object for more time than any other, she is able to select it. *Smart ray* gradually decreases the score of the objects that have lost intersection with the ray. Therefore, it still maintains most of the score of objects that unintentionally lost contact with the ray for a short period of time.

Other approaches: Steed and Parker have proposed *shadow cone-casting* [12], which uses cone casting persistence of pointing along time to define a selection. When the user starts a selection, all objects inside the cone are selectable. The user must disambiguate among these objects by moving the origin of the cone while trying to always maintain the desired object(s) inside the cone. If an object falls outside the cone, it is cut out of this selection process. This technique allows the selection of multiple targets. However, it relies on the proximity of objects for this. Additionally, this technique is likely to be very time consuming in a cluttered environment, where high precision is required.

Grossman and Balakrishnan have proposed the *lock ray* [10], which expands their own technique of *depth ray*. The depth ray uses forward and backward hand movements to disambiguate which of the objects intersected by ray-casting will be selected. In the lock ray, these steps are performed in sequence, assuring higher precision control as the ray-casting becomes locked, while in the depth ray they are performed simultaneously. Both techniques still require high precision of pointing in order to hit a target with the ray-casting metaphor.

Stellmach and Dachselt presented *Look & Touch*, which has more similarities with our approach [16], although being designed for 2D selection. With *Look & Touch* the user uses his gaze to control the direction of the lower precision phase of pointing; objects intersecting the circular area of their cursor are pre-selected, and the user can disambiguate cycling through the objects on an *iPod touch* touchscreen. Besides this disambiguation technique, they also propose a relative and an absolute control of the cursor using the touchscreen, similar to [17, 8].

3 Technique Design

This section exposes our design decisions while developing the *disambiguation canvas*, its particularities (such as objects distribution over the plane) and its implementation details.

3.1 Volume Casting Techniques

The most common approaches for volume casting are the *cone-casting* and the *sphere-casting* techniques. Which one of these is best fit for our technique may depend on the application. Thus, we decided to support both volume-casting techniques for the disambiguation canvas.

Using the cast of a sphere, it is likely that the amount of objects intersecting the volume is smaller, as the sphere has a limited depth. However, it is also harder to control the first step of the selection as the depth must be somehow provided; there are two common approaches to determine the sphere depth. The first uses the near intersection of the sphere, and sets its distance to the intersection position. The second casts a ray through the center of the sphere; the distance of the first intersection of this ray is used to set the sphere depth. In *SQUAD* for instance, Kopper et al. [5] favored the *sphere-casting* with the depth of the sphere determined by *ray-casting*, but this may be due to the type of environment for which they developed the technique, a virtual supermarket application. In the supermarket environment, objects were very cluttered and organized as stacks in many shelves. Such organization facilitates the task of pointing out a cluster of objects with *sphere-casting*.

On the other hand, *cone-casting* allows reaching objects even if an intersection occludes them from the casting origin point of view, which sphere-casting is unable to do. Cone-casting can also reduce the necessary precision during the first stage, as no depth input is required. However, *cone-casting* may intersect too many objects if the scene is very cluttered. In the supermarket case, *cone-casting* would require constraints in order not to select objects behind the shelves; otherwise a huge amount of objects may fall inside the conic volume.

Based on the report by [13] that users achieved up to 4° of error during the coarse precision phase of distal pointing, we suggest the angular size of 12° for the sphere/cone casting technique. Six degrees from the center of the ray to the sphere/cone borders. The sphere/cone always rescales to achieve the angular size of 12° from the casting position point of view.

3.2 Graphic Representation

We use an *arrow* shape for the cursor, and a semi-transparent *rectangle* to represent the canvas. Based on its widespread use, we concluded that the arrow would be the most natural and intuitive representation for the cursor pointing position, while the rectangle helps to easily match and associate the mobile device touchscreen with an area on the virtual environment. These shapes are only visible during the second step, which is the disambiguation step. The mobile device pointing direction is represented by a semi-transparent sphere or a cone, depending on the volume-casting technique used. The volume-casting shape in use is always visible, so the user can always have feedback on his/her pointing direction, even when performing the second step of a selection (disambiguation step).

The rectangle uses an absolute mapping with the mobile device touchscreen. It is drawn to use 30° of the total 45° standard vertical FOV of the camera. We positioned it 70*cm* away from the camera on our immersive display implementation, thus we obtained the size of ≈38*cm*. However, in order to make better use of the FOV, this size should be decided according to the available display, allowing the user to inspect the objects more efficiently, and therefore reducing the visual search time. The distance on which it is drawn may also vary in order to avoid occlusion with other objects in the scene. In our specific implementation, we adopted the distance of 70*cm* to maintain a pleasant stereoscopic rendering when switching from background to canvas.

3.3 Mapping Objects to the Canvas

When entering the second step of the *disambiguation canvas*, the subset of objects must be reorganized side by side over the canvas plane. As most users might have trouble reaching the whole touchscreen with the thumb, we propose two standard layouts on which the objects are reorganized so the user can easily reach them. The first consists of ≈53.4% of the total area, and is oriented to user handedness (Figure 2a). This layout takes 5% from the right, the left and the top, and 25% from the bottom out of the useful area, as well as 1/8 and 1/16 of the remaining that is too close to the palm and far from the thumb reach respectively. The second layout consists of ≈42.4% of the total touch screen area. This layout takes 5% from the right and the left, 10% from the top, and 30% from the bottom out of the useful area. The final layout consists of a circle inside the remaining area, as illustrated in Figure 2b.

For a preliminary (not presented) and the first evaluation (Section 4.1), the layout presented in Figure 2a was used. Although in general it has worked properly for most subjects, two users from the preliminary evaluation had difficulty to reach a large portion of the layout, and in consequence obtained significantly higher error rates. Thus, we have also decided to approach a layout calibration method. To calibrate, the user performs circular movements with the thumb on the touchscreen within their range of comfortable motion (Figure 3a). A flood fill algorithm identifies the outermost bounds for that user's specific layout. This approach was used in the second evaluation, presented in Section 4.2.

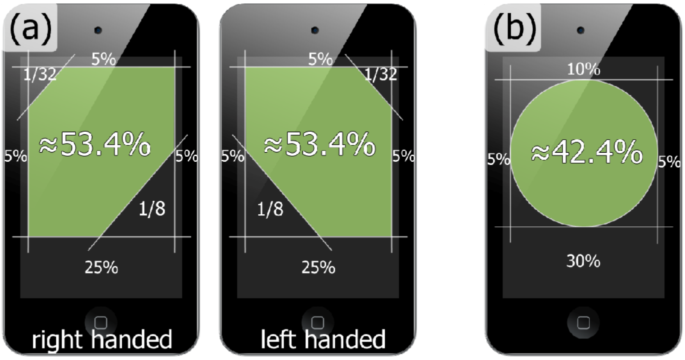


Fig. 2. Standard layouts of useful touchscreen area proposed for the disambiguation canvas techniques. The (a) standard layout was used for the first presented evaluation.

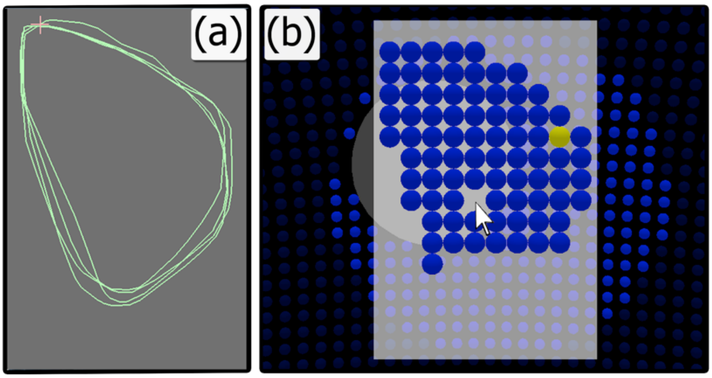


Fig. 3. Proposed layout calibration process, (a) contour of the reachable area defined by a user. (b) Arrangement of objects inside the reachable area, notice that the starting position of the arrow is kept empty. This procedure was used for the second evaluation (Section 4.2).

When switching to the disambiguation phase, a matrix fitting every pre-selected object inside the layout is computed, and each object is designated to a valid slot of the matrix (Figure 3b). A slot is considered valid if its center is located inside the usable area defined by the layout. In order to fit every object inside their designed slot, the objects are rescaled so their bounding box does not pass beyond that space. However, this may make some visual attributes less apparent or even impossible to be perceived, such as when the user wants to select a target of specific size within a group of similar objects. To overcome this issue the rescale factor may also be proportional to the largest and smallest target, being linearly remapped between a minimum and maximum threshold. That is, if the smallest object is ten times smaller than the largest one, this proportion factor (1:10) will be lowered to a maximum of half of the size (1:2) so both objects remain visible and distinguishable. Objects with intermediate sizes are proportionally rescaled in between these thresholds. By default, our current implementation uses this approach.

In our current implementation, which was used for the second user evaluation (Section 4.2), the position where the user starts a touch is not superposed by objects. This design facilitates leaving a selection procedure if the user does not want to select any of the objects pre-selected by the first phase (Figure 3b). In addition, it also avoids accidental selection in the case of an unintentional touch by the user.

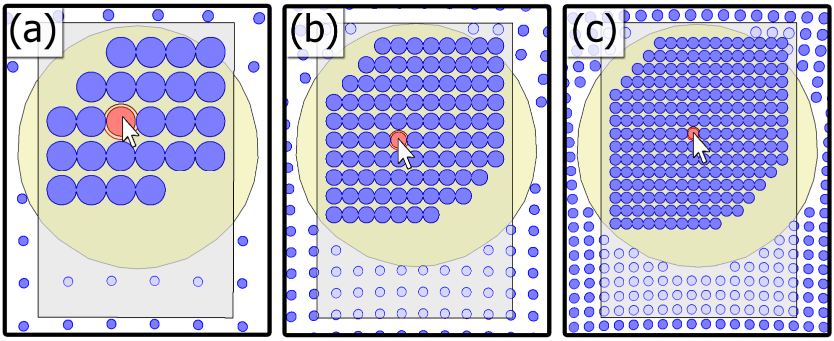


Fig. 4. Difficulty of selection is proportional to the amount of objects pre-selected by the first step; in (a) there are 25 objects on the canvas, in (b) 97 objects, and in (c) 224 objects.

A regular immediate selection technique usually has its difficulty of pointing increased by reducing the target object size. However, when using the *disambiguation canvas* the difficulty increases according to how many objects have been pre-selected by the volume-casting during the first phase. This is an expected behavior of progressive refinement selection techniques, and might make the refinement slower and/or harder. However, as the *disambiguation canvas* relies on the mobile device touchscreen for disambiguation, we are able to align hundreds of objects within a single disambiguation step while still ensuring high precision. During the first technique evaluation (Section 4.1), we have used three distinct object densities; they are shown in Figure 4. For the worst case depicted in Figure 4c, 224 objects went to the disambiguation phase. Still, our technique offered the very convenient sensing area of ≈230 pixels of the touchscreen (the orthogonal projection of the spherical object over an area of 19×19 pixels), which has a total input area of 320×480.

3.4 Prototype Implementation

We used an *Intel Core i7* computer, equipped with two *AMD Radeon HD 5870 Eyefinity6*. The immersive display is a *Sensics zSight Integrated SXGA Head Mounted Display* (HMD) (Figure 5a). It provides stereoscopic vision using two 1280 x 1024 displays, and has a FOV of 60°. This HMD also provides the orientation of the head. For the mobile device, we have chosen the *Apple iPhone 4/4S* and *iPod touch 4*.

The software displayed by the *zSight HMD* is implemented in *C++*, using *Ogre3D* for graphics [14] (Figure 5b). We support stereoscopy in our application. The mobile device software is an *app* implemented in *Objective-C*. It acquires the sensor readings and communicates them over a Wi-Fi infrastructure through *UDP*. To obtain the orientation of the *iPhone 4/4S*, which has a *magnetometer* – thus providing the recalibration of drift on the *yaw* – we have used the strategy proposed by Madgwick [15] with the adaptations presented in a previous work [8]. In [8], a complete description on the acquisition and processing of sensor data to provide orientation is presented. To obtain the orientation of the *iPod touch 4*, which does not contain a *magnetometer*, we have used the standard orientation provided by the *iOS SDK*.

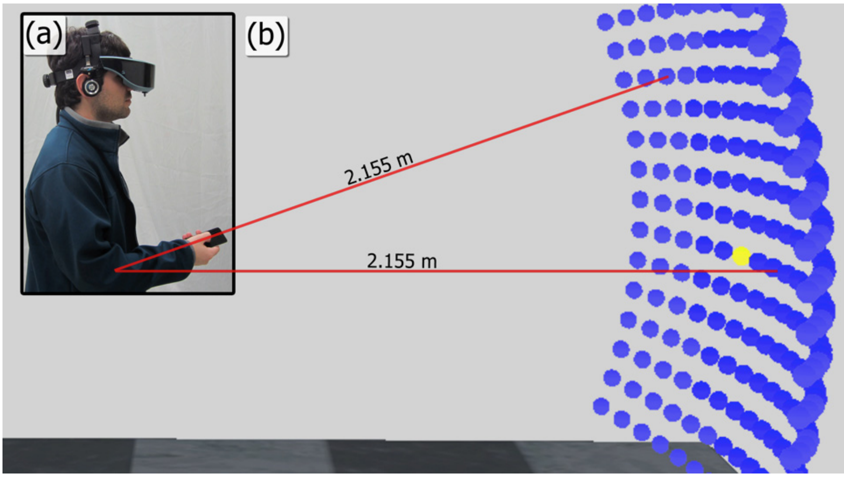


Fig. 5. Merged images of the prototype overview (a), and the test application used for evaluation (b)

4 Disambiguation Canvas Evaluation

We conducted two sets of user tests for the *disambiguation canvas* technique. The main design decisions are common to both evaluations and were based on those used by Kopper et al. for the SQUAD technique evaluation [5]. Both were comparative evaluations and used a within-subject design. In the first we compare *disambiguation canvas* with *ray-casting*, while in the second we compare it with *SQUAD*. The implemented ray-casting relies only on the orientation of the device – assumption of a constant casting position – and is therefore referred to as *ORayCasting* (orientation ray-casting); *disambiguation canvas* and SQUAD also use only the orientation of the device for the volume-casting step. The *disambiguation canvas* is referred to as *DCanvas*. For both evaluations the objects pre-selected by the first phase of DCanvas were animated from their original position to the canvas in a 250ms animation.

An *iPod touch 4* was used in both evaluations. As mentioned earlier, it is not equipped with a magnetometer, and thus is subject to drift on *yaw*, losing its correct orientation. Therefore we used blocks with no more than 11 trials, and applied an offset in order to correct any accumulated error between blocks. On the other hand, it produces an orientation less noisy than using the magnetometer to correct yaw, necessary for reliability of pointing while using ray-casting. Mobile device orientation was filtered using a *dynamic low-pass filter*, interpolating between cutoffs of 0.2Hz and 50Hz, with 60Hz sampling rate. Cutoff is defined according to the angular change speed in degrees: when < 1°/*sec*, the lowest cutoff is used (0.2Hz); when > 50°/*sec*, the highest cutoff is used (50Hz). For any speed between those, a linearly interpolated cutoff value is used.

General goal: the goal was to select a yellow sphere among several distractors of same size represented as blue spheres. These objects were arranged as a matrix. To position them, we use a main sphere of 2.155*m* radius. All the selectable spheres were positioned with their centers intersecting the borders of this main sphere. The origin of the ray/sphere-casting was set to the center of the main sphere. This guarantees the same angular pointing size for all the objects. The virtual camera was positioned 50*cm* above the ray/sphere-casting origin. This configuration is shown in Figure 5.

Evaluation procedure: The same procedure and similar questionnaires were used for both of our evaluations. The procedure was as follows:

1. The subject was asked for any health issue or impairment that could prevent them from participating (such as a history of epilepsy and color blindness)
2. The subject filled in a characterization questionnaire
3. The subject was presented to the first technique on an ordinary screen display (so the experimenter and the subject could share the view while explaining how the technique works)
4. The experimenter presented the HMD and how to adjust it to the head
5. The subject performed practice blocks with the first technique
6. The subject performed evaluation blocks with the first technique
7. The subject answered a questionnaire about the first technique
8. The subject repeated steps 3, 5, 6 and 7 for the second technique
9. The subject filled in a post-experiment questionnaire comparing both techniques

A block consisted of a collection of trials. A trial consisted of a selection task, ending with an activation of selection, which could be successful or not. The number of blocks and trials is different for each evaluation. Subjects were allowed to remove the HMD and rest between the blocks if desired.

4.1 Comparison with ORayCasting

Design: sphere-casting was used for the volume-casting step of selection on DCanvas. Instead of using the suggested standard size of ≈12° for the casted sphere, we used the angular size of ≈26°, so more objects would be pre-selected for the subsequent disambiguation phase. This sphere is represented by a semitransparent sphere. The ORayCasting casted ray is represented by a cylinder with 1*cm* of diameter.

The independent variables are *angular space* between objects (resulting in different distractors density): 5°, 2.5° and ≈1.67°; and *angular size*: ≈0.53°, ≈1.06° and ≈1.6° (2*cm*, 4*cm* and 6*cm* respectively). We used blocks of 10 trials, 9 of them representative of the combination of *density* × *size*, and an initial target which was used to start the block. Training consisted of 5 blocks, while the evaluation consisted of 10. The 9 valid targets within each block were randomly presented, while technique presentation order was counterbalanced.

The target was randomly chosen among the objects with only one constraint, this object should have its center within a range between 52*cm* and 77*cm* from the center of the matrix of objects. We did so in order to keep the possible targets within the user's field of view, thus reducing visual search bias. We have also colored objects farther than 77*cm* to green, so the user knows they are not target candidates. For this evaluation, the collision checking on the disambiguation step was performed using the superposition of the arrow over the projection of the sphere on the canvas (a circular area).

The comparison with *ORayCasting* was also intended to verify design decisions and to evaluate whether the technique was comprehensive and easy to use. The design choices presented in Figure 3 were not used in this evaluation. In fact, they were implemented after the feedback from this experiment and were used for the following *SQUAD* comparison (Section 4.2).

Subjects: six graduate students in Computer Science from our university participated in this experiment (mean age of 29, four right handed). All of them were very experienced in managing mobile device touchscreens, and had at least some experience using natural pointing devices. In a 7 points scale, only two reported experience with virtual reality equipment of 3 or more points. Each test took from 25 to 40 minutes to be performed. We have obtained a total of 1,080 valid trials: *2 techniques* × *6 subjects* × *10 blocks* × *9 trials*.

Results: overall mean selection times with *DCanvas* and ORayCasting were respectively: 2.37 and 2.29 seconds. One-way ANOVA showed that *DCanvas* was slower with statistical significance when compared to *ORayCasting* ($F(1.1078)=4.43$, $p<0.036$). See Figure 6 for detailed mean time for each combination of target size and density. Error rate with ORayCasting was significantly higher than with *DCanvas* ($F(1.1078)=70.34$, $p<0.0001$). Error rates for each combination of size and density are presented on Figure 6.

Figure 7 presents the time and error rate per user. For this experiment, *DCanvas* obtained a lower mean time for two users. Four users have not made any selection error while using our technique. Subject 6 presented an exceptionally low error rate for *ORayCasting*, but still higher than with *DCanvas*.

The intermediate questionnaire asked users to rate each technique concerning: ease of learning and ease of use; how well it performs for small, medium and large targets; and how much fatigue was felt on their wrist, hand, fingers, back and legs. Results are presented in Figure 7. Both were considered very easy to learn, while our technique was considered easier to use. *DCanvas* was preferred over *ORayCasting* for small and medium targets, while large targets received equivalent ratings for both techniques. Overall fatigue was lower for *DCanvas*; its mean of the 5 related questions was 2.1, against 2.5 of *ORayCasting*.

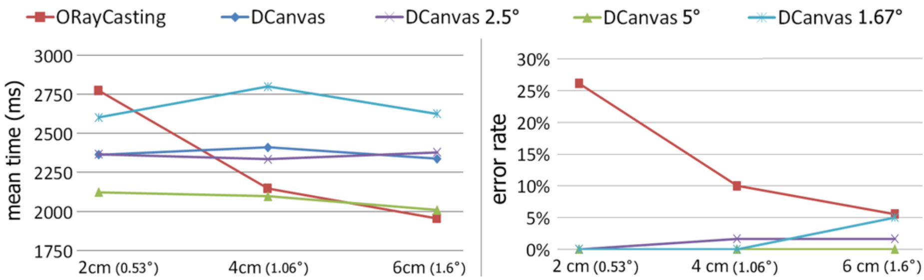


Fig. 6. Mean trial completion time and error rate for each combination of target angular size and density for the comparison with ORayCasting

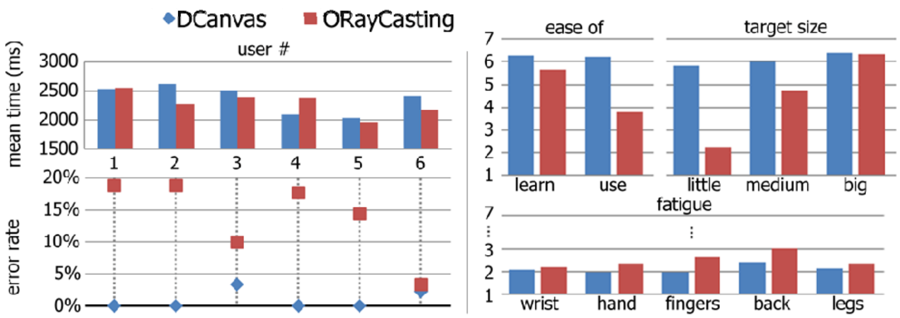


Fig. 7. Mean time and error rate per user and subjective questionnaire scores for DCanvas and ORayCasting

Regarding the comparative questionnaire, *DCanvas* presented higher scores for all the questions. It was considered more accurate (6.7 against 1.6), faster (5.9 against 2.7), less tiring (4.8 against 2.6) and easier to use (5.6 against 2.5). Curiously, users felt that *DCanvas* was faster, which is true for ≈.53° angular size targets, but false for the overall evaluation. *DCanvas* was also preferred by all the users.

4.2 Comparison with SQUAD

Design: for the comparison between DCanvas and SQUAD we have eliminated the target size from the set of independent variables. Our previous evaluation, as well as the one performed by Kopper et al. to test SQUAD, showed that the target size is not significant for time of selection or error rate. Thus we used the constant angular size of ≈1.06° (4*cm*) for all the objects. For the independent variable of angular space between objects (distractors density) we used: 6°, 3°, 2°, 1.5° and 1.2°.

We used blocks of 11 trials, 2 trials of each density, which were randomly presented, and an additional initial target used to mark the start of the block. Training consisted of 3 blocks, while the evaluation consisted of 4. Technique presentation order was counterbalanced. The target object was randomly chosen with the same constraint as before, but accepting a range between 45*cm* and 55*cm* from the center of the matrix of objects. Objects beyond 60*cm* from the center were colored in dark blue, which is less distractive than the green used for the previous evaluation.

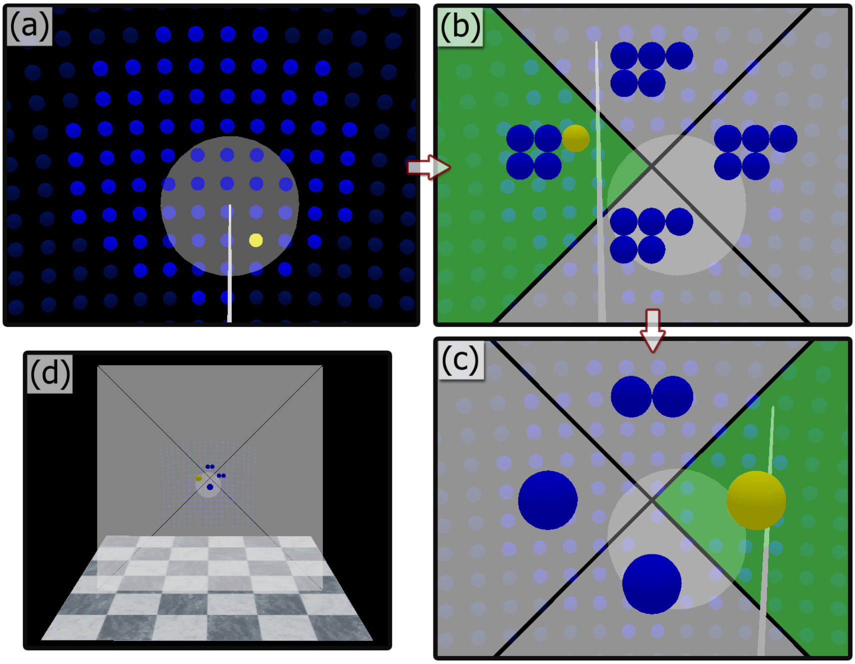


Fig. 8. SQUAD implementation used for comparison: (a) the group of objects inside the sphere-casting volume is selected by a tap gesture; (b) these objects are rearranged into quadrants in a 250ms animation, to select a quadrant the user perform a tap gesture while intersecting it with ray-casting; (c) the subgroup is rearranged in new quadrants with a 200ms animation, a new tap gesture while intersecting the target quadrant results in a successful selection. Note that although the animations impose some time constraints, it also avoids the need for visual search at each new step, and the user can start the repositioning of the ray during the animation. (d) Overview of the quadrants area of selection, a tap gesture while not intersecting any quadrant – or intersecting an empty quadrant – may be performed to leave the selection procedure.

We used *sphere-casting* for the first step of selection with DCanvas and SQUAD. Instead of adopting the suggested standard size of ≈12° for the casted sphere, we used the angular size of ≈17° for both techniques, so more objects would be pre-selected for the subsequent disambiguation phase. Combined with the possible angular space between objects for this evaluation, the sphere-casting phase could pre-select ≈6, ≈25, ≈55, ≈100 or ≈160 objects for the disambiguation phase. The sphere-casting step of the DCanvas and SQUAD was represented by a semitransparent sphere.

The SQUAD casted ray – disambiguation phase – was represented by a cylinder with 1*cm* of diameter. Figure 8a-c presents the walkthrough of our SQUAD implementation. The quad menu is drawn at a distance of 100*cm* from the camera, and is oriented to face the camera. The quad menu is composed of four triangles that assemble a square of 4×4 meters (Figure 8d). To recover from a mistaken pre-selection, the user can point outside of the quad menu and perform a tap gesture, or select an empty quadrant. An error only occurs when the user selects a wrong quadrant that contains only one object inside.

For this evaluation, the collision checking on the disambiguation step of DCanvas was performed using the superposition of the arrow over the designated objects slot (instead of the object projection) in the canvas. This approach increases the effective selection size of the object, and also allows a simpler collision test when dealing with objects with a mesh more complex than those tested (spheres). Additionally, the improvements presented in Figure 3 were also used.

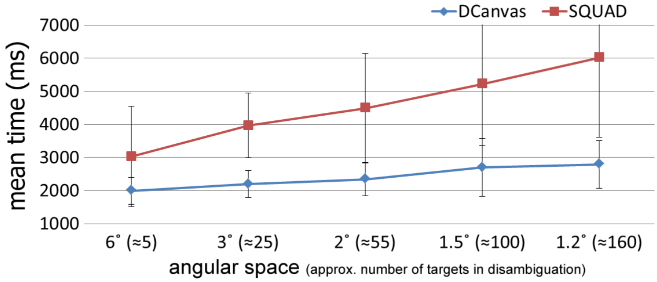


Fig. 9. Mean trial completion time for each angular space (density) of objects distribution for the comparison with SQUAD

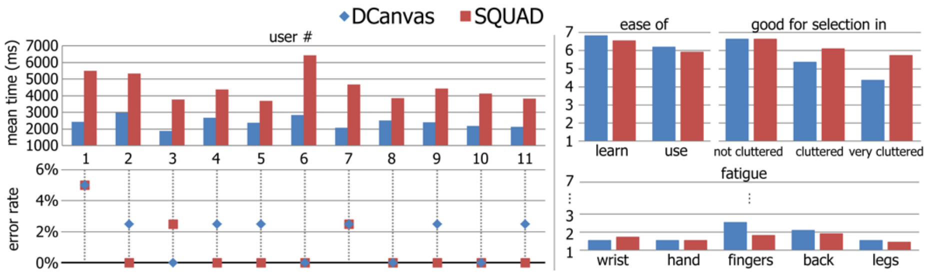


Fig. 10. Mean time and error rate per user and subjective questionnaire scores for the comparison with SQUAD

Subjects: eleven subjects participated on this evaluation, all of them students or professors in computer science or electrical engineering (mean age of 31). On a seven points scale, only one reported little experience with mobile device touchscreens (below 5), and three reported experience with pointing devices equal or above 4 points. Only one subject reported high frequency of use of virtual reality devices.

Results: as demonstrated by Figure 9, DCanvas performed significantly faster than SQUAD for all conditions. We also highlight that the increase in selection time for DCanvas was less steep than SQUAD. For the error rate, both techniques achieved good marks. DCanvas had an error rate of 0.018 errors per trial, while SQUAD achieved 0.009 errors per trial. Figure 10 show the individual performance of mean time and error rate for each user.

Concerning the comparative questionnaire, 6 subjects preferred the DCanvas, while 5 liked SQUAD most. The mean scores of comparative questions were very similar between techniques. DCanvas was considered less precise by a difference of ≈1.26 points on the 7 points Likert scale, and more difficult to learn and use by a difference of ≈0.45 and ≈0.39. DCanvas was regarded as faster by a difference of ≈0.19.

However, on the questionnaire specific for each technique, when no direct comparison was required, the DCanvas obtained higher scores concerning its ease of learn and use (Figure 10). On the other hand, SQUAD received higher absolute scores concerning the level of cluttering of the environment. Concerning the fatigue, answers were generally very similar, except by the fatigue on the fingers, which was higher for DCanvas. The mean of these scores are reported in Figure 10.

5 Discussion and Final Remarks

5.1 Transition to the Canvas

The most recurrent feedback left by the users regards the transition of the subset of objects from its original context to the control canvas. Users frequently had the conviction that positioning the intended target near the center of the sphere during the sphere-casting step would take that target near the center of the canvas when switching to disambiguation. This intuition may arise from the arrangement of objects as a matrix, which would be easily fitted inside the layout. However, on a more complex scenario, with targets spread in depth, such organization is not so obvious. We are currently working on this issue, as it could reduce user effort of reaching objects mapped to distant regions of the touchscreen and reduce visual search time.

5.2 Keeping the Context

As for being a progressive refinement technique based in menu disambiguation, objects that go from the first to the second phase lose their original context. This could make it difficult to distinguish the intended object in real applications if they are very similar in shape or if the selection depends on their original topology. We propose three possible solutions for such limitation of the menu disambiguation approach. The first solution is to control the instant of interpolation that animates the objects while bringing them over to the disambiguation menu. To cast a ray or a volume for pointing, only 2 degrees of freedom (DOF) among the 3 provided by the device orientation are required. Our proposal is to use the 3rd DOF to dynamically control the instant of the interpolation. The mapping from orientation into instant of interpolation can be achieved with an absolute relation, where a certain orientation always results on the same instant, or with a relative relation, where after a threshold the orientation controls acceleration forward or backward on the interpolation instant. This strategy was implemented, and showed to be functional; however it was not yet evaluated. Figure 11 shows four frames of an animation, moving the objects from their original to new position in the disambiguation canvas.

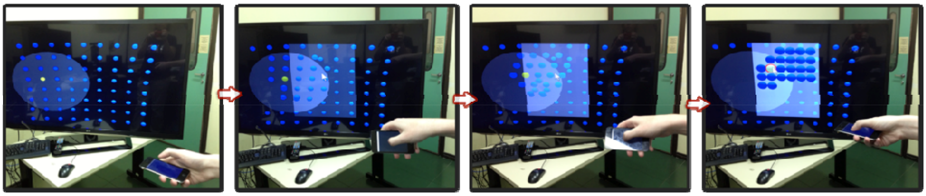


Fig. 11. Four frames illustrating the animation of the objects from their original positions to the new ones on the disambiguation canvas. The animation can be reproduced in both directions by twisting the mobile device in the corresponding direction.

The second technique consists in duplicating the original object into the canvas – instead of moving the original – and using the copy superposed by the arrow cursor to highlight the original object. It can indicate whether the user is pointing to the desired object when there is more than one object with the same or similar shapes.

The third approach is to draw a trajectory curve to connect the original position of an object with its final position on the canvas. This approach allows the simultaneous observation of all connections between original and final positions at the same time. However, this can result in cluttering and may overwhelm the user with information when too many objects are taken for disambiguation.

Notice that these suggestions are not exclusive and can be combined among them. Given that this is a general problem of menu based progressive refinement selection techniques, we intend to investigate these approaches further in future works.

5.3 Immersive Tool

The tests we led, as well as the prototype implementation described in Section 3.4 considered the use of a head-mounted display (HMD) and 3D stereo visualization to enhance the realism and immersion. Subjects were comfortable with this set up. However, even if the camera is driven by the movement of the user's head, we are not taking much advantage of this, since the objects were concentrated in a relatively small area, in front of the user. Our objective so far was to compare the disambiguation canvas with other techniques – as ray-casting and SQUAD – accordingly. Then, we tried to avoid any other independent variable.

Informally, we also tested the disambiguation canvas with a regular display. Our intuition is that its use with a regular display depends strongly on the layout of the objects. On the other hand, using the HMD provides easy control of the camera, which comfortably overcomes this limitation.

We are also aware that, to verify the robustness of the technique, more tests should be done with other layouts for the objects in the scene. Currently, they are all disposed on the surface of the sphere that surrounds the user.

6 Conclusions and Future Work

In this paper we have presented the *disambiguation canvas*, a technique for fast and high accuracy selection of objects in a 3D space. It relies on *progressive refinement* to address the lack of accuracy common to immediate selection techniques, and uses distinct input hardware to optimize the control over its two steps and overall time performance on a selection procedure.

By using the touchscreen on the second step of the *disambiguation canvas*, users were able to select objects represented in a motor area of $\approx 7.9mm^2$ of the touchscreen surface very efficiently during evaluation, allowing consistent disambiguation among a group of 150≈250 objects. However, the limits for efficient disambiguation with our technique are still unknown. Perhaps the simultaneous exhibition of so many objects to the user may be more limiting than the precision of input of the mobile device touchscreen. In [8], precision above 60% was obtained on a touchscreen area as small as $\approx 0.6mm^2$. If we transfer this parameter to the *disambiguation canvas*, the whole device touchscreen surface would allow the disambiguation of up to 6,144 objects in one step. Are we able to display meaningful objects in the order of thousands to the user? Is the user able to search for a specific object within such a large group? If this is the case, we also intend to adapt our technique to ensure a reliable selection with one additional step, such as an area selector that points a group of objects within a radius from the thumb position. With this strategy, and taking advantage of the touchscreen precision, we expect to reduce the selectable objects by a factor of at least 10, instead of the factor of 4 used by SQUAD.

Nevertheless, we emphasize that the user's subjective rating assumes that our technique was faster than *ray-casting*, while it had in fact performed slightly slower. This might be a clue to how unpleasant it is to perform a difficult selection with full attention. We have observed that even the breath had to be controlled for some users.

Acknowledgments. We would like to thank CNPq-Brazil for the financial support through projects 311547/2011-7, 485820/2012-9, 302679/2009-0 and 305071/2012-2. Thanks are also due to Microsoft Interop Lab at UFRGS. We finally thank the volunteers who kindly agreed to participate in the experiments.

References

1. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc., USA (2004)
2. Mine, M.: Virtual environment interaction techniques. Technical report, UNC Chapel Hill CS Dept. (1995)
3. Steed, A.: Towards a general model for selection in virtual environments. In: Proceedings of the 3D User Interfaces, 3DUI 2006, USA, pp. 103–110. IEEE Computer Society (2006)
4. Haan, G.D., Koutek, M., Post, F.H.: Intenselect: Using dynamic object rating for assisting 3d object selection. In: Virtual Environments 2005, pp. 201–209 (2005)

5. Kopper, R., Bacim, F., Bowman, D.A.: Rapid and accurate 3d selection by progressive refinement. In: Proceedings of the 2011 IEEE Symposium on 3D User Interfaces, 3DUI 2011, pp. 67–74. IEEE Computer Society, USA (2011)
6. Bacim, F., Kopper, R., Bowman, D.: Design and evaluation of 3d selection techniques based on progressive refinement. International Journal of Human-Computer Studies (to appear, 2013)
7. Baudisch, P., Chu, G.: Back-of-device interaction allows creating very small touch devices. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, USA, pp. 1923–1932. ACM (2009)
8. Debarba, H., Nedel, L., Maciel, A.: Lop-cursor: Fast and precise interaction with tiled displays using one hand and levels of precision. In: 2012 IEEE Symposium on 3D User Interfaces, 3DUI, pp. 125–132 (2012)
9. Dang, N.T., Le, H.H., Tavanti, M.: Visualization and interaction on flight trajectory in a 3d stereoscopic environment. In: The 22nd Digital Avionics Systems Conference, DASC 2003, vol. 2, pp. 9.A.5 –91–10 (2003)
10. Grossman, T., Balakrishnan, R.: The design and evaluation of selection techniques for 3d volumetric displays. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, UIST 2006, pp. 3–12. ACM, USA (2006)
11. Liang, J., Green, M.: Jdcad: A highly interactive 3D modeling system. Computers & Graphics 18(4), 499–506 (1994)
12. Steed, A., Parker, C.: 3D selection strategies for head tracked and non-head tracked operation of spatially immersive displays. In: 8th International Immersive Projection Technology Workshop (2004)
13. Argelaguet, F., Andujar, C.: Visual feedback techniques for virtual pointing on stereoscopic displays. In: Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, VRST 2009, USA, pp. 163–170. ACM (2009)
14. Ogre 3D: Ogre 3D (2012), `http://www.ogre3d.org`
15. Madgwick, S.O.H., Harrison, A.J.L., Vaidyanathan, R.: Estimation of IMU and MARG orientation using a gradient descent algorithm. In: 2011 IEEE International Conference on Rehabilitation Robotics (ICORR), pp. 1–7 (2011)
16. Stellmach, S., Dachselt, R.: Look & touch: gaze-supported target acquisition. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems (CHI 2012), pp. 2981–2990. ACM, USA (2012)
17. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., Mackay, A.: Mid-air pan-and-zoom on wall-sized displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 177–186. ACM, USA (2011)

Full Semantic Transparency: Overcoming Boundaries of Applications

Andrea Kohlhase, Michael Kohlhase, Constantin Jucovschi, and Alexandru Toader

Jacobs University Bremen, Germany
{a.kohlhase,m.kohlhase,c.jucovschi, a.toader}@jacobs-university.de

Abstract. Complex workflows require intelligent interactions. In this paper we attack the problem of combining user interfaces of specialized applications that support different aspects of objects in scientific/technical workflows with semantic technologies. We analyze the problem in terms of the (new) notion of full semantic transparency, i.e., the property of user interfaces to give full access to an underlying semantic object even beyond application lines. In a multi-application case full semantic transparency is difficult, but can be achieved by representing the semantic objects in a structured ontology and actively supporting the application-specific framings of an object in a semantic interface manager. We evaluate the proposed framework in a situation where aspects of technical constructions are distributed across a CAD system, a spreadsheet application, and a knowledge base.

Keywords: Full semantic transparency, multi-application SEMANTIC ALLIANCE, frame shifts, spreadsheets, CAD systems, semantic services.

1 Introduction

As the objects we create and interact with get more and more complex, the corresponding workflows increase in complexity as well. Part of this complexity is mitigated by specialized support systems for specific sub-goals. But this only shifts the underlying problem from mastering complexity to mastering sequences of autonomous steps to smoothly perform a given task. Even though sub-goals might get simpler to reach, individual subtasks often require distinct support systems. In particular, users need to use multiple applications offering support to address specific rather unpredictable aspects of a task. Therefore, this user-driven workflow, we call it "**useflow**", has become a topic of research interest in recent years.

To clarify the issues, let us look at a concrete multi-application useflow, which will serve as a running example for this paper and for which we implemented a fully semantically transparent interface.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 406–423, 2013.

1.1 Running Example

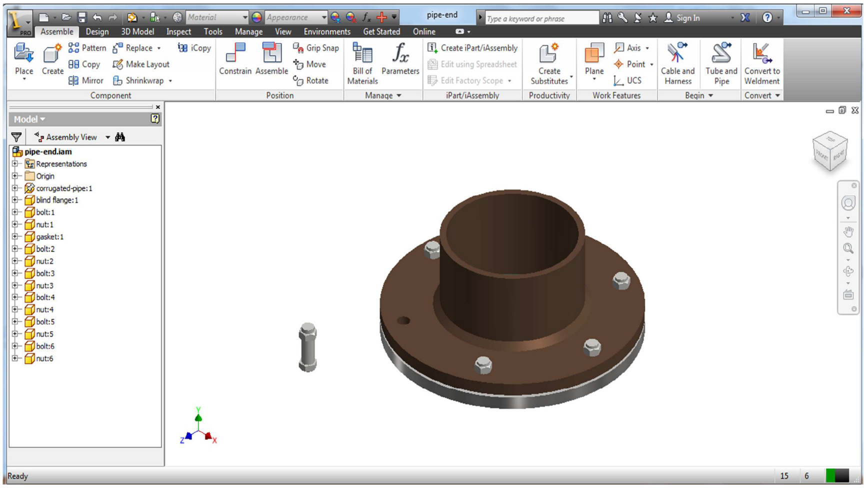


Fig. 1. A Pipe End Representation in a CAD System

Charles A. is a design engineer in a small company that specializes in the production of small and fast luxury yachts. As it is a small company Charles needs to analyze customer needs in interviews, has to create design solutions with a CAD system like Autodesk Inventor, and also decides on production issues.

The design tasks range from the yacht's shape down to pipe fittings (see Fig. 1 – here a blind flange fitting, a particular pipe end that allows allows access to the piping system for revision and cleaning; we will refer to it as a "**pipe end**"[1] in the following).

As customers often set tight budget restrictions, Charles uses several spreadsheets to inform himself about vendors' pricing schemes for frequently needed items as for example pipe ends. In Fig. 2 we see such a spreadsheet for pricing details of pipe ends (e.g. in MS Excel).

G7 fx {= INDEX(INDIRECT("'"&G$5&"'!$F$8:$F$1000");MATCH($A7&$B7&$C7&$D7&$E7;INDIRECT("'"&G$5&"'!$A$8:$

	A	B	C	D	E	F	G	H	I	J	K	L
1	Pricing of Pipe-Ends (Blind Flange Fittings)											
2												
3	Quantity (Pipe Ends)	104										
4												
5							Vendor A		Vendor B		Vendor C	
6	Component	Thread	Color	Head	Type	Quantity per flange	by piece	total	by piece	total	by piece	total
7	bolt	*M15*	*black*	*machine*		*6*	0,297 €	185,328 €	0,336 €	209,664 €	0,340 €	211,848 €
8	nut	*M15*	*black*			*6*	0,499 €	311,351 €	0,480 €	299,520 €	0,485 €	302,640 €
9	gasket				*standard*	*1*	2,020 €	210,038 €	1,920 €	199,680 €	2,910 €	302,640 €
10	flange	*M15*	*black*			*1*	1,069 €	111,197 €	1,344 €	139,776 €	1,071 €	111,372 €
11	blind flange	*M15*	*black*			*1*	0,879 €	91,428 €	0,192 €	19,968 €	0,776 €	80,704 €
13						**Price**	8,74 €	909,34 €	8,35 €	868,61 €	9,70 €	1.009,20 €

Fig. 2. A Spreadsheet for the Calculation of Pipe End Prices

[1] This assembly is chosen for purely expository reasons, nothing in this paper hinges on this choice.

Charles can ‘play with the numbers’ by entering distinct quantities or specifying different attributes of the components of a pipe end to compare prices between vendors. The price conditions for components like flanges and bolts are stored in other worksheets in the same workbook (see example in Fig. 3). In this industrial sector it is customary that vendors offer percentual discounts for sale items depending on the quantity of units to be ordered. Charles’ useflow forces him to switch between the CAD program and the spreadsheet system quite frequently.

Price List of Vendor A					
Conditions depending on Quantity of Pipe End Units					
100	1000	10000	>10000		
100,000%	99,000%	95,000%	90,000%		
Component	Thread	Color	Head	Type	Basic Price
bolt	M15	silver	carriage		0,450 €
bolt	M15	silver	stove		0,460 €
bolt	M15	black	machine		0,300 €
bolt	M15	silver	machine		0,310 €
bolt	M15	red	machine		0,340 €
bolt	M15	black	machine		0,350 €
bolt	M16	black	machine		0,300 €
bolt	M16	black	machine		0,350 €
nut	M15	black			0,504 €
nut	M16	black			0,498 €
gasket				standard	2,040 €
flange	M15	black			1,080 €
flange	M15	silver			1,080 €
flange	M16	black			1,090 €
blind flange	M15	black			0,888 €
blind flange	M16	black			0,888 €
blind flange	M17	black			0,888 €

Fig. 3. A Spreadsheet as Database

1.2 Contribution and Structure of the Paper

We note that application switches in a useflow become necessary to access different aspects of an object currently in focus (e.g. pricing for the currently active CAD assembly). In particular, Charles would like to have the application he switches to already focused on the target object: that object in the target application that represents the respective aspect of the source object (e.g. the price of the type of bolt focused in the CAD system). Normal application switchers (e.g. Alt-Tab) do not provide this quality.

In this paper, we use sense-making theory to better understand this “aspect switching” resulting in a model of “*frame shifts*” and an interface requirement we call “*full semantic transparency*”, i.e., a UI property that provides users access to the full meaning of information objects. Moreover, we present a fully semantically transparent interface. In particular, we strive to support frame shifts via a semantic “mash-up” of user interfaces on a fine-grained object level. Concretely, we make use of an existing semantic interaction framework, which links the semantic objects in applications to concepts in a structured background ontology. These concepts act as a “pivot table” that allow us to predict and focus target objects in frame shifts.

Before we present the interface solution in Section 4, we survey related work (Section 2) and establish a model for frame shifts and develop the concept of “full semantic transparency” (Section 3). Section 5 concludes the paper.

2 Related Work

Generally there are three ways to address useflow smoothing and application interoperability:

U1 Merge applications into a single large system.
U2 Make applications interoperable on the data level.
U3 Make applications interoperable on the user interface level.

Option **U1** is only feasible for very established and high-volume workflows. A good example for this option are the office suites that combine word processors, spreadsheets, slide presenters, and drawing applications under one joint interface. Other examples consist in integrated trade-specific solutions that cover the main workflows of a specific industry sector. The workflows covered by such vertically-integrated systems are usually quite rigid, since any change in workflow needs to be reproduced in the system and thus needs a system extension. Thus, such systems cannot cater to the needs of single users; moreover, Charles' small company could probably not afford one of these large-scale integrated systems.

Option **U2** is facilitated by the trend towards offering open APIs that allow access to its underlying domain and event models. In [7] e.g., Diaz et al. describe a method that uses deep annotation, i.e., an ontology representation, to define the data flow from one portlet to another. Therefore, objects are related semantically on a data level. Most existing approaches though don't pay attention to the resulting user interface challenges, that is a users' access to their full meaning. In [8] this approach is taken up on a conceptual level, stating for example, that a document " will no longer be a mere file, but rather a knowledge base resource", but a closer reading still reveals a similar bias towards machine-understanding. Another approach is the Gnowsis Semantic Desktop framework [18], "a tool for personal information management" [ibd., p. 887]. This collects data from various desktop-based applications in an RDF-based "personal information management ontology" (PIMO), which services can tap e.g. to mash-up data by distinct applications. Even though the data modeling component is similar to what we will present below, the focus of the Semantic Desktop is on information retrieval and semantic information management services (e.g. cataloging or sorting images according to the PIMO) and – to the best of our knowledge – not on UI-level support for useflows, which is what we concentrate on.

Option **U3** has gotten less general attention, except for a call for uniform look-and-feel across applications driven by usability concerns. The underlying reason consists of the idiosyncrasies of user interfaces based on the particular technologies used and tailored task support. User interface mash-ups only seem simple to realize if the applications share the same base technology, e.g. for web applications. But there are many non-web applications supporting subtask achievement that are designed as insular systems that excel in their specific domain.

The coordination of multiple visualizations or *mash-up of visualizations* of information is often used when the "information is sufficiently complex to require different types of visualizations for different aspects or layers" [15, p. 715]. North & Shneiderman note that such visualization requirements tend to be rather unforeseeable. Thus, they devised "Snap-Together Visualization" [ibid.], an approach to dynamically support multiple coordinated visualizations based on a relational database model that generates views, which in turn are linked together through different events. In contrast, Bull models the coordination of visualizations via software engineering modeling approaches [2]. Tudorache et al. note in [21] that such approaches are essentially based on dynamic communication between independent components, but envisioned beforehand by a software designer and not necessarily employable by an application user involved with situated actions [20, 14]. In particular, they are mainly targeting the workflow rather than the useflow.

The unpredictable flexibility requirements for the useflow may be explained by the difference between a user's task environment and his information environment, the background environment that allows users to acquire more information to take better actions (see [16, p. 20]). Pirolli discusses multiple information foraging models of humans, one of which is the "patch model" [ibd., 31ff]. Here, information is assumed to be distributed in a 'patchy' manner. In our running example, Charles finds information needed in his useflow at different locations in distinct applications. Therefore, we can apply the patch model to Charles' useflow. Obviously, an information forager needs to optimize his use of information patches. In which cases or at which steps in the useflow, for instance, should Charles switch applications in order to get price information about pipe ends? If, for example, Charles wants to minimize the patch distance, he makes use of keyboard shortcuts like "Alt Tab" to switch applications. In contrast, if he wants to maximize information gain on pricing issues for pipe ends, he authors a spreadsheet to support him. To keep customers within their applications, companies draw on this trade-off by enriching environments with added-value services like providing piece-lists of assemblies in grids from within a CAD system (with all the disadvantages listed in Option **U1**).

Pirolli shows in [16, p. 38] that reducing between-patch costs improves the overall average rate of information gain and moreover that the optimal gain is achieved by spending *less* time within a patch. Thus, information gain can be improved by mashing up patchy interfaces to reduce between-patch distance and at the same time supporting application-switching. Therefore, in this paper we concentrate on Option **U3** departing from a concept called "semantic transparency" [9], that assumes a user's need for semantic alignments of applications in his useflow, towards "full semantic transparency" based on sense-making theory. In particular, we will apply the notion of "frames" [22] to semantic transparency and propose to address useflow-centered mashing up of user interfaces by semantic technologies.

3 Full Semantic Transparency

If we want to optimize between-patch costs in Charles' information environment, then we have to better understand how humans conceive information. This is a central question in *Sense-Making Theory* (see e.g. [19], [6] for an overview). In [22] Weick introduces the terms "**frame**", "**cue**", and "**connection**":

> Frames and cues can be thought of as vocabularies in which words that are more abstract (frames) include and point to other less abstract words (cues)

> that become sensible in the context created by the more inclusive words. Meaning within vocabularies is relational. A cue in a frame is what makes sense, not the cue alone or the frame alone. Said differently, the substance of sense-making starts with three elements: a frame, a cue, and a connection.

To anticipate trajectories of information (and act accordingly) Charles needs to make sense of the information in his focus. In a nutshell, this means that he continuously strives to understand the potential connections of the information cues he is perceiving with respect to possibly different frames via the UI of applications. Kolko notes the long-term importance of frames in sense-making as a frame can be viewed as “an active perspective that both describes and perceptually changes a given situation” [10].

In the following we suggest a model of sense-making in the process of application switching in UI terms.

Cues via Information Objects Information is carried in a user interface via “**information objects**”. Each application supports the presentation of this information in a different way, particularly via different means. In a CAD system, the assemblies and the parts they are made of are information objects. Consider for instance a hex bolt such as the one on the left of Fig. 1, which is made of a cylindrical shaft and a hexagonal head: while the former is a geometric object the latter two are not. In contrast, in a spreadsheet document cells and tables are information objects: they are mere data containers on a technical level, but their associated properties like the type give their content (e.g. a specific number) a meaning. This meaning of a cell might be implied by other information objects. In particular, the position of a cell within a table consisting of row and column specification carries specific meaning itself. Therefore, we can view information objects with their contents as cues.

Frames via Applications Frames conceptualize knowledge in a way that enables information system stakeholders to make use of the elements of information systems (see [11]). Usually, the terms “frame” and “cue” are employed with respect to people’s cognitive abilities, but they are also expressed in the design of their tools. In particular, we can consider applications as tools that frame information objects. They are manifestations of frames, thus, we can treat them as frames themselves.

Connections via Conventions The information objects in the applications can be seen as constituting a symbolic, domain-specific language for application authors and readers based on common conventions. These can be general conventions like the interpretation of a table as a grid having specific properties or application-dependent conventions like cell content being of a specific type. For making sense of given information the connection between information object and application is established with these conventions.

Sense-Making across Frames via Semantic Objects Frame shifts in a certain useflow assume the existence of a meaningful object that aligns these frames. Charles, for example, looks at the object “bolt” from two viewpoints: under the frame of geometric construction from within a CAD system and under the frame of pricing of a product from within a spreadsheet system (see Fig. 4). To stress that an object carries *several* meanings depending on the perspective it is looked at, this is called "**semantic object**". A semantic object functions as a pivot for application shifting, thus, if we establish a semantic object as a node in an according ontology, this node can serve as a pivot for frame shifts.

For each frame the representation of a semantic object is stored separately in a document. Each application in turn is a "**player**" of certain document-types (i.e., it plays special documents), it *presents* the representations of semantic objects. The "pipe end" e.g. is represented as *.iam file and as *.xlsm file. Note that such representations also include representations of other semantic objects like "bolts" or "flanges", that e.g. in the CAD system Autodesk Inventor are stored as independent part-files (*.ips). In the following we restrict applications to be document players as only those enable application switching as described above.

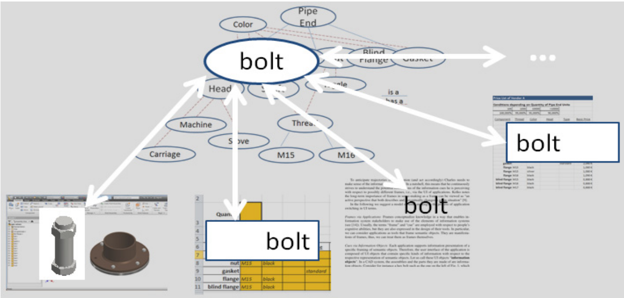


Fig. 4. A Semantic Object as Pivot For Sense-Making

Sense-Making of Frame Shifts via Full Semantic Transparency [9] introduces the term "**semantic transparency**": "a user interface [is called] semantically transparent, if it enables a user to access its semantic objects and their relations via the corresponding UI objects". But what does "access its semantic objects" mean? In [9] it means access to more information about the semantic object. But in our running example we have observed that respective information objects in the CAD system and the spreadsheet application only handle their very own specific frame of the semantic objects Charles has to deal with in his daily work. In particular, access to a semantic object should be extended to mean access to all *available* frames of a semantic object — enabling a "**full semantic transparency**". Even though – as our example above shows – this is exactly what Charles would need, neither of the user interfaces MS Excel nor Autodesk Inventor involved in our running example allow access to both frames of the object under consideration, that is, neither UI is fully semantically transparent.

Full semantic transparency hinges on the sense-making process during frame shifts for a semantic object. As we mentioned above, sense-making has three components: a frame, a cue and a connection. In this paper, we assume that frames are given by particular documents and their players. Moreover, we can view information objects with their contents as cues. But who or what establishes the connection across frames? Charles aligns the applications, that is, his cognition establishes the connection between a frame and a cue, so that a specific aspect of a semantic object can be communicated. This relation can also be called an **interpretation**, i.e., "the assignment of

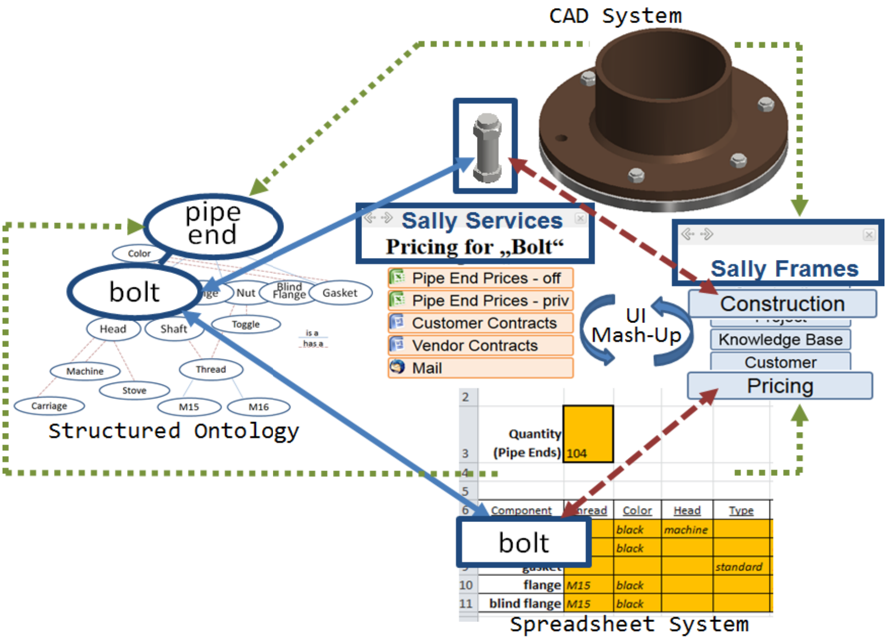


Fig. 5. Mashup of Interfaces

meaning to data" [1], which in turn can be interpreted as subjective extension or enhancement of the given data [10]. In the following we formalize this connection across frames so that we obtain computer support for frame shifts.

4 Realizing Full Semantic Transparency

The very basic idea to reach full semantic transparency consists in the fact that semantic objects can be modeled in structured ontologies. So instead of trying to connect e.g. the information object "bolt" in the CAD system directly with the information object "bolt" in the spreadsheet system, we formally align both by the detour through a structured ontology (see Fig. 5) which acts as a pivot structure. Semantic technologies handling these ontologies can provide intelligent services and user interfaces.

We choose the open source "Semantic Alliance" framework [4], which allows to superimpose semantic services over an existing (and possibly proprietary) application provided that it gives open-API access to user events as a basis for a joining, fully semantically transparent user interface. The central advantage for our purposes is that the interaction intelligence is encapsulated in an autonomous, application-independent semantic ally called "Sally", which remains invisible to the user, but which orchestrates the joint user interface. In particular, information objects of applications are linked to concept specifications of the semantic object in a structured background ontology via an "**interpretation mapping**" which can serve as a model for the sense-making component "connection" between a frame and a cue. In turn we can model frame shifts by aligning information objects via their interpretation mapping to a presentation of a semantic object in an ontology. This way, we can support useflows by extending the SEMANTIC ALLIANCE framework towards integration of applications of different media types.

4.1 Mashing Up User Interfaces

We have implemented the "MA SEMANTIC ALLIANCE" ("MA" stands for "multi-application") framework, an extension of SEMANTIC ALLIANCE which allows to integrate multiple applications. It enables mediating frame switches between applications of different media type in complex useflows.

It has been stressed according to [18, p. 1815] that "sense-makers benefit not just from good sources of information but also from good sources of structure". Moreover, information objects are the UI objects that carry meaning for the user. In particular, clicking those objects is a natural entry-point for access to the semantic object. Therefore, to "unleash […] sense-making potential" [6, p. 36], the Sally-based UI adds interaction elements at two levels:

Sally Frames is a frame selection menu (Fig. 6) which allows users to cognitively shift frames very comfortably without necessarily yet shifting applications. The idea is that the user focuses on an information object in one application, which triggers awareness about other meanings of the underlying semantic object. If the interest in another meaning is large enough, he can choose a frame shift via the menu "Sally Frames" and all available services for an information object under the new frame will be shown on demand via Sally Services. Sally draws on an ontology of frames (including an empty one), which services can register for. Therefore, Sally can gather services with respect to a certain frame.

Sally Services is a service selection menu at the information-object level (Fig. 7[2]). An information object is assumed to be in the user's focus when he left-clicked it, for example the "bolt" in the assembly in Fig. 1. Once he has opted for a frame, the respective "Sally Services", a menu for services currently registered with Sally and available for the semantic object "bolt" in the current frame, replaces the Sally Frames menu. These services provide intelligent user assistance at a very fine-granular level. They can range from a Google search with keyword "bolt" over a definition look-up service in a company-wide knowledge base up to a service in a spreadsheet to calculate prices for the selected bolt in the CAD system (see [4]). Additionally, the respective services can add their own custom interactions, either through custom menus or via popup windows provided by the SEMANTIC ALLIANCE framework.

[2] Note that this layout is still a preliminary proof-of-concept realization that will evolve, for instance, Sally Frames could be realized as a dropdown menu that doesn't occupy as much screen space.

To realize the MA SEMANTIC ALLIANCE, we modeled the abstract document models (ADM; called "abstract document types" in [4], for a concrete example see [12]) for spreadsheets, CAD assemblies, etc. in abstract document ontologies, so that multiple documents and their interpretation mappings can be jointly represented in the form of RDF triples in an in-memory triple store in Sally. Whenever a document is opened in the MA SEMANTIC ALLIANCE, salient aspects of its contents are exported to the triple store together with information about their interpretations. Instead of direct ADM API calls, MA SEMANTIC ALLIANCE services now use SPARQL [17] queries to access the respective information objects. This redesign allows queries across information objects and their interpretations from all open documents. In particular, services can use queries to determine whether they are applicable to a chosen information object (and frame); this information can be used to populate the menus Sally Frames (Fig. 6) and Sally Services (Fig. 7).

Our experiments show that the query-based approach is expressive enough to realize the smooth task-switching capabilities in all studied cases, including our running example which we will now revisit.

4.2 A Use Case with a Fully Semantically Transparent UI

To fortify our intuition about these interface elements on the one hand and to evaluate the expediency of full semantic transparency on the other, let us see how an instance of the MA SEMANTIC ALLIANCE combining a CAD system UI, a spreadsheet system UI, and a knowledge base UI transforms Charles' application-switching useflow from our running example to an interaction model allowing full semantic transparency via frame shifts.

We will go through Charles' new useflow step by step (*italicized text*) and discuss the UI repercussions (normal text).

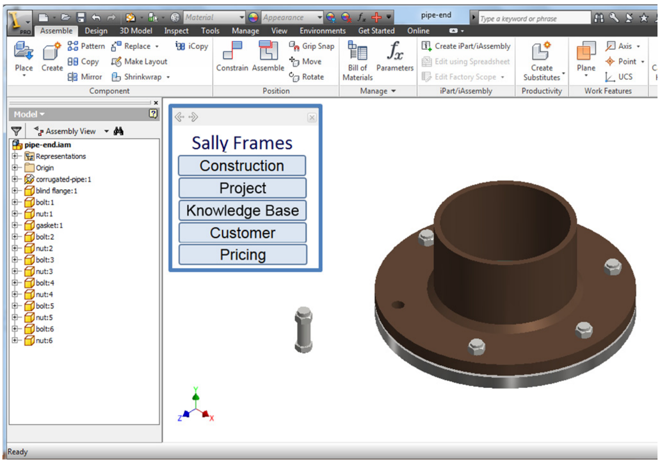


Fig. 6. The Sally Frames Menu for a Pipe End

1. **CAD System UI** *Initially, Charles works on the design of a specific pipe end fulfilling some geometric side conditions like strength and pipe diameter. Satisfied, he wants to verify his design in terms of the customer requirements and senses that his pipe end design may violate financial constraints. He has a closer look at the kind of bolt he used (by clicking on one bolt) and verifies that it is a rather expensive carriage bolt (Fig. 6).* Normally, Charles would have had to switch to his spreadsheet application, load the worksheet in Fig. 2, enter part quantities, and evaluate the result.
    a. **Frame selection** *Now, Charles looks at Sally Frames for the "Pricing" frame.* In this context menu Sally offers the user access to all current frames, here e.g. "Pricing", "Knowledge Base", or "Customer Contracts". *Under the frame "Knowledge Base" he expects for instance services that deepen his understanding for the concept "pipe end", whereas under "Customer" he wants to get information about exact customer requirements given in the contract for pipe ends. But for now Charles selects the "Pricing" frame.* His choice replaces the Sally Frames with the Sally Services menu (containing currently available services for the semantic object "bolt") yielding Fig. 7.
    b. **Service Selection** *Charles studies the Sally Services menu (as seen in Fig. 7).* To determine the available pricing services, Sally needs to know the pertinent information object; as Charles has chosen the "bolt" component, Sally gathers all pricing services available in an extra window overlaying the CAD application. If those services depend on cue *x* like "Give me the definition for *x*", then these services are offered for the content of the selected information object, here for "bolt".

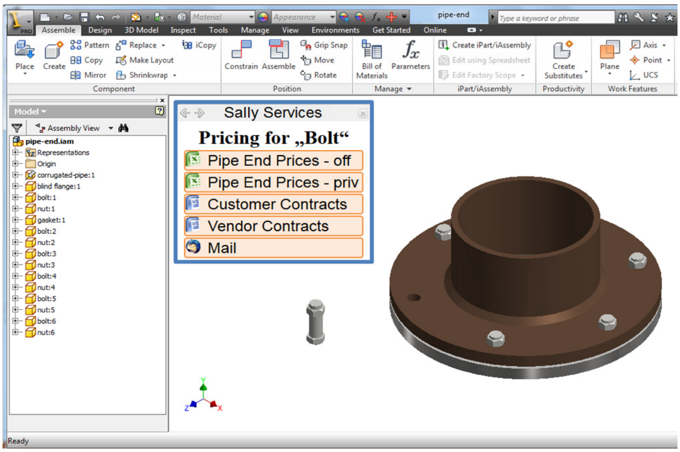
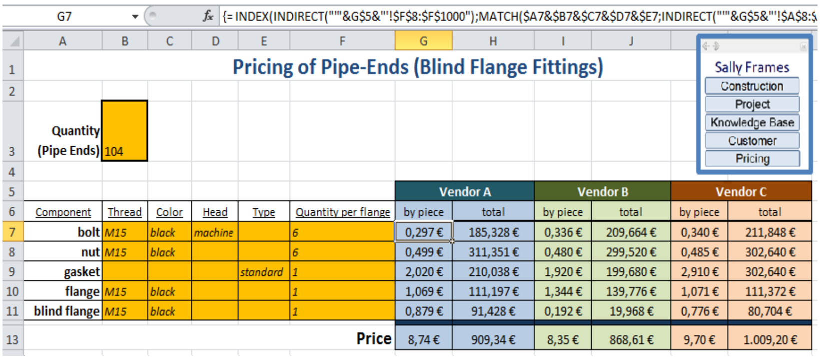


Fig. 7. Sally Services for Selected Bolt under Pricing Frame

Concretely, Charles has the choice between a company-wide, official spreadsheet for pipe end prices, a private one, that he authored himself as a calculation template, the price conditions for the component "bolt" *documented in the customer or vendor contracts for pipe ends, or he can simply check his mail for more informal communications about bolts. Charles opts for the private pricing spreadsheet …*

2. **Spreadsheet System UI** …which orders Sally to open the private "pipe end pricing" worksheet with the resp. properties of the selected bolt, which in turn becomes the top most UI, that is, *Charles switched the application (Fig. 8).* Here, Sally makes use of the interpretation mapping for the Sally Services menu in two ways: it can focus on the pertinent worksheet, since the respective table is mapped to the pipe end via the interpretation mapping, and in particular, the quantity cell to the concept of multiplicity of parts in assemblies. *Charles glances over the values and decides that his suspicion was warranted: the design is over the price range set by the customer.*



G7 {= INDEX(INDIRECT("'"&G$5&"'!$F$8:$F$1000");MATCH($A7&$B7&$C7&$D7&$E7;INDIRECT("'"&G$5&"'!$A$8:$

Pricing of Pipe-Ends (Blind Flange Fittings)

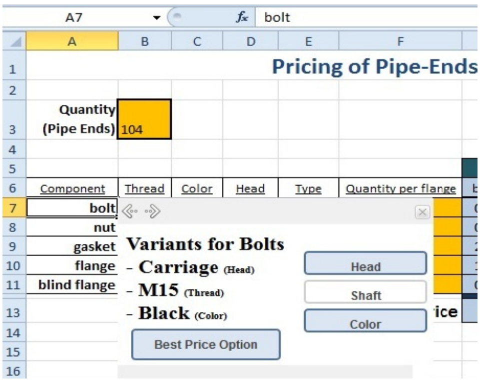
Quantity (Pipe Ends): 104

Component	Thread	Color	Head	Type	Quantity per flange	Vendor A by piece	Vendor A total	Vendor B by piece	Vendor B total	Vendor C by piece	Vendor C total
bolt	M15	black	machine		6	0,297 €	185,328 €	0,336 €	209,664 €	0,340 €	211,848 €
nut	M15	black			6	0,499 €	311,351 €	0,480 €	299,520 €	0,485 €	302,640 €
gasket				standard	1	2,020 €	210,038 €	1,920 €	199,680 €	2,910 €	302,640 €
flange	M15	black			1	1,069 €	111,197 €	1,344 €	139,776 €	1,071 €	111,372 €
blind flange	M15	black			1	0,879 €	91,428 €	0,192 €	19,968 €	0,776 €	80,704 €
					Price	8,74 €	909,34 €	8,35 €	868,61 €	9,70 €	1.009,20 €

Fig. 8. Pipe End Pricing with Sally Frames Menu

a. **Frame Selection** *He tries to reduce the overall price by looking for alternatives for the pricey carriage bolts. Therefore, Charles now selects the frame "Knowledge Base", since he needs information on homologous bolts, and clicks cell* [A7].Sally in turn opens a window next to cell [A7] offering all available knowledge base services for "bolts".

b. **Service Selection** *Here, Charles chooses the "variants" service.*



3. **Variants Service Menu** The variants service uses a complex user interface of its own inside the Sally-originated UI window to determine the full specification of the bolts. In particular, Sally replaces the menu in the window by a list of bolt properties that can be altered without affecting functionality. *This list*

tells Charles that he can use arbitrarily colored bolts with arbitrary head types as long as he stays with the M15 thread.

4. **Direct Interaction** *This variants menu stays on top even though Charles now selects* [D7] *and modifies the head type to "machine"*, which triggers the spreadsheet system to recalculate the numbers.
5. **Variants Service Menu** *The price is already better, but he isn't yet satisfied, so he selects the color entry in the variants menu.* In consequence a temporary worksheet is created with a number of tables that display prices for the different color options for M15 machine bolts as in Fig. 9. *Alternatively, Charles could have chosen the "Best price option", then the input data from the variant with the best price had replaced the original values.*

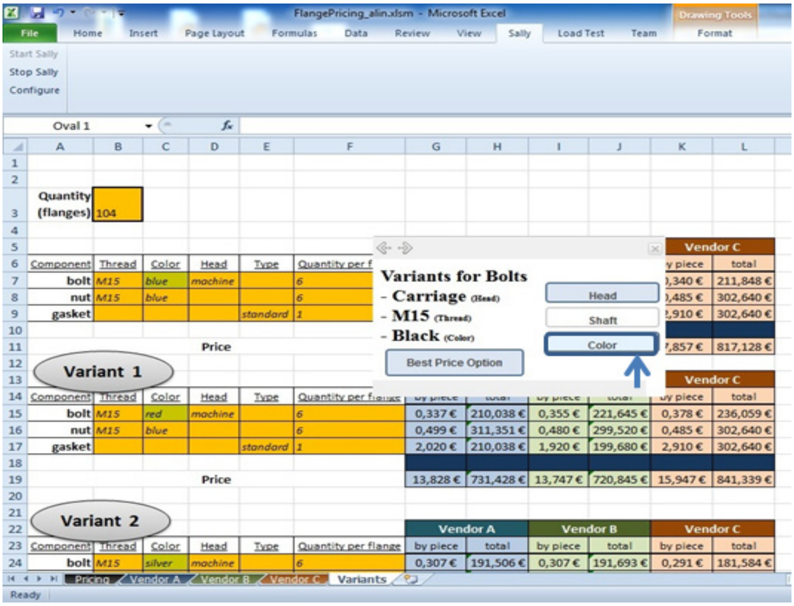


Fig. 9. Variants for Bolts in Terms of Color

We see that the interactions in the Sally-based interface are triggered by the interpretation mapping and the target concepts in the ontology. The advantage of this approach is that some aspects of the user interface can be declaratively specified in the ontology, but the MA SEMANTIC ALLIANCE-based approach also incurs requirements on the ontology and costs, which we will discuss next.

4.3 Requirements for the Ontology

The ontology has to include conceptualizations of the data models of the (information objects in the) applications themselves, e.g. concepts of CAD assemblies that are made of "parts" – which may be geometric objects like nuts or sub-assemblies themselves. Fig. 10 shows the taxonomic part of the ontology used in our example enriched with the "has part" relation (for the structure of assemblies). In our case, the ontology also contains the fact that a bolt is characterized by its head type, thread, and color.

Note that in the development of the ontology, we chose to only represent the object types, not the object instances themselves. For instance, the ontology only has the concept of a machine bolt (and its characteristic dimensions), but not the six individual instances in the pipe end assembly. Similarly, the ontology only specifies the notion of a volume pricing function of an object, but not the actual prices for the objects in the respective assemblies. The instance data stays in the applications themselves: the CAD system for the parts of the assemblies and the spreadsheet for the pricing information. This seems like a generally applicable heuristic for the development of engineering domain ontologies, as this information still justifies data import/export between frames: For the frame switch from the CAD frame to the pricing frame, we only need to know that the six bolts in the pipe end assembly co-reference the M15 machine bolt to export the data into the spreadsheet, which also references it in a functional block that takes the bolt characteristics as inputs.

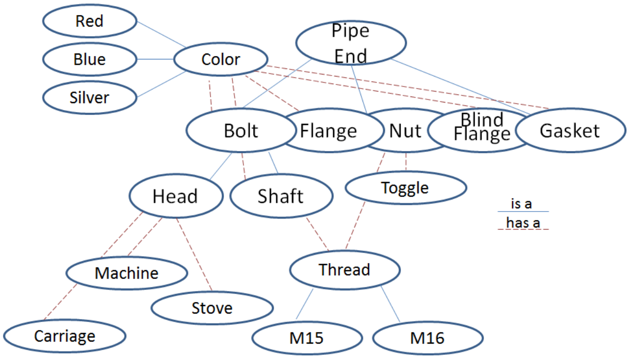


Fig. 10. A Simple Pipe End Ontology

4.4 Cost-Benefit Evaluation

The MA Semantic Alliance architecture consists of a network of semantic services that provide value, but need an interpretation mapping and background ontology to operate. Like any network-based approach, the potential benefits grow in proportion to the size of the network while the costs – often called **network joining costs** – grow with the size of the joining node. Our approach incurs three kinds of joining costs:

a) the applications have to be equipped with a Semantic Alliance API,

b) the domain has to be represented in a structured background ontology that specifies the semantic objects concerned, and

c) each document must be equipped with an interpretation mapping into the ontology.

The costs for *a)* are almost negligible: [4] report that the effort required is in the order of developer weeks, which we can confirm for the API for Autodesk Inventor that was necessary for our example (the spreadsheet APIs already existed). Moreover, the API costs are only incurred once per application and are therefore domain-independent.

The costs for *b)* are non-trivial, and are incurred by every useflow as long as it involves new semantic objects. But note that the costs for establishing an ontology are network costs, since we can assume the ontology to be shared, either globally or in the specific institution (in our example Charles' company). In our experience, large parts of the ontology concern very general topics (engineering parts, measurements, and legal transactions in our example), and can therefore be shared widely. In the future, they might even be outsourced to existing ontologies like DBpedia [5]. But of course, the useflow-specific parts of the ontology have to be hand-coded.

Finally, the costs for *c)* are document-specific, i.e., they are incurred by every document and service to be added to the useflow. Even though these costs can be somewhat mitigated by the interpretation mapping editors under development in the SEMANTIC ALLIANCE context, they have to be taken into account in the evaluation of the framework.

Another kind of costs involved in our approach is the provision and maintenance of services. The current implementation of the MA SEMANTIC ALLIANCE framework restricts the content of the RDF triple store to the contents of already open documents, which enables smooth application switching for a semantic object that is already in the user's focus. A persistent RDF triple store not only could offer more services, it also might allow an integration of MA SEMANTIC ALLIANCE and already existing Semantic Desktop services offered, for example, by the Gnowsis Semantic Desktop framework [18]. The synergies could drive down service development and maintenance costs quite heavily.

5 Conclusion

In this paper, we have addressed the problem of integrating complex useflows spanning multiple applications at the interface level. We have concentrated our investigation on applications that can be described as document players, where the objects addressed are represented in "documents" and the applications give the user access to these objects – "playing the documents", where each application is responsible for different aspects of the objects under consideration. We analyze the underlying problem as one of (lacking) full semantic transparency: Particular useflows are sense-making processes in which an information becomes sensible via frame (here: application), cue (here: information object) and connection (here: interpretation mapping), and switching between distinct applications force frame shifts with respect to the semantic object in the user's focus. In traditional useflows, the connection between the cues remains in the mind of the beholder and cannot be called upon to integrate the respective user interfaces.

We propose to solve that with (more) semantic technologies and materialize the connection – but not as a direct connection between information objects. Rather we make use of an existing technology that establishes a connection between the application's information objects and a semantic object represented in a structured background ontology and uses this mapping to mash semantic services into existing user applications, effectively turning them into user frames. We show that this architecture can be used for the task-switching and information integration requirements of user interface integration on a real-world engineering use-flow involving a CAD system for construction tasks and spreadsheets for calculations.

The point of this paper is not that user interfaces can be integrated at all, but that the integration can be achieved modularly, where new applications (and semantic services) can be integrated flexibly as the useflows warrant. Indeed most business applications opt for integration in a single system (see Option **U1** in the introduction) with the well-known lock-in problem for useflows. Our CAD system (Autodesk Inventor) already supports the integrations of spreadsheets, but this integration is at the data integration level only (Option **U2**).

In our semantic integration we can modify useflows easily by adding frames and services or replacing existing frames and services by compatible ones. For instance the spreadsheet-based pricing service in our example workflow could be replaced by a more scalable OLAP-based one if volumes and resources warrant.

Future Work The most significant shortcoming of the work presented here is that, even though our use cases and experiences with the system make it plausible that smooth application switching will enhance productivity in knowledge-intensive useflows, we have not subjected the system to evaluation with real users. We plan to do this once the system has stabilized further. Then we will also see whether our current assumption that semantic objects in the ontology are sufficient for pivoting between applications, or whether we need to extend pivoting to a two-step operation, which can utilize closely aligned semantic objects in the ontology. Our ontology representation already has candidates for such alignments, but we will have to see which ones can be used for pivoting – clearly not all ontology relations would be adequate.

We can improve the reach of our method and the implementation: For instance, one of the key ingredients of the solution proposed in this paper is the materialization of the connection between information objects across frames via interpretation mappings. This is also one of the largest cost factors in deploying it – assuming that the background ontologies will be shared and thus stabilize for a given domain. To reduce the costs we want to experiment with the powerful background ontology framework the MA SEMANTIC ALLIANCE employs. This provides a notion of "views" (mappings between semantic object representations) that may be used to materialize the semantically interesting parts of interpretation mappings into the ontology by representing information objects in the ontology and establishing ontology views to the semantic objects. This would have the consequence, that the interpretation mapping, i.e., the mapping between an information object in an application to its ontology representation, would be extended by an intra-ontology view which can be shared by the resp. Community of Practice or the institution. In our experience, the necessary views are often already implicitly present in the ontology. Take for instance the bolts in our running example: as they are physical objects, they have a time-space extent (their geometry, which is the aspect modeled in the CAD system) and they are possible goods in financial transactions (the aspect modeled in the pricing spreadsheet). Relying on such views made explicit in the ontology would simplify the interpretation mappings sufficiently that they may be amenable to heuristic methods for automating interpretation mapping creation.

So far, the user's focus is assumed to be on all relevant frames for a semantic object, as the documents have to be open to become available to Sally. With a persistent RDF triple store, not only this limitation would disappear, but new, e.g. collaborative services can be offered. Collected abstract document ontologies can be used for discovery of relevant data sources (e.g. other pricing resources) and smooth application switching to them. Thus, if the synergies with the Gnowsis framework are made use of, then the MA SEMANTIC ALLIANCE framework could even be employed for realizing full semantic transparency for information objects on the Semantic Desktop.

Acknowledgements. We want to thank the anonymous reviewers for their constructive suggestions. This work has been funded by the German Research Council under grants KO-2484-10-1 and KO-2484-12-1.

References

1. Beyer, H., Holtzblatt, K.: Contextual Design. ACM Interactions 6(1), 32–42 (1999)
2. Bull, R.I.: Integrating Dynamic Views Using Model Driven Development. In: Proceedings of CASCON 2006, Riverton, NJ, USA (2006)
3. Jeuring, J., Campbell, J.A., Carette, J., Dos Reis, G., Sojka, P., Wenzel, M., Sorge, V. (eds.): CICM 2012. LNCS, vol. 7362. Springer, Heidelberg (2012)
4. David, C., Jucovschi, C., Kohlhase, A., Kohlhase, M.: Semantic Alliance: A Framework for Semantic Allies. In: [3], pp. 49–64 (2012), http://kwarc.info/kohlhase/papers/mkm12-Sally.pdf
5. DBpedia, http://dbpedia.org
6. Dervin, B.: Sense-Making Theory and Practice: An Overview of User Interests in Knowledge Seeking and Use. Journal of Knowledge Management 2(2), 36–46 (1998)
7. Díaz, O., Iturrioz, J., Irastorza, A.: Improving Portlet Interoperability through Deep Annotation. In: WWW 2005, New York, NY, USA, pp. 372–381 (2005) ISBN 1-59593-046-9
8. Iturrioz, J., Diaz, O., Anzuola, S.F.: Toward the Semantic Desktop: The seMouse Approach. IEEE Intelligent Systems 23, 24–31 (2008) ISSN 1541-1672
9. Kohlhase, A., Kohlhase, M.: Semantic Transparency in User Assistance Systems. In: [13], pp. 89–96 (2009), http://kwarc.info/kohlhase/papers/sigdoc09-semtrans.pdf
10. Kolko, J.: Sensemaking and Framing: A Theoretical Reflection on Perspective in Design Synthesis (2010), http://www.jonkolko.com/projectFiles/preso/kolko_2010_07_sensemaking.pdf
11. le Roux, D.B., le Roux, G.P.: People Frames: The Social Construction of Information Systems. In: CHiMiT 2010, New York, NY, USA, pp. 1:1–1:9 (2010) ISBN 978-1-4503-0447-4
12. Liguda, C.: Modeling the Structure of Spreadsheets. In: LWA 2012, pp. 13–17 (2012)

13. Mehlenbacher, B., Protopsaltis, A., Williams, A., Slatterey, S (eds.): Proceedings of the 27th Annual ACM Int. Conference on Design of Communication (SIGDOC). ACM Press, ACM Special Interest Group for Design of Communication, New York (2009)
14. Nardi, B.A.: Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition, pp. 69–102 (1995) ISBN 0-262-14058-6
15. North, C., Shneiderman, B.: Snap-together Visualization: a User Interface for Coordinating Visualizations via Relational Schemata, New York, NY, USA, pp. 128–135 (2000) ISBN 1-58113-252-2
16. Pirolli, P.: Information Foraging Theory, Oxford Series in Human-Technology Interaction. Oxford University Press (2009)
17. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF, W3C Recommendation World Wide Web Consortium (W3C) (2008)
18. Sauermann, L., Grimnes, G.A., Kiesel, M., Fluit, C., Maus, H., Heim, D., Nadeem, D., Horak, B., Dengel, A.: Semantic Desktop 2.0: The Gnowsis Experience. The Semantic Web 6, 887–900 (2006)
19. Sharma, N.: Role of Available and Provided Resources in Sense-Making. In: CHI 2011, New York, USA, pp. 1807–1816 (2011) ISBN 978-1-4503-0228-9
20. Suchman, L.A.: Plans and Situated Actions: The Problem of Human Machine Communication, Learning in Doing: Social, Cognitive, and Computational Perspectives. Cambridge University Press (1987, 1994)
21. Tudorache, T., Noy, N.F., Falconer, S.M., Musen, M.A.: A Knowledge Base Driven User Interface for Collaborative Ontology Development. In: IUI 2011, New York, NY, USA, pp. 411–414 (2011) ISBN 978-1-4503-0419-1
22. Weick, K.E.: Sense-Making in Organizations. Sage Publications Inc. (1995)

A Comprehensive Study of the Usability of Multiple Graphical Passwords

Soumyadeb Chowdhury, Ron Poet, and Lewis Mackenzie

School of Computing Science, University of Glasgow
{soumc,ron,lewis}@dcs.gla.ac.uk

Abstract. Recognition-based graphical authentication systems (RBGSs) using images as passwords have been proposed as one potential solution to the need for more usable authentication. The rapid increase in the technologies requiring user authentication has increased the number of passwords that users have to remember. But nearly all prior work with RBGSs has studied the usability of a single password. In this paper, we present the first published comparison of the usability of multiple graphical passwords with four different image types: Mikon, doodle, art and everyday objects (food, buildings, sports etc.). A longitudinal experiment was performed with 100 participants over a period of 8 weeks, to examine the usability performance of each of the image types. The results of the study demonstrate that object images are most usable in the sense of being more memorable and less time-consuming to employ, Mikon images are close behind but doodle and art images are significantly inferior. The results of our study complement cognitive literature on the picture superiority effect, visual search process and nameability of visually complex images.

Keywords: usability, user authentication, multiple image passwords.

1 Introduction

Information security systems must permit only legitimate users to gain access to the system and use its resources. This is done by a two step process. *Identification* verifies the user's right to access the system and, once the user is identified, they have to prove their identity by *authentication*. In computer security mechanisms, people are often required to authenticate themselves by using a secret known as a *password or authenticator*. In current practice, alphanumeric passwords are the most widely used mechanism to authenticate users. According to the studies reported in [1, 2] most people find it difficult to remember these passwords. As the number of passwords per user increases, the rate of forgetting them also increases [3]. In order to cope with multiple passwords, users tend to adopt unsafe strategies, which include writing down the passwords, reusing the same passwords and sharing them with others [1, 3].

RBGSs are an alternative type of mechanism where images are used as passwords. The idea is that humans can remember images better than recalling alphanumeric text [4, 5] and so this may be a way of devising more memorable passwords. Given the need for alternative usable authentication systems and existing interest in image

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 424–441, 2013.

passwords [3, 6, 7, 8, 9, 10] as a potential solution, we identify an important limitation of existing work: most prior studies with images as passwords, except [11, 12], have focused on the usability of a single password. However, people will need to remember and use multiple image passwords in the same way that they currently use multiple alphanumeric passwords. The usability of multiple image passwords has not been explored except using faces and everyday objects (flowers, food, sculptures, nature etc.). So it is necessary to conduct experimental studies with other popular image types to compare their usability, when multiple image passwords are used. The motivation of our work is to investigate '*which image type (s) performs best in terms of usability, when multiple graphical passwords are used?*' We believe, unless such large-scale studies are done, image passwords will always look good (usable) on paper but fail in real life and cannot be considered as a viable alternative to alphanumeric passwords.

This paper presents a study with 100 participants, who used multiple image passwords over a period of 8 weeks. The study compares the usability of 4 image types: Mikon, doodle, art and everyday object, when used as graphical passwords. In this paper: (1) we identify the need to study and compare the usability of multiple image passwords using different image types; (2) we show that the image types used as passwords by the participants significantly affect the effectiveness as well as efficiency of RBGSs, when multiple image passwords are used; (3) We discuss the implications of our findings, regarding the attempts to record image passwords motivating a need for future studies examining the vulnerability (guessability) of these passwords to descriptions given by the users.

2 Background Work

Existing studies have reported the usability of faces [3, 7], Mikons [6, 10], doodles [6, 9], abstract art [8] and pictures of daily objects [7]; however, these studies have focused only on the use of single passwords.

We are aware of only two pieces of prior work that studied multiple image passwords: [12] studied the use of multiple facial passwords and [11] compared multiple picture passwords to that of multiple PIN's.

The work reported in [12] studied the effects of frequency of facial password usage, the effects of interference resulting from the use of multiple graphical passwords and the effects of different patterns of access while training with multiple facial passwords. The study also demonstrated that long term memorability of the multiple facial passwords issued by the system is low.

The work reported in [11] compared the memorability of multiple graphical passwords to that of multiple alphanumeric passwords. The images used in this study were photographic and were drawn from different categories such as food, music, sports etc. In the study, there was a single challenge set which contained 4 authenticator images and 6 decoy images. However, this limited set may not be secure and may make the authentication procedure somewhat too easy. In the experimental framework, the participants used the password during training session just after completion of the registration and then three retention tests were carried out with a gap of two weeks between each of them. In such a scenario, the encoding of password information in the long term memory may not be entrenched enough to aid memorability. This is evident from the login success rate reported in the paper, which just after training is almost three times higher than the success rate after 4 weeks. We also find that passwords are issued by the system rather than chosen by the user themselves, and this may hamper the usability of the system.

Given the significant interest of researchers in image passwords, it is clear that additional work is needed using other image types to enhance knowledge of usability, when multiple image passwords are employed. Therefore we designed a usability study to explore the potential of 4 different images types (graphical passwords) Mikon, doodle, art and everyday objects with: (1) Common experimental settings including user tasks for all graphical passwords; (2) Same evaluation parameters; (3) Same design of the authentication system interfaces ; (4) Testing in both conditions i.e. passwords frequently used and less frequently used.

3 User Study Overview

The main aim of the study was to compare the usability of multiple image passwords in RBGSs. A three stage study was designed, which consisted of: (1) a pre-study questionnaire evaluating the participant demographics and current password strategies; (2) an 8-week online study of participants using multiple image passwords; and, (3) a post-study questionnaire regarding participant experiences. In our usability study we used the following image types (See Appendix): (1) Mikon: These are icon-like images which have been drawn by users in studies reported in [6, 10] using a tool called the Mikon engine developed by Mikons.com. (2) Doodle: These images are drawn by a user using pen on paper, in studies reported in [6, 9]. (3) Art: These images were collected from a range of free websites and comprised of paintings from different styles like cubism, abstract, modernism etc. (4) Object: These images comprised of pictures of food and drinks, sculpture and buildings as well as sports and leisure activities, collected from a range of free websites.

3.1 Experimental Conditions and Images

We used independent measures (between subjects) style of experimental design with four conditions (equal number of participants in each condition) namely Mikon (m), doodle (d), art (a) and objects (o). Each participant was assigned to only one of the conditions randomly. The participants in the Mikon condition used Mikon images as their password. They created 4 Mikon passwords and authenticated using them. Each password comprised of 4 Mikon images. So the participants had to remember 16 images in total. Similarly participants in the doodle, art and object conditions used the respective types as passwords. The participants were given a task information sheet which contained the information on the steps to register with the system i.e. select four images to create a password and steps to login to the system after successful registration. However, the participants were neither given any instructions regarding the strategy they should use to select their password images nor the strategy they should employ to remember them.

Each condition was associated with a mock online study website that had a distinct logo and address. Each website had 4 hyperlinks (each link corresponding to one password, See Fig 1 step 1): My jokes; My movies; My news; My status. All hyperlinks except My status had a collection of 150 different images of the same image type (Table 1 and Appendix). My status had a collection of 150 images (50 images each from My jokes, My movies and My news respectively). Each hyper link had a distinct name and background color. The participants could post information/content in the link upon successful authentication: This made sure that the participants had a context to use in differentiating their multiple passwords.

Table 1. Image categories in each link

	My jokes (MJ)-150	My movies (MM)-150	My news (MN)-150	My status (MS)-150
Mikon	Colorful images (No tags)	Black and white. (No tags)	Colorful as well as Black and white images with tag	50 images each, from first three links
Doodle	Black and white (No tags)	Black and white - not same as MJ (No tags)	Black and white with tags	50 images each, from first three links
Art	Abstract paintings	Paintings different from MJ (cubism/ modernism etc)	Paintings different from MJ and MM (cubism/ modernism etc)	50 images each, from first three links
Object	Food and drinks	Sculpture and buildings	Sports and leisure	50 images each, from first three links

3.2 Tasks

Registration

The online study website had 4 links. Each link would correspond to one password (made up of 4 images). The participants could register in each of the links by entering a username and selecting 4 images as their password from the given collection. Figure 1 shows the registration screens of the RBGS developed for our experiment. Each link had a collection of 150 images, presented on the screen as six sets of 25 images in the form of 5X5 grids. The participants could browse from one set to other using the 'change set' button on the web page. The participants could choose all the 4 images from a single set or each image from a different set. The experiment was designed such that each of the participants would use a different collection of the same image type while registering for each of their passwords. For example, each participant in the Mikon condition created 4 passwords, with 1 password selected from the image collection of the first link (My jokes), one from the second link (My movies) and so on. The image archive in each link of the Mikon website consisted of the same image type (Mikon).

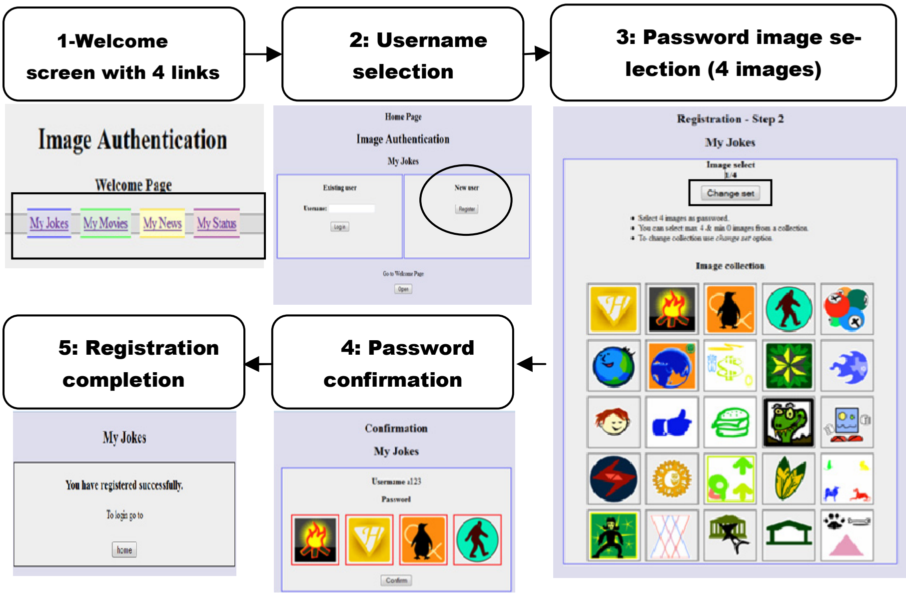


Fig. 1. Steps to register with a password in RBGS

Authentication Procedure

Upon accessing the online study website, the participants had to select the correct password images (the ones they selected during registration) from a sequence of 4X4 grids at each step of a four-step procedure (Figure 2). All the images other than the password image in the sequence are known as *decoy images*. The grid consisting of 15 decoy images and 1 password image is called a *challenge set*. The decoy images for each step were fixed during registration to ensure that the intruder would not be able to guess the correct password image, merely refreshing the web page. The decoy images for each password image were chosen randomly from the collection of 150 images. The decoy images for each of the 4 target images were distinct and never repeated. If the participants at any step during the login procedure selected wrong password image, then they would never get any of their registered password images in the subsequent steps. So in this case, 16 decoy images (without the password image), different from the original challenge set were displayed to the users. The participants were given feedback on the result of the login only after the last step of the procedure. In case of three continuous failed login attempts, the participants were automatically reminded of their password. The reminder was given only for the first 20 login attempts (week-1) of each password. Once the participants logged in with the correct password images, they could post any information which could be seen by others using the system, who could like it.

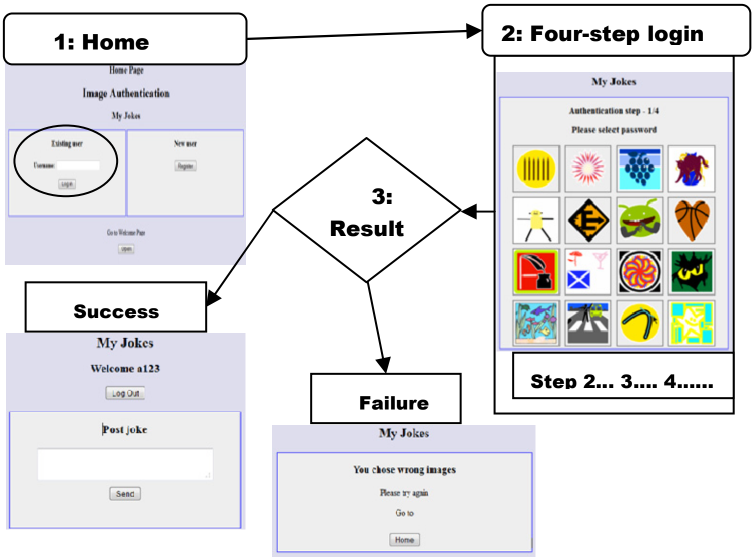


Fig. 2. Steps to authenticate in RBGS

Upon successful authentication, the main task of the participants was to post some content in each of the links. The participants were not forced to use the system and the experimental procedure was designed to allow flexibility in the tasks. All sorts of ethical approvals were taken from to conduct the experimental study.

4 Experimental Framework and Protocol

The essential component of the experimental design is the 8 week online study. We used email prompts to inform the participants about the experimental procedure and tasks. Emails were sent to participants on Day 1, 3 and 5 of each week to notify the progress that they have already made and they are expected to make, to complete the weekly tasks. These emails helped them to keep track of the tasks and made sure they followed the experimental procedures.

There is no standard procedure to design experiments for studying multiple image passwords. In our experiment we tried to vary the frequency of login at regular intervals to simulate a real life scenario. We designed the study, considering the following.

- Ecological validity: participants had a task and a context. A pre questionnaire was conducted for two reasons: (1) to decide the number of multiple graphical passwords to be used for the experiment; (2) to examine, if our participant sample has the same text password behaviour as those of other participants, reported in existing studies [1].

- Confounding variable: the authentication systems (user interface, design), tasks and frequency of login were the same for all conditions.

4.1 Instructions

The participants had to first register with the system. Each participant in each condition had to create 4 passwords, each password comprising of 4 images. Once they finished registration, they had to practice logging in using the password images. The participants were instructed as follows: Week 1: They must create 4 passwords and login 20 times with each of the passwords (total 80 logins in the week). Week 2: They must login 20 times with each of the passwords (total 80 logins in the week); Week 3-4: They must login with each of the passwords 10 times (total 40 logins in the week); Week 5: In this week participants were instructed to login with each of the passwords 2 times (total 8 logins in the week); Week 6: In this week participants were instructed to login with each of the passwords 4 times (total 16 logins in the week); Week 7: In this week participants were instructed to login with each of the passwords twice (total 8 logins in the week); Week 8: In this week participants were instructed to login with each of the passwords thrice (total 12 logins in the week); The participants were instructed to distribute their login sessions over a period of time, instead of finishing them simultaneously. The participants who did not follow the experimental procedure were left out of the experiment.

4.2 Measures – Usability Criteria

We evaluated the usability of the different image types used as password in different conditions using the four measures given below:

1. Effectiveness : This examined the *average/mean successful login percentage (S)* for each of the conditions calculated as,

$$S = \frac{\textbf{Total number of successful login in the condition}}{\textbf{Total number of login in the condition}} \times 100$$

2. Efficiency: It examined the average/mean registration time (R), and average/mean login time of successful login (L). The registration time for each of the passwords is the time taken to go from screen 3 to screen 6 of the registration process as shown in Fig 1. The average registration time (R) for each condition is calculated:

$$\frac{1}{4}\sum_{i=1}^{4} \textbf{Registration time for password } (i), \qquad \textbf{where } i \textbf{ denotes password } 1, 2, 3, 4$$

The login time for a password is the time taken to go from screen 2 to the success notification screen of the authentication process shown in Fig 2. The average time of successful login (L) for each condition is calculated as given below, z represents total number of successful login.

$$\frac{1}{z}\sum_{n=1}^{z} \textbf{Login time for successful login } (\boldsymbol{n})$$

3. Satisfaction: This dimension was assessed from the ratings (1- 5, 1 being highly dissatisfied to 5 being highly satisfied) given by the participants to the different aspects in the post study questionnaire- (sat1) Ease to register; (sat2) Ease to authenticate; (sat3) Meaningfulness/nameability of the image; (sat4) satisfaction with the type of image used as password. These aspects were based on some of the items in SUS (System Usability Scale) questionnaire [13].
4. Stress: This dimension was assessed from the ratings (1- 5, 1 being least stressful to 5 being highly stressful) given by the participants to the different aspects in the post study questionnaire- (str1) level of mental stress; (str2) level of physical stress; (str3) amount of effort required to choose images during registration; (str4) amount of effort required to successfully login. These aspects were based on the items in the NASA Task Load Index questionnaire [14].

5 Results

5.1 Participants

115 undergraduate participants, 30 female and 85 male, of age 20-24 took part in our experiment. They were studying different undergraduate courses: Mechanical Engineering - 22, Electrical Engineering- 19, Aerospace Engineering- 25, Computer Science- 24, Electronics and Communication Engineering- 25. Of the 115 participants who took part, 10 participants had a very low participation rate (did not follow the experimental procedure) and 5 of the participants had to withdraw due to some personal circumstances. So the participation rate was 86.9%.

5.2 Pre Study Questionnaire Results

Given the number and quality of problems associated with alphanumeric passwords, we conducted an online survey to obtain information about password construction techniques and different issues related to them. A total of 150 participants took part in the survey, which included the 115 participants of our usability study. A web based questionnaire was used to obtain data on different aspects of user behavior and perceptions in context of the use of alphanumeric passwords. The framework of the survey was developed using Grounded theory [15]. The framework provided a step by step methodology (Fig 3): (1) identifying the key points of the survey; (2) categorizing the key points depending on the factors influencing them; (3) parameters or concepts to be examined under each category; (4) explanation of the results (qualitative data) to draw conclusions.

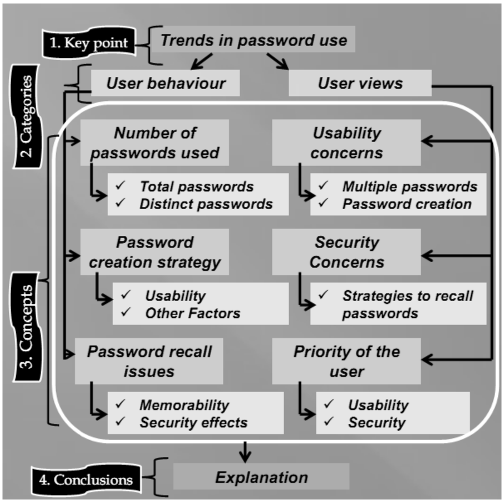


Fig. 3. Grounded theory framework for pre-study survey

The results of the survey revealed that participants used as many as 4-6 passwords in their day to day lives. This result enabled us to make the choice of using 4 graphical passwords with our sample users for the usability study. In context to the password creation strategy, of almost 600 responses: 71.2 % reported using similar passwords for all web accounts; 54.4 % used passwords that could be linked to their personal likings. The users felt that having similar passwords across all accounts aids their memorability. The results also demonstrated that 80% of the participants forget their passwords, either due to the strategy used to aid memorability, or constraints imposed by the system while creating alphanumeric passwords. The results of the survey are in line with the findings of similar research reported in [1]. Thus we can confirm that the sample population used for usability studies has the same password behaviour as reported in other studies. Hence, the sample would represent an accurate reflection of the general population.

5.3 Effectiveness Results

Our first planned analysis compares the effectiveness of the different image types used as password. We analyzed the data from week 2 to week 8. We eliminated the data from week 1, as it was considered a training week, where participants would get used to the system. The dependant variable average/mean login success percentage (for 7 weeks) for each of the conditions (image types) was normally distributed as assessed by the Shapiro-Wilk test. Levene's test indicated that the assumption of homogeneity of variance has not been violated ($F(3, 96) = 2.083$, $p= 0.108$). Given the use of the independent measures (between subjects) experimental protocol with four conditions and the normal distribution of the data, we chose One-way independent measures ANOVA to examine statistical significance. The results of the ANOVA shows that there is a statistically significant difference between the conditions F

(3, 96) = 129.659, p<0.01. These indicate that the type of images used as password by the participants significantly affect the average successful login percentage. The effect size (r = 0.83, large effect) indicates that the effect of the images used as authenticator on the average successful login percent is substantial. The results of the Tukey post hoc tests revealed significant differences between all conditions (p < 0.001 for all tests) except between Mikon and Object (p = 0.059). Thus there was no significant difference between the effectiveness of Mikon and Object images.

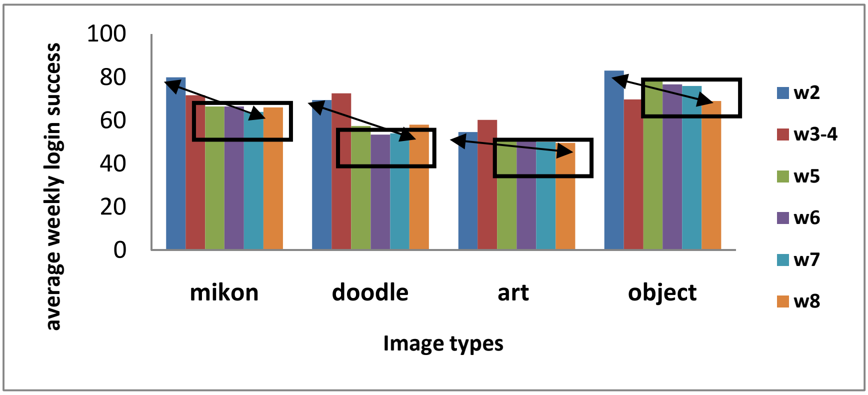


Fig. 4. Weekly login success percentage

We also analyzed the weekly login success percent for each of the image types (Fig 4). We found that the login success percentage for each of the image type falls from week 2 to week 8, as the frequency of usage of the passwords decreases. If we compare the average weekly login success percentage in week 2 and week 8 for each image type (shown by bidirectional arrows in Fig 4), it is found that they fall by 11.44 % in case of Mikon, 12.55 % in case of doodle, 7.74 % in case of art and 14% in case of object. The results of the study show that the average weekly login success percentages for Mikon, doodle and art passwords remain almost the same after week 4 (indicated using black rectangles). In case of object passwords the average weekly login success percentage remains almost same from week 5-7 but drops in week 8. Thus the decrease in memorability for the best performers i.e. Mikon and object are almost the same. Similar characteristic are shown by doodle passwords. In case of art passwords (lowest average login success percent), the difference is comparatively low which clearly suggests that people had problems remembering them in the initial as well as final stages of the study. This reflects that once the participants had used the passwords for a considerable amount of time i.e. at least for a month, memory interference does not hamper the memorability of the image passwords. The memorability of multiple image passwords after a considerable amount of time would depend on the encoding of the passwords in the human memory and frequency of their usage. A two-way ANOVA was conducted with week and image types as the two independent variables. The dependant variable was average weekly login success percent. The week x image interaction was significant, (F (15,576) = 6.102, p < 0.001). This indicated that the average weekly login success significantly varied for each of the image types in different weeks.

5.4 Efficiency Results

Efficiency is a measure of convenience: since a time consuming process will be a barrier, to the repeated use of the authentication system by the user.

Registration Time

The mean registration time of the passwords (in seconds) for each condition is as follows: Mikon (mean: 72.18, SD: 5.48, SE: 1.17), doodle (mean: 75.40, SD: 4.27, SE: 0.88), art (mean: 84.44, SD: 4.91, SE: 0.99), object (mean: 70.61, SD: 3.84, SE: 0.76). We also find that the registration time decreases as users get used to the system in each of the conditions (registration time decreases from p1- first registered password to p4- last registered password) (Fig 5).

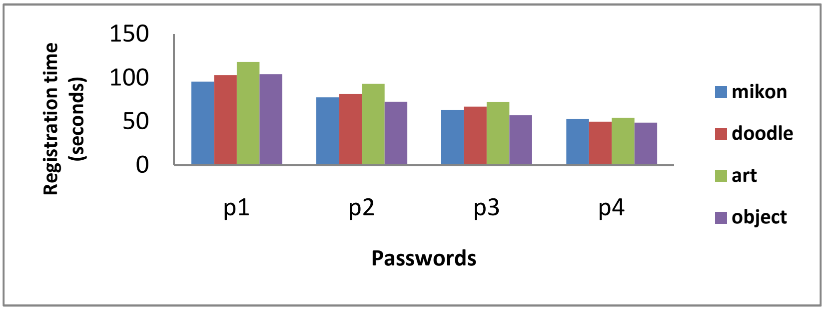


Fig. 5. Registration time for each password created in each condition

The mean registration time (of 4 passwords)) for each of the conditions (image types) was normally distributed as assessed by the *Shapiro-Wilk test*. Levene's test indicated that the assumption of homogeneity of variance has not been violated ($F(3, 96) = 1.968$, $p= 0.127$). We used One-way independent measures ANOVA as the statistical test. The result of the ANOVA shows that there is a significant difference between the conditions ($F(3, 96) = 41.277$, $p<0.001$). This indicates that the type of images used as a password by the participants affect the average registration time significantly. The effect size ($r = 0.78 > 0.5$ represents large effect) indicates that the effect of the images used as authenticator on the average registration time is substantial. The results of the *post hoc* tests revealed significant difference between all groups ($p < 0.05$) except Mikon-Doodle ($p=0.091 > 0.05$) and Mikon -Object ($p = 0.658 > 0.05$). In other words, the differences between the average registration time of Mikon and doodle as well as Mikon and object passwords is not significant.

Authentication Time

The average authentication time of the successful login (in seconds) for each of the conditions is as follows: Mikon (mean: 19.52, SD: 3.60, SE: 0.72), doodle (mean: 22.16, SD: 3.75, SE: 0.75), art (mean: 24.56, SD: 4.8, SE: 0.96), object (mean: 18.28, SD: 2.84, SE: 0.59). The average login time (of 7 weeks) for each of the conditions (image types) was normally distributed as assessed by the Shapiro-Wilk test. Levene's test indicated that the assumption of homogeneity of variance has not been violated (F (3, 96) = 1.791, p= 0.124). The results of the one way ANOVA shows that there is a statistically significant difference between the conditions (F (3, 96) = 13.199, $p < 0.001$). This indicates that the type of images used as password by the participants affect the average login time significantly. The effect size ($r = 0.61 > 0.5$ represents large effect) indicates that the effect of the images used as password on the average login time is substantial. The results of the *post hoc* tests revealed significant differences between all groups ($p < 0.05$) except Mikon-Doodle ($p=0.091 > 0.05$) and Mikon -Object ($p = 0.658 > 0.05$). In other words, the differences between the average login time of Mikon and doodle as well as Mikon and object passwords is not significant.

5.5 Satisfaction Results

According to the box plot in Fig 6, the satisfaction scores of objects range from 13-15 which is better than Mikons ranging from 12-14.5. The doodles have a range of 10.5-12 whereas art have a range of 8-10.5. Hence objects have the best satisfaction score distribution followed by Mikons. But the doodle and art images have inferior satisfaction scores.

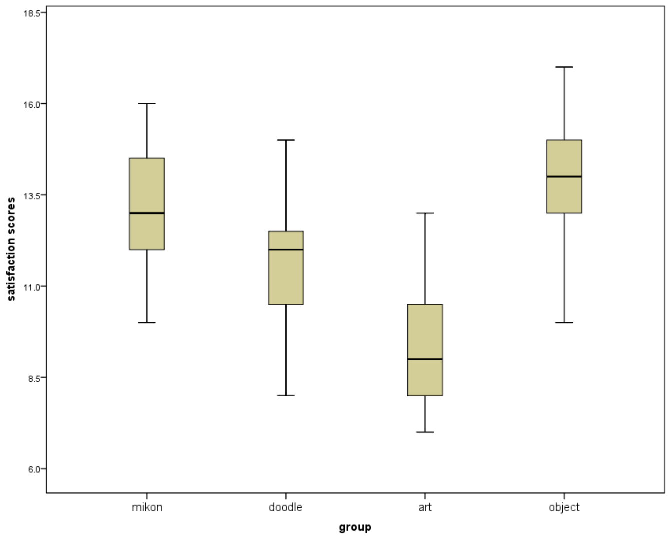


Fig. 6. Box plot for satisfaction scores

Given the ordinal scale of data (user ratings) and the independent measures (between subjects) experimental protocol with four conditions, we used Kruskal-Wallis test to establish statistical significance of data. The result shows that the satisfaction scores for each of the conditions were statistically significant [H (3) = 52.37, p < 0.001). In other words, the satisfaction of the participants was significantly affected by the type of images used as password. We conducted a Mann-Whitney test to follow up our findings by applying a Bonferroni correction, to report all the effects at a 0.008 level of significance. The results reveal that the satisfaction scores were significantly different in all conditions (p < 0.008 for all tests) except for the Mikon-Object (p = 0.156 > 0.008). So we conclude that users are most satisfied with object and Mikon passwords (no significant difference), followed by doodles and least satisfied with art passwords.

5.6 Stress Results

According to the box plot in Figure 7, the stress scores of objects range from 11-13 which is same as that of Mikons. The doodles have a range of 12-14, whereas art have a range of 13-16. Hence the art images have the highest stress score distribution closely followed by doodles. The object and Mikon images have the lowest stress score distribution.

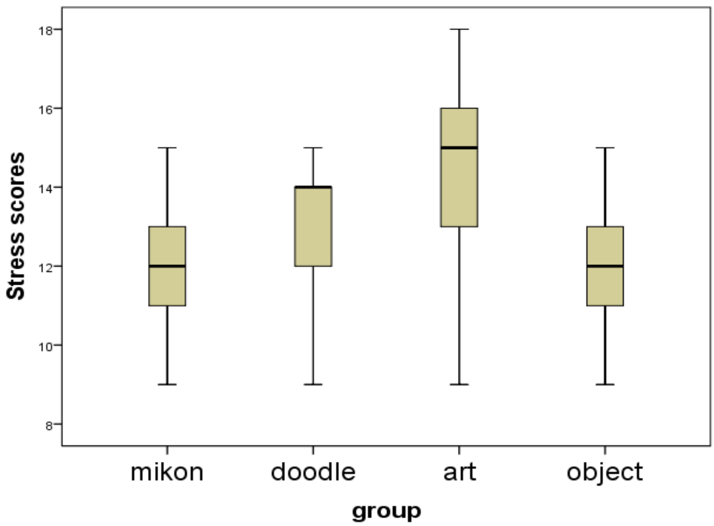


Fig. 7. Box Plot for stress scores

Given the ordinal scale of data (user ratings) and the independent measures (between subjects) experimental protocol with four conditions, we used Kruskal-Wallis test. The result demonstrates that the stress scores for each of the conditions were statistically significant [H (3) = 23.314, p < 0.001). In other words, the stress scores given by the participants were significantly affected by the type of images used as password. We conducted a Mann-Whitney test to follow up our findings by applying a Bonferroni correction, to report all the effects at a 0.008 level of significance. The results show that the stress scores were significantly different in all conditions ($p < 0.008$ for all tests) except for Mikon-Object ($p = 0.32 > 0.008$). Hence we conclude that art images are most stressful to use, followed by doodles, whereas Mikons and objects (no significant difference) are least stressful to use.

5.7 Post Study Questionnaire Results

Rate Images: In the post study questionnaire the participants were asked to rate the image types on a scale: will use, not sure and never use. We find most of the Mikon (17/25) and object (20/25) users would like to use these images as passwords in the future. The art users (17/25) disliked these images to be used as password and the doodle users had a split opinion.

Strategy Used for Password Creation: The participants were asked to provide information on the strategy/method they used to create their passwords (Fig 9). The results reveal that most Mikon and doodle users either used a story/pattern to remember their passwords or they chose passwords according to their personal likings (Fig 8). Most art users chose passwords either based on their personal likings- favorite color, objects, scene or visual and aesthetic quality of the images i.e. attractiveness. But, these strategies for creating memorable passwords may either make them guessable to an intruder who knows the user quite well, or could be disclosed and thus shared with ease.

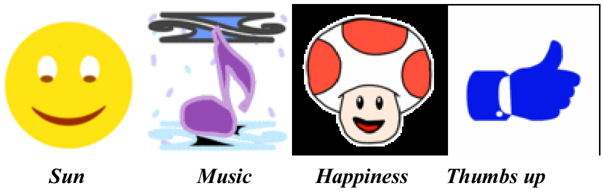


Fig. 8. A Mikon password created using pattern strategy

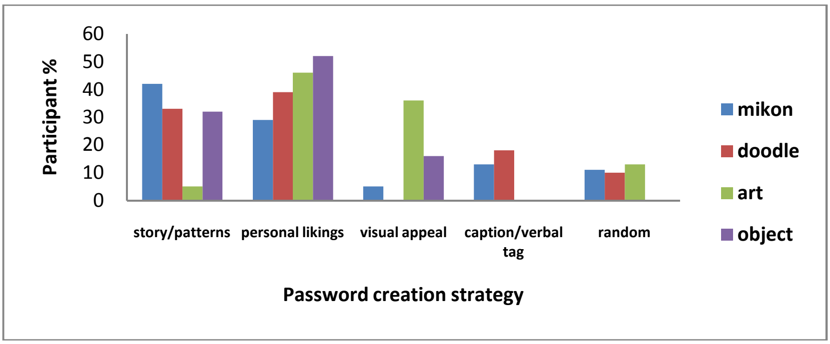


Fig. 9. Strategy employed to create image passwords

Recording Passwords: None of the participants said they made an attempt to record their passwords but almost 84 % of the participants who used art passwords, 68 % of doodle users, 36% of Mikon users and 32% of object users reported that they might use screen captures, sketches or notes (written descriptions) to store their image passwords.

6 Discussion

This is the first study that compared the usability of multiple image passwords using 4 different images types- Mikon, doodle, art and objects in RBGS. The use of different experimental framework, dependant variables and image types in [11, 12] makes it difficult to allow systematic comparison of our results with them.

Our findings show that the effectiveness of graphical passwords is significantly affected by the type of images used. In this context, the results show that the mean login success percentage is highest for objects, closely followed by Mikons, then doodles and lowest for art images. According to the cognitive studies, dual coding theory [5] and guided search process [16], an elaborative encoding of an image in human memory makes it memorable. Thus, an image which is easily associated with a name (nameable) or can be interpreted in a meaningful way is likely to be more memorable due to superior encoding in human memory. In this context, mean scores of the sat 3 parameter (meaningfulness of the image) is: highest for objects (3.65/5); closely followed by Mikon (3.5/5); then doodles (2.8/5); lowest for art images (2.14/5). Thus we find that the results of the satisfaction parameter are in line with the mean successful login percentage. Hence the higher memorability of the object and mikon images can be attributed to the fact that users find these images meaningful and easy to associate with something. The doodle images are black and white line drawings and do not convey much meaning to aid memorability. Hence these images may not be encoded in an elaborate way in the human memory. The art images were very difficult to remember because according to the users, it was not only difficult to associate them with something meaningful but they were visually complex, containing a lot of information and color which would lead to information overload in memory. This complements the work reported in [17] which suggested that visual complexity of an image is linked to the ease of associating it with a name. The work has indicated that it is difficult to assign names to visually complex images. The results also reveal that the mean registration time is: lowest for objects; closely followed by Mikons; then doodles; highest for art images. In this context, the mean scores of the sat 1 parameter (ease to register) are: highest for objects (3.24/5); closely followed by Mikons (3.12/5); then doodles (2.85/5); lowest for art images (2.45/5). Thus the qualitative data obtained from the participants through the questionnaire complement the mean registration time obtained from the online study. These results can be attributed to the fact that users find it difficult to choose meaningful images in case of doodle and art, which they could use as passwords. The authentication time follows the same trend as that of the registration time: lowest for objects; closely followed by Mikons; then doodles; highest for art images. The mean scores for the sat-2 parameter (ease to authenticate): highest for objects (3.56/5); closely followed by Mikons (3.36/5); then doodles (2.95/5); lowest for art images (2.45/5). Thus the mean scores of the sat 2 parameter complement the results of the login time data obtained from the online study. The above discussion suggests that the effectiveness and efficiency results complement each other. So we can conclude that images which are meaningful or can be associated with something easily are: effective in the sense of being memorable; efficient i.e. less time consuming to employ. This conclusion is also supported by the mean satisfaction scores obtained through the questionnaires: highest for objects (13.91/20); closely followed by Mikons (13.16/20); then doodles (11.8/20); lowest for art images (9.24/20) and mean stress scores: lowest for objects (11.87/20); closely followed by Mikons (12.2/20); then doodles (13.12/20); highest for art images (14.66/20).

7 Conclusions and Future Work

We have presented the first study with 100 participants over a period of 8 weeks to compare the usability: mean success percentage for 7 weeks; mean registration and login time; the qualitative data i.e. participant's opinion (satisfaction as well as stress), of multiple image passwords using 4 different image types- Mikon, doodle, art and objects. The results of the study revealed that object and Mikon passwords performed best in each of the usability criteria compared to doodle and art passwords. So we conclude that meaningful or easily nameable image types are most usable when multiple graphical passwords are used. The experimental design in our study is: valid, it answers our research question through the data we collected for each measure; reliable, it can be reproduced by the research community; most importantly, such a study for the stated research problem has not been conducted in the past. In terms of improvement, the same study could be reproduced with a user-group other than students. The post study questionnaire results demonstrate that most people chose password images either by making a pattern/story or something which is related to them. These results underscore the need to examine, whether passwords created by using patterns aid memorability, when multiple graphical passwords are used and assess the ease of employing such a strategy these different image types. The results suggest that though meaningful images would aid usability when multiple graphical passwords are used, users may engage in insecure coping mechanisms like recording them through digital or non digital media. So in our ongoing work, we are investigating the vulnerability of image passwords to user descriptions (non digital attempt to describe the images by writing them or verbally communicating them). We also find that the statistical analysis alone does not unambiguously identify the most suitable image type to be used for graphical password. In our ongoing work, we are developing a framework that would help quantify the usability value for each image type, taking into account qualitative as well quantitative data obtained for all criteria from the experiment.

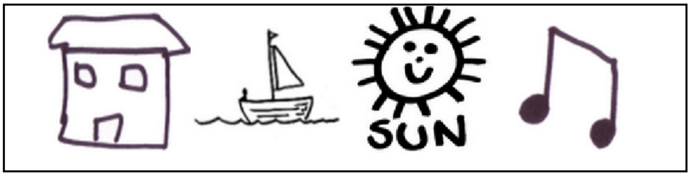
Acknowledgement. We are thankful to SICSA (Scottish Informatics and Computer Science Alliance) for funding this research. We are also thankful as well as grateful to Dr. Balvinder Shukla (Vice Chancellor Amity University, India) for helping us get the participants for the study and permitting us to perform the experiments with them.

References

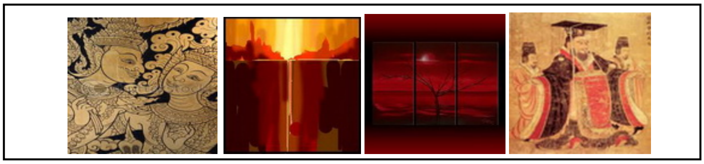
1. Adams, A., Sasse, M.: Users are Not the Enemy. Communications of the ACM, 40–46 (1999)
2. Florencio, D., Herley, C.: A large-scale study of web password habits. In: Proceedings of 16th International Conference on World Wide Web, University of Calgary, pp. 657–666 (2007)
3. Brostoff, S., Sasse, M.A.: Are passfaces more usable than passwords? A field trial investigation. In: Proceedings of HCI on People and Computers XIV, pp. 405–424 (2000)
4. Madigan, S.: Picture Memory Imagery, Memory, and Cognition: Essays in Honor of Allan Paivio. In: Yuille, J. (ed.). Lawrence Erlbaum Associates, Hillsdale (1983)
5. Paivio, A.: Mental Representataions: A Dual Coding Approach. Oxford Press, UK (1986)
6. Chowdhury, S., Poet, R.: Comparing the usability of doodle and Mikon images to be used as authenticators in graphical authentication systems. In: Proceeding of Conference on User Science and Engineering, pp. 54–58 (2011)
7. Davis, D., Monrose, F., Reiter, M.: On user choice in graphical password schemes. In: Proceedings of the Conference on USENIX Security Symposium, vol. 13, p. 11 (2005)
8. Dhamija, R., Perrig, A.: Déjà vu: A user study using images for authentication. In: Proceedings of the Conference on USENIX Security Symposium, vol. 9, p. 4 (2000)
9. Renaud, K.: A Visuo- Biometric Authentication Mechanism for Old Users. In: Proceeding of British HCI, pp. 167–182 (2005)
10. Renaud, K.: Web authentication using Mikon images. In: Proceedings of World Congress on Privacy, Security, Trust and the management of E-Business, pp. 1–10 (2009)
11. Moncur, W., Leplâtre, G.: Pictures at the ATM: exploring the usability of multiple graphical passwords. In: Proceedings of SIGCHI Conference on Human Factors in Computing Systems, pp. 887–894 (2007)
12. Everitt, K., Bragin, T., Fogarty, J., Kohno, T.: A comprehensive study of frequency, interference and training of multiple graphical passwords. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 889–898 (2009)
13. SUS questionnaire (2013), `http://www.usabilitynet.org/trump/methods/satisfaction.htm` (accessed January 19, 2013)
14. NASA Task Load Index, `http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf` (accessed January 19, 2013)
15. Strauss, A., Corbin, J.: Basics of qualitative research: Grounded theory procedures and techniques. Sage, Newbury Park (1990)
16. Wolfe, M.: Guided search 2.0 a revised model of visual search. Psychonomic Bulletin & Review 1(5), 202–238 (1994)
17. Szekely, A., Bates, E.: Objective visual complexity as a variable in picture naming, University of California. CRL Newsletter Center for Research in Language, pp. 3–33 (1999)

Appendix

Sample Mikon password



Sample doodle password



Sample art password

Sample object password

Security for Diversity: Studying the Effects of Verbal and Imagery Processes on User Authentication Mechanisms

Marios Belk[1], Christos Fidas[1,2], Panagiotis Germanakos[1,3], and George Samaras[1]

[1] Department of Computer Science, University of Cyprus, 1678 Nicosia, Cyprus
[2] Interactive Technologies Lab, HCI Group, Electrical and Computer Engineering Department University of Patras, 26504, Patras, Greece
[3] SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
{belk,christos.fidas,pgerman,cssamara}@cs.ucy.ac.cy

Abstract. Stimulated by a large number of different theories on human cognition, suggesting that individuals have different habitual approaches in retrieving, recalling, processing and storing verbal and graphical information, this paper investigates the effect of such processes with regard to user performance and preference toward two variations of knowledge-based authentication mechanisms. In particular, a text-based password authentication mechanism and a recognition-based graphical authentication mechanism were deployed in the frame of an ecological valid user study, to investigate the effect of specific cognitive factors of users toward efficiency, effectiveness and preference of authentication tasks. A total of 145 users participated during a five-month period between February and June 2012. This recent study provides interesting insights for the design and deployment of adaptive authentication mechanisms based on cognitive factors of users. The results and implications of this paper are valuable in understanding and modeling user interactions with regard to authentication mechanisms.

Keywords: User Authentication, Cognitive Factors, Efficiency, Effectiveness, Preference, Usable Security, Diversity, User Study.

1 Introduction

The World Wide Web has become gradually a platform for deployment of complex applications of increased interactivity, as it takes the form of a medium used for complex and important tasks including commercial and governmental transactions, collaborative work, learning and information retrieval. Within this realm, security and privacy issues of interactive systems are considered of paramount importance as it is known that the consequences of a security breach can harm the credibility and legal liability of an organization, decreases users' trust and acceptance, while it exponentially increases maintenance and support costs. In this context, one of the most important and challenging issues is to support users, engaged on tasks related to authentication, through usable computer-human interface designs [1-8].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 442–459, 2013.

Nowadays, user authentication over the Internet is achieved primarily with the use of text-based passwords. It is estimated that more than 80% of US and UK companies apply some form of text-based password authentication; in many cases it is their solely method for user authentication [9]. Despite this familiarity factor of users with regard to text-based authentication mechanisms, a considerable amount of research has focused on the design and implementation of graphical authentication schemes. Graphical authentication mechanisms claim to preserve security and improve usability and memorability, as they leverage the vast capacity and capabilities of the human visual memory system [10, 11, 12]. Principally, graphical authentication mechanisms require from a user to enter an authentication key represented by images in a specific sequence. A recent comprehensive review of graphical authentication schemes [13] revealed that numerous ideas have been proposed over the last decade which focus on graphical authentication, and how to address the security and usability issues of text-based passwords, however, few schemes exist that deliver on the original promise of addressing the known problems with text-based passwords since many of the same problems continue to exist. For example, in an early study of Brostoff and Sasse [14], which compared the usability of a text-based password and a recognition-based, graphical authentication mechanism (Passfaces [15]), results demonstrated that although fewer login errors were recorded in graphical authentication than in text-based password interactions, they tended to log in less frequently because the login process was not as efficient as through the text-based password mechanism. A more recent study of Everitt et al. [16] revealed a dramatic decline in memorability and performance when using multiple graphical authentication keys. Also, from a security perspective, the greatest concern of graphical authentication mechanisms is guessing attacks since usable graphical authentication keys usually have a weaker level of security strength than traditional text-based passwords. A study of Stobert et al. [17] indicated that increasing the required authentication key length in graphical authentication mechanisms negatively affects the memorability of the authentication key, and thus, the overall usability of the authentication process.

Graphical authentication schemes can be classified into the following three categories: *Recall-based* authentication mechanisms require that users remember information and reproduce a secret drawing on a static image as their authentication key. Examples include Draw-a-Secret (DAS) [18], its variations [13], YAGP (Yet Another Graphical Password) [19], and Pass-Go [20]. *Cued-recall* authentication mechanisms require users to identify specific locations on a static image and are intended to reduce the memory load on users, since specific cues are utilized in order to assist the recall of information. Examples include PassPoints [21], Persuasive Cued Click Points [22] and gaze-based authentication [23]. *Recognition-based* authentication mechanisms require that users create an authentication key by selecting and memorizing specific images, and then recognize the images among decoys to authenticate. Examples include Passfaces [15], ImagePass [24] and Tiles [25].

A variety of studies have been reported that underpin the necessity for increasing usability of authentication mechanisms [1-8]. The literature reveals many proposals, such as educating and influencing users to create more secure authentication keys [26], improving existing recall-based password approaches with recognition of text

[27], enforcing the creation of secure authentication keys through policies [2, 3], automatically generating secure authentication keys and mnemonic passphrases [28]. Furthermore, password managers [29, 30] have been proposed to minimize users' cognitive load. Nevertheless, ineffective practice of usability in user authentication does not naturally embed the users' characteristics in the design process, and usually adopts a "one-size-fits-all" approach when concerning user authentication designs, ignoring the fact that different users have different characteristics and develop different structural and functional mental models, and thus need individual scaffolding. In this respect, supporting usability of user authentication mechanisms with user-adaptive technologies [31] is based on the promise that understanding and modeling human behavior in terms of structural and functional user requirements related to such security tasks can provide an alternative to the "one-size-fits-all" approach with the aim to improve the system's usability and provide a positive user experience.

Consequently, a first step toward designing an adaptive user authentication mechanism is to identify which individual characteristics (e.g., knowledge, previous experience, lingual characteristics, cognitive characteristics, etc.) are considered important for adapting such mechanisms. Bearing in mind that human computer interactions with regard to authentication mechanisms are in principal cognitive tasks that embrace to recall and/or recognize, process and store information, we argue that these interactions should be analyzed in more detail under the light of cognitive theories. Accordingly, the purpose of this paper is to investigate whether there is an influence of specific individual characteristics of users targeting on cognitive factors, toward efficiency, effectiveness and user preference of authentication mechanisms.

The rest of the paper is structured as follows: next we present the underlying theory of this work. Furthermore, we describe the context of an empirical study, methods, and developed hypotheses. Thereafter, we analyze and discuss our findings. Finally, we summarize our paper and outline the implications of the reported research.

2 Theoretical Background

One of the most widely accepted theory of human cognition is the Dual Coding Theory [32, 33]. It suggests that visual and verbal information is processed and represented differently and along two distinct cognitive sub-systems in the human mind; the visual and verbal cognitive sub-systems. Each sub-system creates separate representations for information processed which are used to organize incoming information that can be acted upon, stored, and retrieved for subsequent use.

Many psychology studies have reported that pictures are better recognized and recalled by the human brain than textual information, referred as the picture superiority effect [34, 35, 36]. Paivio's Dual Coding Theory explains the picture superiority effect that pictures are more perceptually rich than words which lends them an advantage in information processing. The picture superiority effect might be explained by the fact that pictures are mentally represented along with the features being observed, whereas text is visually sparse and represented symbolically, where symbols might have a different meaning depending on the form of the text, which requires an additional processing for the verbal sub-system. For example, 'X' may represent the Roman numeral 10, or the multiplication symbol [13].

The advantage of the picture superiority effect has led to the design of graphical authentication schemes as a viable alternative to text-based authentication mechanisms aiming to increase efficiency and effectiveness of interactions. In graphical authentication mechanisms, human memory is leveraged for visual information in hope of a reduced memory burden that will facilitate the selection and use of more secure authentication keys [13]. Given also that the current user authentication paradigm expands from login on single desktop computers, to login on smart devices, the added difficultly of typing text-based passwords on small screens has amplified the attention to graphical authentication mechanisms as an alternative to traditional text-based password schemes.

However, several research findings claim that the picture superiority effect does not always hold and is affected by various factors [36, 37, 38]. Oates and Reder [37] claim that the picture superiority effect only occurs when a picture affords a meaningful textual label that discriminates it from other pictures. Results of these studies reveal that abstract pictures are not memorable as single words since the visual stimuli is difficult to identify, and hence, a generation of a consistent textual label is not easy or possible. Robertson and Köhler [38] have further provided evidence that the ability to label a picture affects its processing and memory. Another recent study of Mihajlov and Jerman-Blazic [24], that assessed the memorability rates of abstract, face and single-object images in recognition-based, graphical authentication mechanisms, supports that single-object images have the highest memorability rates in graphical authentication mechanisms, since single-object images can be easily labeled (e.g., "picture of a football"), and thus processed and remembered.

As an effort to explain the aforementioned empirically observed differences in users' mental representation and processing of information, many researchers have developed theories of individual differences in *cognitive style* from the perspective of dual coding theory [39], and consequently, argue that individuals have differences in the way they process and remember information. In particular, individuals may process and organize verbal information more efficiently than visual information, whilst others the opposite [39-44]. Within this context, a number of psychometric tests and questionnaires have been developed that elicit verbal-imagery cognitive style, mainly through self-reported experiences and preferences, and response times on verbal vs. visual aptitude tasks (OSIVQ questionnaire: [41]; VICS test: [42]; CSA test [43]; see also [39, 44] for a review on older questionnaires and psychometric tests). Self-reported questionnaires usually ask the participants to rate their preference toward a verbal versus visual mode of processing. Example ratings would be "I have a photographic memory" or "My verbal skills are excellent" [41]. For the reason that questionnaires showed relatively low internal reliability and poor predictive validity [42, 45], objective measures through the development of psychometric tools have emerged, such as response time in solving cognitive tasks that require verbal or visual processing. In particular, psychometric tools have been proposed that typically require from the participant to provide an answer to text-based or image-based statements. Depending on the response time of each answer, the ratio of means or medians between the verbal and visual statements is computed and further used to classify the participant to a particular group; Verbal or Imager group.

To this end, with the aim to investigate the relation among individual differences in cognitive style, and user authentication mechanisms in terms of efficiency, effectiveness and preference, we next describe an empirical study which entailed a psychometric-based survey for extracting users' cognitive styles, combined with a real usage scenario with two variations of user authentication mechanisms; text-based password and recognition-based, graphical authentication mechanism. Such an endeavor is considered valuable as it entails investigating and modeling human behavior in the context of user authentication tasks. Understanding structural and functional user mental models assists the deployment of usable computer-human interaction designs and workflows, whilst minimizing user cognitive loads, perceptual and learning efforts and erroneous interactions with regard to authentication mechanisms.

3 Method of Study

3.1 Procedure

A Web-based environment was developed within the frame of various university courses. During each course enrolment process, students were required to create their authentication key that was used for accessing the courses' material (i.e., course slides, homework exercises) and for viewing their grades. The experimental procedure has been divided in three phases:

- **Phase A:** The type of authentication (text-based password or graphical mechanism) was randomly provided during the enrolment process. At the end of the enrollment process the sample was divided of half of the students having enrolled with a text-based password and the other half having enrolled with a graphical authentication mechanism. The users' interactions during this phase were recorded for a period of one and a half month.
- **Phase B:** After Phase A, the system altered the students' authentication type; students that had enrolled with a text-based password in Phase A were prompted to create a new graphical authentication key and vice versa. The new authentication key would be used for the same period as in Phase A (one and half month). The main aim of Phase B was to engage the whole sample on both authentication mechanisms for the same period of time.
- **Phase C:** After Phase B, the system gave the users the option to choose the preferred authentication type to access the system until the end of the study. In particular, the students were first asked to choose between the two variations of authentication (i.e., text-based password or graphical) for which they had already used in Phase A or Phase B during the semester, and then entered the preferred authentication key. Aiming to avoid the effect of users' familiarity with the previously (Phase B) used authentication mechanism, with regard to preference, the system was available for open access during a one month period (without the presence of any authentication mechanism, the users could download the learning material without to authenticate themselves).

Controlled laboratory sessions were also conducted throughout the period of the study to elicit the users' cognitive styles through a psychometric test. With the aim to apply the psychometric tests in a scientific right manner, we conducted several sessions with a maximum of five participants by following the protocol suggested by the inventors of the psychometric tests.

3.2 Users' Cognitive Style Elicitation

Users' cognitive styles where elicited by exploiting Riding's Cognitive Style Analysis test (CSA) [43] since it is considered one of the most credible psychometric tests to elicit cognitive style of users [43, 46]. In particular, we used the CSA test for assessing the Verbal-Imager dimension which indicates an individual's tendency to process information verbally or in mental pictures. An individuals' style on the Verbal-Imager dimension is obtained by presenting a series of 48 questions about conceptual category and appearance (i.e., colour) to be judged by the participants true or false. 24 statements require participants to compare two objects conceptually (e.g., "Are ski and cricket the same type?", and 24 statements require participants to compare the colour of two objects (e.g., "Are cream and paper the same colour?").

The psychometric test records the response time of each given answer to the questions and then uses a three-phase algorithm to determine the participant's cognitive style: i) calculate the average response time on each section (24 questions) of the CSA test, ii) calculate the ratio between the average response times on the verbal (conceptual category) and imagery (appearance) items, and iii) associate the value of each subject's Verbal-Imager ratio with a style category. A low ratio (<1.02) classifies the participant as a "Verbal", a high ratio (>=1.02) classifies the participant as an "Imager" [43].

3.3 Authentication Mechanisms Used in the Study

One text-based password mechanism and one recognition-based, graphical authentication mechanism were developed. Figure 1 and Figure 2 respectively present the text-based password mechanism and the graphical authentication mechanism used in the study.

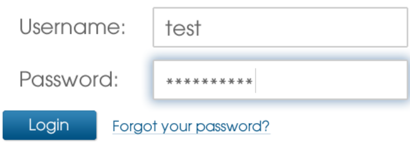


Fig. 1. Text-based password mechanism used in the study

The text-based password mechanism involved alphanumeric and special keyboard characters which could be chosen by the user. A minimum of eight characters including numbers, a mixture of lower- and upper-case letters, and special characters were required to be entered by the users. An additional option for resetting the text-based password was available in case the users forgot their authentication key. In that case, users had to enter their username and a hyperlink was sent to their email that led to a Web-page for resetting their text-based password.

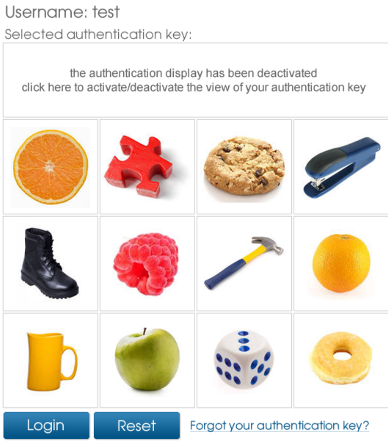


Fig. 2. Graphical authentication mechanism used in the study

A graphical authentication mechanism that involved single-object images was developed based on the recognition-based, graphical authentication mechanism proposed by Mihajlov and Jerman-Blazic [24]. The choice of this particular graphical authentication mechanism was based on the fact that its theoretical assumption (i.e., single-object images are better memorized and recognized by the human mind) is closely related to the dual coding theory [32, 33], which is considered the basis upon which the Verbal/Imager cognitive styles were developed [39]. During the authentication key creation, users could freely select between eight to twelve images, in a specific sequence out of a random subset of thirty images that were retrieved from a large image database. In case the user was not satisfied with the presented choices, an option to load a different random image subset was available. Repetitions of images were also possible in the sequence (in this case the key length could be longer than twelve images). After the graphical authentication key was created, a fixed image set of sixteen images, containing the user-selected authentication images and system-selected decoy images were permanently attached to the username in order to increase security, since if the decoy images were to change every authentication session, the authentication key could be easily revealed by eliminating the non-repeated images through subsequent sessions [24]. During authentication, a 4 x 4 grid containing the user-selected and system-selected decoy images were presented (Figure 2). The image positions in the selection grid were randomly positioned in each authentication session. Thereafter, users had to select their images in the specific sequence, as entered in the enrolment process in order to get permission for accessing the system. An additional option for resetting the authentication key was also available which was similar to the text-based password reset process.

3.4 Hypotheses

The following hypotheses were formulated for the purpose of our research:

- H_1. There is general preference of users toward text-based password mechanisms or recognition-based, graphical authentication mechanisms.
- H_2. Cognitive styles of users have a main effect on users' preference toward text-based password mechanisms or recognition-based, graphical authentication mechanisms.
- H_3. There is a significant difference with regard to time (efficiency) and total number of attempts (effectiveness) needed to authenticate through a text-based password mechanism or a recognition-based, graphical authentication mechanism among users belonging to the Verbal and Imager class.

3.5 Participants

A total of 145 people participated in the study between February and June 2012. Participants varied from the age of 17 to the age of 26, with a mean age of 22 and were undergraduate students of Computer Science, Electrical Engineering, Psychology and Social Science departments of the University of Cyprus. The participants' native language was Greek, and had learned English as a second language. A total of 2605 authentications have been recorded during the five-month period, with an average of 19.79 (SD 11.62) logins per participant.

3.6 Data Captured

Both client-side and server-side scripts were developed to monitor the users' behaviour during interaction with the authentication mechanisms. The following data was captured:

- **Performance (Efficiency and Effectiveness):** The total time (efficiency) and total number of attempts (effectiveness) required for successful authentication was monitored on the client-side utilizing a browser-based logging facility that started recording time as soon users entered their username for identification, until they successfully completed the authentication process. Additional performance data included total number of authentication key resets.
- **Preference:** The users' choice between the two variations of authentication during the last one and a half month of the study (Phase C) was used to draw conclusions about their preference toward a particular type of authentication.

Complementary self-reporting data for users' performance and preference toward a particular type of authentication was extracted by conducting semi-structured focus group sessions at the end of the study.

3.7 Analysis of Results

For our analysis, we separated users into two categories based on their cognitive style: Verbal (N=68, f=46.89%), and Imagers (N=77, f=53.1%), which consisted of participants that belong to the Verbal and Imager class, respectively.

Efficiency Related to User Authentication. A two by two way factorial analysis of variance (ANOVA) was conducted aiming to examine main effects between the users' cognitive style (i.e., Verbal, Imager) and authentication type (i.e., text-based password vs. graphical) on the time needed to accomplish the authentication task. Figure 3 illustrates the means of performance per cognitive style group and authentication type. In addition, Table 1 summarizes the descriptive statistics of each user group (Verbal and Imager) per authentication type (text-based password and graphical).

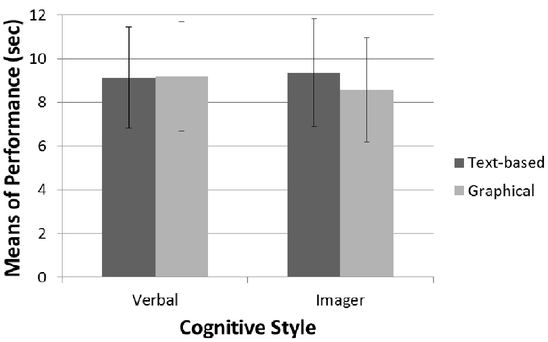


Fig. 3. Means of performance per cognitive style group and authentication type

The analysis revealed that there is a significant effect on the time needed to accomplish an authentication task with regards to users' cognitive style and authentication type used (F(1,896)=6.635, p=0.01). In particular, users belonging to the Imager class performed faster in graphical authentication than in text-based password authentication mechanisms. In contrast, users belonging to the Verbal class did not perform significantly different in either of the two authentication types (F(1,491)=0.074, p=0.785).

Table 1. Efficiency per cognitive style group and authentication type

	Verbal		Imager	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Text-based	9.13	2.31	9.36	2.47
Graphical	9.19	2.50	8.57	2.39

A further comparison between the cognitive style groups revealed that Imagers performed significantly faster in graphical authentication than Verbals ($F(1,514)=8.292$, $p=0.004$). However, performance in text-based password mechanisms revealed no significant differences in time since both cognitive style groups performed similarly ($F(1,381)=0.838$, $p=0.361$), with Verbals being slightly faster. Given that both types of user groups were more familiar and experienced with text-based passwords, this might explain the fact that no significant differences have been observed in the time to authenticate through the text-based password mechanism. On the other hand, since the familiarity factor has not affected the graphical authentication mechanism, results indicate that the different approach in processing information mentally of each user group (visual approach for Imagers and verbal approach for Verbals) has affected their performance in the graphical authentication mechanism.

Effectiveness Related to User Authentication. For each user authentication session, the total number of attempts made for successfully authenticating in each type was recorded. Table 2 summarizes the means of attempts of each user group (Verbal and Imager) per authentication type (text-based password and graphical).

Table 2. Effectiveness per cognitive style group and authentication type

	Verbal		Imager	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Text-based	1.15	0.40	1.30	0.63
Graphical	1.23	0.50	1.09	0.34

Shapiro-Wilk tests revealed that these distributions do not follow the normal distribution. Regarding the text-based password, although Verbals needed on average less attempts to authenticate, no significant differences were observed between the two user groups, as the Mann-Whitney U test revealed ($p=0.183$). In the case of graphical authentication, on average, users belonging to the Verbal class needed more attempts to authenticate than the Imager group. The Mann-Whitney U test revealed that the differences between Verbal and Imager user groups were statistically significant ($p<0.001$). In this respect, Imager users' enhanced ability of processing visual information has positively affected their effectiveness compared to Verbal users. This is further strengthened by the fact that the majority of authentication key resets were initiated by Verbal users for graphical authentication keys. In particular, a total of 7 authentication key resets were initiated by Imagers for graphical authentication mechanisms while in the case of Verbals the number of key resets rose to 24. In case of text-based passwords, a total of 11 password resets were performed; 4 by Imagers and 7 by Verbals.

Preference Related to User Authentication. During Phase C, the system provided the option to the users to choose which authentication type to use for accessing the system; either use the current authentication type that was active after Phase B, or use the first authentication type of Phase A. In Table 3 we summarize the users' preferences toward a particular authentication type based on cognitive style. Figure 4 illustrates the total number of authentication sessions in Phase C per cognitive style group.

Table 3. Preference per cognitive style group and authentication type

	Verbal	Imager	Total
Text-based	29	23	52
Graphical	39	54	93
Total	68	77	145

A binomial statistical test was conducted to examine whether there is a general preference relating text-based passwords and graphical authentication mechanisms (H_0: p(text-based password)=0.5 and p(graphical)=0.5). The results revealed that overall, users preferred graphical authentication (p=0.001).

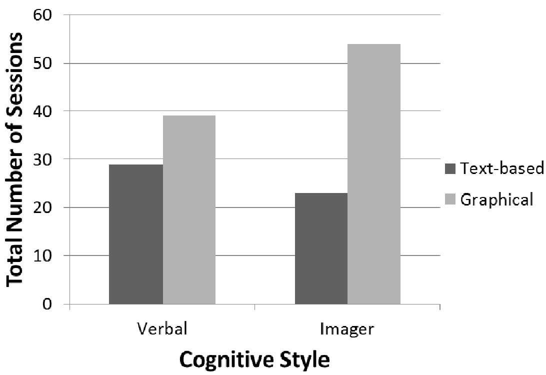


Fig. 4. Users' preference per cognitive style group

Next, a Pearson's chi-square test was conducted to examine whether there is a relationship between users' cognitive style and their preference toward a specific type of authentication (i.e., text-based password or graphical). The results revealed that in general there is no significant relationship between these two variables (Chi square value=2.563, df=1, p=0.109). Furthermore, examining each cognitive style group individually with respect to preference toward a particular authentication mechanism, it has been identified that users belonging to the Imager class significantly prefer graphical authentication mechanisms (p=0.001). In contrast however, users belonging to the Verbal class do not significantly prefer a particular type of authentication (p=0.275).

3.8 Focus Groups

Semi-structured focus-group sessions were concentrated around the participants' subjective preference and perception of the authentication mechanisms. The question structure of the sessions was focused around the following points: i) which type of authentication the users prefer, ii) which type of authentication was more efficient, iii) which type of authentication was more effective.

The first question for each session was intended to explore the users' preference on authentication mechanisms. The participants were asked to rate their preference toward a particular type of authentication mechanism (text-based password or graphical). A binomial statistical test was conducted to examine whether there is a general preference relating text-based passwords and graphical authentication mechanisms (H_0: p(text-based password)=0.5 and p(graphical)=0.5). The results revealed that overall users preferred graphical authentication (p=0.008) with a total of 89 users preferring graphical authentication mechanisms compared to 56 users who preferred the text-based password mechanism (Table 4).

Table 4. Authentication preference per cognitive style group

	Verbal	Imager	Total
Text-based	33	23	56
Graphical	35	54	89
Total	68	77	145

A binomial test was conducted separately for each cognitive style group to examine whether a particular cognitive style group prefers specific types of authentication mechanisms. Results revealed that Imagers significantly prefer graphical authentication mechanisms (p=0.001). This supports the aforementioned analysis of results that indicated a significant effect of cognitive style toward preference of authentication mechanisms. In contrast, regarding the Verbal group no clear preference toward text-based passwords or graphical authentication mechanisms was recorded (p=0.904).

As for efficiency and effectiveness of the authentication mechanisms, the participants were asked to rate the efficiency and effectiveness of each mechanism. Table 5 and Table 6 respectively summarize the users' rating regarding efficiency and effectiveness for each authentication type.

Table 5. Authentication efficiency per cognitive style group

	Verbal	Imager	Total
Text-based	38	33	71
Graphical	30	44	74
Total	68	77	145

A Pearson's chi-square test did not reveal a significant difference in efficiency between the text-based password and graphical authentication mechanism (Chi square value=2.452, df=1, p=0.117). Nevertheless, a noticeable difference in opinions about the efficiency of the two authentication types was observed by participants who belonged to the Verbal class since the majority found the text-based password mechanism more efficient.

Table 6. Authentication effectiveness per cognitive style group

	Verbal	Imager	Total
Text-based	28	22	50
Graphical	40	55	95
Total	50	95	145

Finally, when asked to rate the difficulty of each authentication type, the majority claimed that graphical authentication mechanisms were easier to recall and recognize the information.

3.9 Validity and Limitations of the Study

With the aim to increase internal validity we recruited a sample of participants that were rather experienced and average than novice users with respect to user authentication and therefore, the research design was setup in order to avoid inference errors. There has also been an effort to increase ecological validity of the research since the user authentication tasks were integrated in a real Web-based system and the participants were involved at their own physical environments without the intervention of any experimental equipment or person. In addition, participants were required to authenticate in the system throughout the semester during real-life tasks (i.e., access their university course's material). Finally, given that future studies will contribute to the external validity of the reported research, we argue that providing personalized user authentication mechanisms, adapted to users' cognitive characteristics, as well as other individual characteristics (e.g., cognitive processing abilities, working memory capacity) could improve the overall user experience with regard to user authentication tasks.

The limitations of the reported study are related to the fact that participants were only university students with an age between 17 to 26 years, and shared common cultural backgrounds. Given that Asian-type alphabets are primarily processed by the visual cognitive sub-system, in contrast to the Western-type alphabets that are primarily processed by the verbal cognitive sub-system [47], cognitive styles might have affected differently the user performance and preference of the authentication mechanisms in case participants had knowledge of non-Western-type languages. In addition, carrying out a single assessment of users' cognitive style might not fully justify the users' classification into specific cognitive-based groups since individuals might be influenced by other circumstances over time such as emotions, urgency, etc.

Finally, the results might be affected by the fact that the two types of authentication mechanisms diverge in term of user interface since the text-based password keys are entered through keyboard clicks, while the graphical keys are entered through mouse clicks on the images.

To this end, future studies need to be conducted with a greater sample of varying profiles, cultures and ages, as well as investigate the effect of contextual factors (i.e., user device) in combination with cognitive styles of users in order to reach to more concrete conclusions about the effect of individuals' cognitive style on their performance and preference related to authentication mechanisms. In addition, considering that in real life, users have to memorize multiple authentication keys [16], another important aspect for investigation is to study the effect of cognitive styles on user performance and preference when using multiple text-based password and graphical authentication mechanisms over time.

4 Conclusions

The purpose of this paper is to present results of an ecological valid user study which was designed with the aim to investigate whether there is a main effect of specific individual characteristics of users targeting on cognitive styles, towards performance and preference of two complementary types of user authentication mechanisms (text-based and graphical). Such an endeavor is considered valuable for the design and the deployment of more usable computer-human interaction processes with the aim to offer personalized and adaptive authentication mechanisms aiming to assist users to accomplish efficiently and effectively comprehensive and usable authentication tasks. For the purpose of this research we have designed a three phase experimental study which entailed a credible psychometric-based test for eliciting users' cognitive style based on the dual coding theory that suggests two sub-cognitive systems for processing and representing verbal and visual information in parallel threads.

The results of this paper can be interpreted under the light of this theory as they demonstrate a main effect of cognitive styles on both performance and preference related to authentication mechanisms. Regarding performance, results revealed that users of the Imager class performed faster in graphical authentication than in text-based password authentication, whereas users of the Verbal class did not authenticate significantly faster in either of the two authentication types. Furthermore, Imagers performed significantly faster in graphical authentication than Verbals did, however in the case of text-based passwords performance of both cognitive style groups was not considerably different, with Verbals being slightly faster than Imagers. An interpretation of this result can be based on the fact that all users were more familiar and experienced interacting with text-based passwords, hence no significant difference was observed between the Verbal and the Imager. On the other hand, since the familiarity factor did not affect the graphical authentication mechanism, we have observed that the visual approach of processing and organizing information of the Imagers has positively affected their performance compared to the Verbals.

Regarding effectiveness (i.e., total number of attempts), we conclude that both authentication mechanisms were effective in use throughout the study. Nevertheless, both number of attempts and number of authentication key resets reveal that graphical authentication mechanisms have positively affected user belonging to the Imager class as they needed significantly less attempts and key resets in graphical authentication mechanisms compared to Verbal users.

Finally, participants in general preferred graphical authentication mechanisms. Results also demonstrate that users categorized in the Imager group significantly prefer graphical authentication mechanisms. A possible interpretation of this result might be based on the novelty effect of graphical authentication. However, results suggest that if this would be the main factor that influences users' preference then it would be observed across all user groups regardless their cognitive style, which in the current sample is not the case, since users categorized into the Verbal group did not significantly prefer a particular authentication type.

Based on the presented results which embrace objective quantitative data (captured data during experimentations) as well as subjective qualitative self-reporting data (focus group studies), we argue that following a user-centered design methodology, it is necessary that designers of authentication mechanisms should clearly bear in mind individual differences of users while interacting with the system. Currently, there is a strong underlying design assumption that text-based passwords are the most comprehensive way for user authentication [48]. The results of this study suggest enhancing current authentication mechanisms aiming to embrace both text-based and recognition-based, graphical authentication mechanisms. As the results suggest, such an approach would have many positive implications from a usability and user experience point of view since, recommending authentication mechanisms, personalized to the users' cognitive style (especially in the case of Imagers, as results indicate) would increase the users' memorability and information processing efficiency of the authentication key, and thus improve task completion efficiency and effectiveness, and user satisfaction. At the same time, graphical authentication mechanisms provide similar security protection levels as text-based passwords from a service provider point of view, taking into consideration that they are encrypted properly on the service provider database layer and submitted securely on the transmission layer [24, 13]. On the other hand, from a user's point of view, they entail similar threats as text-based authentication mechanisms with regard to guessing (e.g., brute-force attacks) or capturing attacks (e.g. shoulder surfing, malware, phishing attacks, etc.) [24].

A practical implication of this work could be either to allow a user to explicitly declare the preferred authentication mechanism or by implicitly recommend the "best-fit" authentication mechanism based on historical usage data of the user in regard with efficiency and effectiveness of authentication tasks, or based on user interactions with a psychometric-based, cognitive style elicitation test. A more sophisticated architecture solution could be based on a recommendation engine as part of an adaptive system specialized on user authentication tasks aiming to extract from interaction data, the individuals' characteristics by use of statistical means.

Studies like the reported one can be useful for improving usable security on the World Wide Web through adaptivity in user interface designs with regard to authentication mechanisms, aiming to organize and present information and functionalities related with security tasks in an adaptive format to diverse user groups, by using different levels of abstractions through appropriate interaction styles, terminology, information presentation and user modeling techniques.

Acknowledgements. We would like to thank all the participants of the study for their time and qualitative feedback in the frame of the focus group sessions. The work is co-funded by the PersonaWeb project under the Cyprus Research Promotion Foundation (ΤΠΕ/ΠΛΗΡΟ/0311(ΒΙΕ)/10), and the EU projects Co-LIVING (60-61700-98-009) and SocialRobot (285870).

References

1. Shay, R., Kelley, P., Komanduri, S., Mazurek, M., Ur, B., Vidas, T., Bauer, L., Christin, N., Cranor, L.: Correct Horse Battery Staple: Exploring the Usability of System-assigned Passphrases. In: ACM Symposium on Usable Privacy and Security, Article 7, 20 pages. ACM Press, New York (2012)
2. Komanduri, S., Shay, R., Kelley, P., Mazurek, M., Bauer, L., Christin, N., Cranor, L., Egelman, S.: Of Passwords and People: Measuring the Effect of Password-composition Policies. In: ACM International Conference on Human Factors in Computing Systems, pp. 2595–2604. ACM Press, New York (2011)
3. Inglesant, P., Sasse, A.: The True Cost of Unusable Password Policies: Password use in the Wild. In: ACM International Conference on Human Factors in Computing Systems, pp. 383–392. ACM Press, New York (2010)
4. Florencio, D., Herley, C.A.: Large-scale Study of Web Password Habits. In: ACM International Conference on World Wide Web, pp. 657–666. ACM Press, New York (2007)
5. Adams, A., Sasse, A.: Users are not the Enemy: Why Users Compromise Security Mechanisms and How to Take Remedial Measures. J. Communications of the ACM 42(12), 40–46 (1999)
6. Shay, R., Komanduri, S., Kelley, P., Leon, P., Mazurek, M., Bauer, L., Christin, N., Cranor, L.: Encountering Stronger Password Requirements: User Attitudes and Behaviors. In: ACM Symposium on Usable Privacy and Security, Article 2, 20 pages. ACM Press, New York (2010)
7. Bonneau, J., Herley, C., van Oorschot, P., Stajano, F.: The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In: IEEE Symposium on Security and Privacy, pp. 553–567. IEEE Computer Society, Washington, DC (2012)
8. LeBlanc, D., Forget, A., Biddle, R.: Guessing Click-based Graphical Passwords by Eye Tracking. In: IEEE International Conference on Privacy, Security and Trust, pp. 197–204. IEEE Computer Society, Washington, DC (2010)
9. Zhang, J., Luo, X., Akkaladevi, S., Ziegelmayer, J.: Improving Multiple-password Recall: An Empirical Study. J. Information Security 18(2), 165–176 (2009)
10. Angeli, A.D., Coventry, L., Johnson, G., Renaud, K.: Is a Picture Really Worth a Thousand Words? Exploring the Feasibility of Graphical Authentication Systems. J. Human-Computer Studies 63(1-2), 128–152 (2005)
11. Everitt, K.M., Bragin, T., Fogarty, J., Kohno, T.A.: Comprehensive Study of Frequency, Interference, and Training of Multiple Graphical Passwords. In: ACM International Conference on Human Factors in Computing Systems, pp. 889–898. ACM Press, New York (2009)

12. Tullis, T.S., Tedesco, D.P., McCaffrey, K.E.: Can Users Remember their Pictorial Passwords Six Years Later. In: ACM International Conference on Human Factors in Computing Systems, pp. 1789–1794. ACM Press, New York (2011)
13. Biddle, R., Chiasson, S., van Oorschot, P.: Graphical Passwords: Learning from the First Twelve Years. J. ACM Computing Surveys 44(4), 41 pages (2012)
14. Brostoff, S., Sasse, M.A.: Are Passfaces More Usable than Passwords: A Field Trial Investigation. In: BCS International Conference on People and Computers, pp. 405–410. British Computer Society, UK (2000)
15. Passfaces Corporation. The science behind Passfaces, `http://www.passfaces.com/enterprise/resources/whitepapers.htm`
16. Everitt, K., Bragin, T., Fogarty, J., Kohno, T.: A Comprehensive Study of Frequency, Interference, and Training of Multiple Graphical Passwords. In: ACM International Conference on Human Factors in Computing Systems, pp. 889–898. ACM Press, New York (2009)
17. Stobert, E., Forget, A., Chiasson, S., van Oorschot, P., Biddle, R.: Exploring Usability Effects of Increasing Security in Click-based Graphical Passwords. In: ACM International Conference on Computer Security Applications Conference, pp. 79–88. ACM Press, New York (2010)
18. Jermyn, I., Mayer, A., Monrose, F., Reiter, M., Rubin, A.: The Design and Analysis of Graphical Passwords. In: USENIX Security Symposium, p. 1. USENIX Association, Berkley (1999)
19. Gao, H., Guo, X., Chen, X., Wang, L., Liu, X.: YAGP: Yet Another Graphical Password Strategy. In: IEEE International Conference on Computer Security Applications, pp. 121–129. IEEE Computer Society, Washington, DC (2008)
20. Tao, H., Adams, C.: Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. J. Network Security 7(2), 273–292 (2008)
21. Wiedenbeck, S., Waters, J., Birget, J., Brodskiy, A., Memon, N.: Authentication Using Graphical Passwords: Effects of Tolerance and Image Choice. In: ACM Symposium on Usable Privacy and Security, pp. 1–12. ACM Press, New York (2005)
22. Chiasson, S., Forget, A., Biddle, R., van Oorschot, P.: Influencing Users Towards Better Passwords: Persuasive Cued Click-Points. In: International Conference, B.C.S. on People and Computers, pp. 121–130. British Computer Society, UK (2008)
23. Bulling, A., Alt, F., Schmidt, A.: Increasing the Security of Gaze-based Cued-recall Graphical Passwords using Saliency Masks. In: ACM International Conference on Human Factors in Computing Systems, pp. 3011–3020. ACM Press, New York (2012)
24. Mihajlov, M., Jerman-Blazic, B.: On Designing Usable and Secure Recognition-based Graphical Authentication Mechanisms. J. Interacting with Computers 23(6), 582–593 (2011)
25. Nicholson, J., Dunphy, P., Coventry, L., Briggs, P., Olivier, P.A.: Security Assessment of Tiles: a New Portfolio-based Graphical Authentication System. In: ACM International Conference on Human Factors in Computing Systems (Ext. Abstracts), pp. 1967–1972. ACM Press, New York (2012)
26. Forget, A., Chiasson, S., van Oorschot, P., Biddle, R.: Improving Text Passwords Through Persuasion. In: ACM International Symposium on Usable Privacy and Security, pp. 1–12. ACM Press, New York (2008)
27. Wright, N., Patrick, A., Biddle, R.: Do You See Your Password?: Applying Recognition to Textual Passwords. In: ACM International Symposium on Usable Privacy and Security, Article 8, 14 pages. ACM Press, New York (2012)

28. Kuo, C., Romanosky, S., Cranor, L.: Human Selection of Mnemonic Phrase-based Passwords. In: ACM International Symposium on Usable Privacy and Security, pp. 67–78. ACM Press, New York (2006)
29. Halderman, J.A., Waters, B., Felten, E.: Convenient Method for Securely Managing Passwords. In: ACM International Conference on World Wide Web, pp. 471–479. ACM Press, New York (2005)
30. Chiasson, S., van Oorschot, P.C., Biddle, R.A.: Usability Study and Critique of Two Password Managers. In: USENIX Security Symposium, pp. 1–16. USENIX Association, Berkley (2006)
31. Brusilovsky, P., Kobsa, A., Nejdl, W.: The Adaptive Web: Methods and Strategies of Web Personalization. Springer, Heidelberg (2007)
32. Paivio, A.: Mind and Its Evolution: A Dual Coding Theoretical Approach. Lawrence, Erlbaum, Mahwah, NJ (2006)
33. Paivio, A., Csapo, K.: Picture Superiority in Free Recall: Imagery or Dual Coding? J. Cognitive Psychology 5(2), 176–206 (1973)
34. Anderson, J.R.: Cognitive Psychology and its Implications, 7th edn. Worth Publishers, NY (2009)
35. Ally, B.A., Budson, A.E.: The Worth of Pictures: Using High Density Event Related Potentials to Understand the Memorial Power of Pictures and the Dynamics of Recognition Memory. J. NeuroImage 35, 378–395 (2007)
36. Brady, T.F., Konkle, T., Alvarez, G.A., Oliva, A.: Visual Long-term Memory has a Massive Storage Capacity for Object Details. J. National Academy of Sciences 105(38), 14325–14329 (2008)
37. Oates, J.M., Reder, L.M.: Memory for Pictures: Sometimes a Picture is not Worth a Single Word. In: Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork, pp. 447–462. J. Psychological Press (2010)
38. Robertson, E.K., Köhler, S.: Insights from Child Development on the Relationship between Episodic and Semantic Memory. J. Neuropsychologia 45(14), 3178–3189 (2007)
39. Riding, R., Cheema, I.: Cognitive Styles – An Overview and Integration. J. Educational Psychology 11(3-4), 193–215 (1991)
40. Peterson, E., Rayner, S., Armstrong, S.: Researching the Psychology of Cognitive Style and Learning Style: Is There Really a Future? J. Learning and Individual Differences 19(4), 518–523 (2009)
41. Blazhenkova, O., Kozhevnikov, M.: The New Object-Spatial-Verbal Cognitive Style Model: Theory and Measurement. J. Applied Cognitive Psychology 23(5), 638–663 (2009)
42. Peterson, E., Deary, I., Austin, E.: A New Reliable Measure of Verbal-Imagery Cognitive Style. J. Personality and Individual Differences 38, 1269–1281 (2005)
43. Riding, R.: Cognitive Styles Analysis. Learning and Training Technology. Birmingham, UK (2001)
44. Kozhevnikov, M.: Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style. J. Psychological Bulletin 133(3), 464–481 (2007)
45. McAvinue, L.P., Robertson, I.H.: Measuring Visual Imagery Ability: A Review. J. Imagination, Cognition and Personality 26, 191–211 (2007)
46. Kinley, K., Tjondronegoro, D., Partridge, H.: Web Searching Interaction Model based on User Cognitive Styles. In: ACM International Conference of SIGCHI Australia on Computer-Human Interaction, pp. 340–343. ACM Press, New York (2010)
47. Tavassoli, N.: Temporal and Associative Memory in Chinese and English. J. Consumer Research 26(2), 170–181 (1999)
48. Herley, C., van Oorschot, P.: A Research Agenda Acknowledging the Persistence of Passwords. J. Security and Privacy 10(1), 28–36 (2012)

Survival of the Shortest: A Retrospective Analysis of Influencing Factors on Password Composition

Emanuel von Zezschwitz, Alexander De Luca, and Heinrich Hussmann

University of Munich (LMU), Media Informatics Group, Munich, Germany
{emanuel.von.zezschwitz,alexander.de.luca,
heinrich.hussmann}@ifi.lmu.de

Abstract. In this paper, we investigate the evolutionary change of user-selected passwords. We conducted one-on-one interviews and analyzed the complexity and the diversity of users' passwords using different analysis tools. By comparing their first-ever created passwords to several of their currently used passwords (e.g. most secure, policy-based), we were able to trace password reuse, password changes and influencing factors on the evolutionary process. Our approach allowed for analyzing security aspects without actually knowing the clear-text passwords. The results reveal that currently used passwords are significantly longer than the participants' first passwords and that most participants are aware of how to compose strong passwords. However, most users are still using significantly weaker passwords for most services. These weak passwords, often with roots in the very first passwords the users have chosen, apparently survive very well, despite password policies and password meters.

Keywords: password, evolution, security, policy, survey, human factor.

1 Introduction

Secret alphanumeric strings, called passwords, have been used to restrict access to specific information or services since the early days of computing. However, while 20 years ago passwords were mainly used by professionals for specific use cases, the introduction of the World Wide Web in the middle of the 1990s led to an extensive spread of passwords in people's daily lives. In the 2000s, the popularity of new technologies like smartphones and tablets and the growing amount of web-based services reinforced the process and thus, users nowadays have to memorize a multitude of passwords compared to a decade ago.

By the end of the 1990s, researchers began to evaluate the influence of user behavior on alphanumeric passwords [1, 2]. Those early studies which were based on self-reported data found that alphanumeric passwords always comprise a trade-off between usability and security. User-chosen passwords are often optimized for memorability and therefore based on dictionary words, birthdays, and so on. This makes them easy to guess for unauthorized persons. Furthermore, most people reuse passwords for multiple accounts and hardly renew passwords they once generated [3].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 460–467, 2013.

More complex passwords are often written down or shared with other people and thus, do not necessarily lead to improved security [4].

To support users in the selection of secure passwords, many companies introduced guidelines, password policies, recommender systems and password meters. The effect of these mechanisms was evaluated in lab studies which found that users often choose the same numbers or symbols and insert these at the same positions to comply with such systems. Thus, password policies do not necessarily increase system security [5]. Furthermore, it was shown that increasing the length of a password has the biggest effect on security and that the recall of policy-based passwords takes significantly more time [6]. This challenges the benefit of extensive password policies.

In addition to self-reported data and lab studies, large databases of user-chosen real-world passwords became available in recent years (e.g. [7, 8]). The analysis of these password lists confirmed that user-chosen passwords are often very short and many passwords are based on names, dictionary words and other trivial strings [7].

The goal of our study was to answer the question if password selection and security awareness evolved since the users' first contact with alphanumeric authentication and which factors influenced this process. Our work contributes to the field by providing valuable insights into the individual-related password evolution. We interviewed 40 people with different demographic backgrounds to get insights into their personal history of password use. We compared early passwords with currently used ones and analyzed what influenced the evolution of these passwords. By using electronic password analysis tools, we were able to quantify the data and gather detailed information about the complexity and the distance of different passwords without actually requiring the password itself.

The results show that the quantity as well as the complexity of used passwords rose in recent years and that password policies did influence this development. However, people still reuse passwords for multiple services and adapt to new policies by simply inserting new characters to old passwords. We found that even if users know how to build secure passwords, they use simple ones more often, some of them being their first-ever created password.

2 Evaluation

The main goal of our study was to find out how passwords evolved over time in terms of length, complexity and quantity and which factors influence this process. Therefore, we conducted one-on-one interviews and collected quantitative data using password analysis tools.

2.1 Password Tools

We implemented two password tools to analyze the composition of a given word and the distance between two words. The tools were built using JavaScript and HTML. Our tools did not store or transmit any data, but displayed statistical results to our participants, who copied them into a questionnaire.

Distance Test. The graphical user interface consisted of two password fields and one text field. When two words were entered into the password fields, the Levenshtein distance of these two words was computed and displayed in the text field. The Levenshtein distance describes the minimum number of changes required to transform one character sequence into the other.

Composition Test. The graphical user interface of the composition test consisted of one password field, one 14 x 2 table and one text field. When a word was entered into the password field, the composition of this word was analyzed and displayed.

We analyzed the word length and counted lower-case letters, upper-case letters, numbers and symbols. In addition, we checked for middle numbers/symbols and for repeated, consecutive and sequential characters. The script was based on the script of www.passwordmeter.com[1].

2.2 Design and Procedure

The study was conducted in a public coffee shop. We used one-on-one interviews in combination with a questionnaire and the password tools. The decision to conduct the study outside of the lab was made to gather a wider demographic spectrum. Participants were recruited via flyers, which were distributed in the coffee shop. As an incentive, we paid each participant one drink from the menu and gave out a 5 Euro shopping voucher. The interview lasted for about 20 minutes on average.

Participants were seated at a table in front of a wall. The examiner sat at the same table and used a laptop to read out the questions and to enter the answers into the questionnaire. At the beginning, we explained the study goals, the technical background (e.g. the password tools) and that there was no way we could steal their passwords from them. In addition, we warned them not to disclose their passwords during the interview. After the introduction, the interview started.

After collecting demographical data, we investigated the general password experience. This involved questions like the year of the first password creation or the amount of actively used passwords. In the case that our participants were not sure about their very first password, they were allowed to use the first password that they could remember. Further questions analyzed the influence of password polices, password meters, etcetera. After the interview, the laptop was handed over to the participants and they were asked to analyze their passwords. For this task, we positioned a screen in front of the participant to prevent shoulder surfing while passwords were entered.

The participants analyzed the requested passwords (see Table 1) and copied the results from the text field of the password tool to the questionnaire. Instructions looked like "*please compare your most secure password to the password that you use most often*". Using this approach, we were able to analyze many aspects of the passwords without actually knowing the participants' passwords.

[1] The script is available for distribution under the GNU General Public License (GPL).

Table 1. Password categories analyzed during the interviews

Password category	Definition
First	The first password ever created
Most used	The password which is used most often
Most secure	The password which is rated most secure (by the participant)
Policy-based	A password which was created based on a given policy
Meter-based	A password which was influenced by a password meter

2.3 Participants

We interviewed 40 participants. The average age of 39 participants was 26 years (18-59). One participant did not reveal his age, but stated to be 40 to 50 years old. 16 participants were female, 24 were male. 38 participants had an academic background, 19 of them had a technical background (e.g. computer science). This might be influenced by the coffee shop being located in a university and business district.

3 Results and Discussion

The results are based on the qualitative answers of the 40 interviewees and the quantitative password analysis. The definition of the analyzed password categories is found in Table 1.

3.1 Experience

On average, the participants had their first contact with passwords in the year 2000 (SE: 1, min = 1994, max = 2008). Their average age at this time was 15 years (SE: 1, min = 5, max = 54). Asked for the reason to create a password, 27 participants stated they had signed up for an email account; seven protected a user account of an operating system and four participants created their first password to protect a mobile phone. Beside these services, gaming and online banking were mentioned. In the first year of password use, our participants had to deal with a mean of 1.5 passwords (SE = 0.1, min = 1, max = 3). Today, the average amount is 14.2 (SE = 3.8, min = 1, max = 150). However, only 5.1 (SE = 0.7, min = 1, max = 28) of these are used frequently.

3.2 Complexity

We conducted a one-way repeated-measures ANOVA to compare the complexity of the different passwords. Fig. 1 shows the results of the complexity analysis; the concrete values can be derived from Table 2.

The results reveal that there is a highly significant main effect of the password category on the length of the password, $F_{2.38,135.99} = 10.33, p < 0.01$. The within-subject contrasts show that the first passwords (M = 7.6) are significantly shorter than passwords of all other categories (all p < 0.05). Most secure passwords (M = 12.1) have the most characters, but policy-based (M = 10.2) and meter-based (M = 10.7) passwords are not significantly shorter (all p > 0.05). In contrast, most-often used passwords (M = 8.7) comprise significantly less characters (all p < 0.05). This result indicates that, according to the amount of characters, passwords became more secure over time. However, even if users know how to create secure passwords, and this creation is supported by policies and password meters, most authentications are still performed using shorter and thus less secure passwords. According to our participants, secure passwords are only used for specific services whose data is rated sensitive (e.g. bank account).

Table 2. Mean values (and SE) of the different characters tested in the password analysis

Password category	Length	Lower-case	Upper-case	Numbers	Symbols
First	7.60 (0.35)	5.68 (0.50)	0.23 (0.10)	1.68 (0.34)	0.03 (0.03)
Most used	8.65 (0.27)	5.83 (0.42)	0.45 (0.14)	2.25 (0.29)	0.13 (0.05)
Most secure	12.13 (0.79)	7.65 (0.70)	0.95 (0.29)	2.85 (0.33)	0.68 (0.24)
Policy-based	10.18 (0.52)	6.44 (0.57)	0.72 (0.16)	2.67 (0.28)	0.36 (0.11)
Meter-based	10.65 (0.60)	6.52 (0.66)	1.03 (0.31)	2.68 (0.36)	0.42 (0.13)

The analysis of the password composition revealed significant main effects on the use of upper-case letters ($F_{2.52,73.09} = 3.80, p < 0.05$), numbers ($F_{2.89,83.68} = 4.64, p < 0.05$) and symbols ($F_{2.01,58.19} = 5.26, p < 0.05$). Interestingly, there is no significant difference on the use of lower-case letters (p = 0.13). This indicates that passwords were always based on lower-case letters, but recently created passwords additionally comprise numbers, upper-case letters and symbols. The post-hoc tests reveal that most secure passwords (M = 1.0) and policy-based (M = 0.7) or meter-based (M = 1.0) passwords include significantly more upper-case letters, than the firstly created ones (M = 0.2), all = p < 0.05. In contrast, most used passwords (M = 0.5) are not based on significantly more upper-case letters than the first passwords (p > 0.05). The analysis of the amount of numbers shows that compared to the first passwords (M = 1.7), passwords of all other categories use significantly more numeric characters (all = p < 0.05). However, most used passwords (M = 2.3), most secure passwords (M = 2.9), policy-based passwords (M = 2.7) and meter-based passwords (M = 2.7) do not differ significantly, p > 0.05. Looking at symbols reveals that both, the first (M = 0.0) and the most used passwords (M = 0.1) are composed of significantly less symbols than the rest of the passwords (p < 0.05). Most secure passwords contain the most symbols (M = 0.7).

An analysis of the simplicity of passwords according to the use of letters only reveals that 50% of the first passwords and 22.5% of the most used passwords consist of letters only. In contrast to that, only 6.3% meter-based, 7.7% policy-based and 7.5% of the most secure passwords use letters only. In addition, 12.5% of the first passwords were based on numbers only. One participant (2.5%) still uses a numeric password as most secure and most often password. No numeric passwords are found in the policy-based or meter-based category.

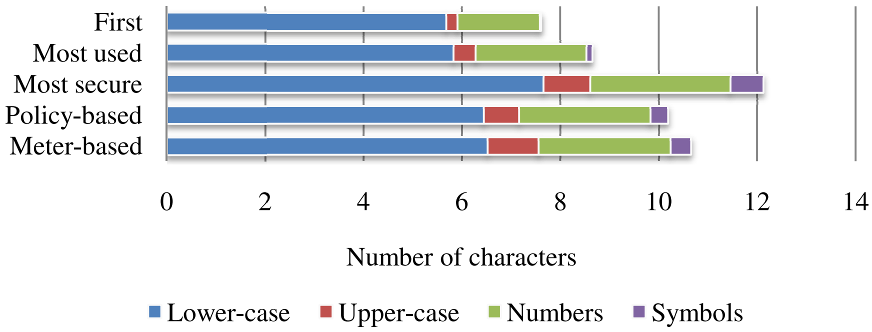


Fig. 1. Password composition by categories (see Table 1) comparing the amount of lower-case letters, upper-case letters, numbers and symbols

3.3 Usage and Behavior

45% of our participants still use their first password. While 30% use it only to authenticate with old services, 70% of them still use their first password to register with new services. This indicates that users rarely change passwords they once created and password reuse is common.

Fig. 2 gives an overview of the mean Levenshtein distances of four comparisons. The distance test reveals that four (10%) participants still use their first password as the most secure password (dist. = 0). In addition, 15% use their unchanged (dist. = 0) first password or a slightly changed (dist. = 1-3) version most often.

A one-way repeated-measures ANOVA comparing the means of the calculated distances reveals a significant main effect of password ($F_{3,87} = 4.61, p < 0.05$), Post-hoc tests show that this effect is caused by the significantly larger distance between the first password and the most secure password (all = p < 0.05). The analysis indicates that people often reuse passwords for multiple services: 40.5% of our participants use an unchanged (dist. = 0) or slightly changed (dist. = 1-3) policy-based password most often. This can be explained by the fact that 67.5% of the participants respond to password policies by adopting an already used password.

A similar conclusion can be drawn when comparing meter-based passwords to most often used passwords: 31.3% of our participants use an unchanged (dist. = 0) or only slightly changed (dist. = 1-3) meter-based password most often. Comparing the distance of the most secure and the most used passwords reveals that 17.5% of our participants use their most secure password most often (dist. = 0). In addition, 5% of the most secure passwords show only minor differences (dist. = 1-3) compared to the most used passwords.

The qualitative data supports the quantitative results as 51.4% stated to reuse passwords that they once created. According to the participants, this behavior is necessary since otherwise, they would have to memorize too many passwords. Even with reuse, 42.5% write down their passwords. A dependent t-test comparing the mean numbers of frequently used passwords of participants, who write down passwords (M = 5.5, SE = 0.7) with those, who do not write down their passwords (M = 3.8, SE = 2.3) reveals that the latter group also uses (and therefore has to memorize) significantly less passwords, $t_{37} = -2.06, p < 0.05, r = 0.32$. This indicates that the number of used passwords has a medium effect on the behavior of writing them down. One participant, who stated to use 28 out of 150 passwords frequently, was excluded from the analysis as this amount was outside of the doubled standard deviation (SD = 4.6) and therefore was assessed as an outlier. This participant mentioned that he uses a master password protected keychain to manage his passwords.

Finally, the results indicate that the influence of password policies on passwords was stronger than the influence of password meters as 92.5% stated that policies did influence their passwords, but only 57.5% stated the same for password meters. This might be based on the fact that password-meters are pure recommender systems, while password-policies often force users to follow specific guidelines.

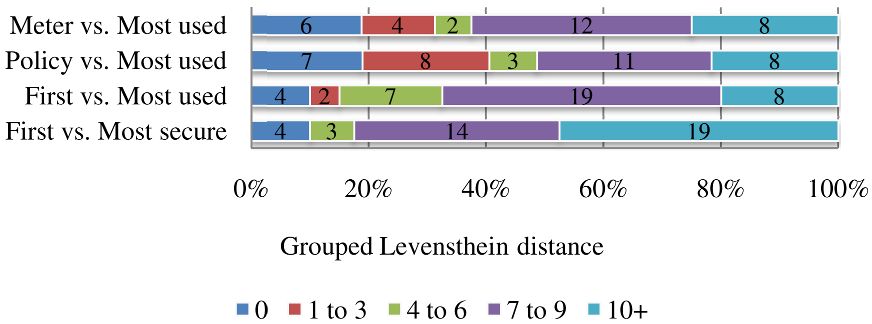


Fig. 2. Grouping of the calculated Levenshtein distances. The biggest differences can be found between the first passwords and the most secure passwords.

4 Limitations

Since the analysis is based on self-reported data, it is possible that the participants did not report their real passwords. Reasons might be that they did not want to expose them or could not perfectly remember some of their passwords. To check the validity of the reported data, we added several control questions into the questionnaire. In addition, we instructed our participants to report any memory gaps. Based on the results of this monitoring, we have no reason to assume that our participants reported contrived passwords and argue that our data is valid.

The complexity analysis is based on 40 participants and can therefore, in terms of generalization, not compete with analyses based on thousands of real user passwords. We like to mention that our goal was not to give an overview of currently used passwords, but to combine quantitative and qualitative data to be able to trace the individual-related evolution of passwords. Therefore, we are confident that the sample size meets the requirements for this approach.

5 Conclusion and Future Work

We conducted one-on-one interviews and used password analysis tools to gather valuable insights into the complexity of different user passwords. We could trace the evolution of our participants' passwords and the factors that influenced this process.

The results indicate that, in contrast to their first passwords, users today know how to build more secure passwords and thus some current passwords are based on significantly more characters. However, most people still rely on weak (e.g. short) passwords for most authentications, especially, when services are not rated sensitive. In addition, recommender systems and password guidelines had only marginal effects on the password strength and the reuse of old-established passwords is still common.

This is a serious security flaw as attackers could start by finding out a password of a low-sensitivity service, just by guessing or by using some deficiencies of the implementation. The reuse of passwords and the small distances to more secure passwords could consequently enable access to more passwords and more sensitive data. As the growth of web-based services will demand memorizing even more passwords in the future, we argue that usable alternatives to alphanumeric authentication have to be found.

Another point for further investigation is the potential influence of long-term mobile device use on password selection. Text input on such devices is cumbersome, which might, in the long run, negatively influence the security of such passwords.

Acknowledgments. This work was partially funded by a Google Research Award.

References

1. Adams, A., Sasse, M.A., Lunt, P.: Making passwords secure and usable. In: Proc. HCI 1997, pp. 1–19. Springer, London (1997)
2. Adams, A., Sasse, M.A.: Users are not the enemy. Commun. ACM 42, 40–46 (1999)
3. Riley, S.: Password Security: What Users Know and What They Actually Do. Usability News 8, 1 (2006)
4. Shay, R., Komanduri, S., Kelley, P.G., Leon, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F.: Encountering stronger password requirements: user attitudes and behaviors. In: Proc. SOUPS 2010, pp. 2:1–2:20. ACM, New York (2010)
5. Komanduri, S., Shay, R., Kelley, P.G., Mazurek, M.L., Bauer, L., Christin, N., Cranor, L.F., Egelman, S.: Of passwords and people: measuring the effect of password-composition policies. In: Proc. CHI 2011, pp. 2595–2604. ACM, New York (2011)
6. Proctor, R., Lien, M.-C., Vu, K.-P., Schultz, E., Salvendy, G.: Improving computer security for authentication of users: Influence of proactive password restrictions. Behavior Research Methods 34, 163–169 (2002)
7. Weir, M., Aggarwal, S., Collins, M., Stern, H.: Testing metrics for password creation policies by attacking large sets of revealed passwords. In: Proc. CCS 2010, pp. 162–175. ACM, New York (2010)
8. Bonneau, J.: The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: Proc. SP 2012, pp. 538–552. IEEE Computer Society, Washington, DC (2012)

Travel Routes or Geography Facts? An Evaluation of Voice Authentication User Interfaces

Alina Hang[1], Alexander De Luca[1], Katharina Frison[1], Emanuel von Zezschwitz[1], Massimo Tedesco[2], Marcel Kockmann[2], and Heinrich Hussmann[1]

[1] University of Munich (LMU), Media Informatics Group
[2] VoiceTrust, Germany
{alina.hang,alexander.de.luca,emanuel.von.zezschwitz, hussmann}@ifi.lmu.de, frison@cip.ifi.lmu.de, {massimo.tedesco,marcel.kockmann}@voicetrust.com

Abstract. Fallback authentication based on voice recognition provides several benefits to users. Since it is a biometric method, there are no passwords that have to be remembered. Additionally, the technique can be used remotely without the user having to be physically present. We performed stakeholder interviews and we iteratively designed and evaluated different voice authentication user interfaces with a focus on ease-of-use. The main goal was to keep embarrassment low and to provide an interaction as natural as possible. Our results show that small changes in the interface can significantly influence the users' opinions about the system.

Keywords: voice user interfaces, voice authentication.

1 Introduction

The number of passwords that users have to remember is steadily increasing. The odds of not being able to recall one's passwords are high and often result in bad habits. In particular, if users frequently have to change their passwords, they counteract these problems by writing them down or reusing the same password or variations of it for all accounts [1]. In the last resort, when even those questionable solutions fail (e.g. the user forgot where she wrote down the password), fallback authentication systems are required that enable users to regain access to their accounts. Well-designed fallback systems are important to encounter the mentioned habits.

Most companies rely on helpdesks, which require administrative personnel and that are expensive in costs (between $10 and $15 per transaction [4]). Automatic password reset systems reduce costs and disburden helpdesk staff. There exist a variety of solutions. Wood [10] categorizes them as something you have [7], something you know [9] or something you are. Systems in the latter category usually rely on biometrics. They require tokens of authentication that are tightly coupled to a person.

Voice authentication systems take advantage of the complex composition of the human speech which creates unique voice characteristics that, in turn, depend on the size of the oral cavity, the throat, nose, mouth and the arrangement of muscles in the

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 468–475, 2013.

according areas [5]. In combination with voice user interfaces, where human and computers communicate over speech, voice authentication systems are a good basis for automated password resets. Firstly, they are capable of recognizing if the speaking person is authorized to perform a password reset. Secondly, the system can guide the user through the process in a communicative manner, making the need of a helpdesk person obsolete. Finally, they do not require the person to be physically present as voice can be transmitted over distance communication channels like a telephone. Being completely automatic, usability aspects are highly important for such a method.

Naturalness and embarrassment are essential keys for the success of voice user interfaces. Embarrassment refers to the feeling of being uncomfortable caused by unnatural dialogs with a voice user interface. A conversation can be considered as natural if the same wording could result from talking with a human partner in daily conversations.

In the late 90s, Delogu et al. [2] showed that users are willing to put up with lower recognition rates in exchange for higher naturalness. The same study made a comparison between speech recognition and dual-tone multi-frequency (DTMF). Users were more satisfied with the speech recognition technique, highlighting the benefits of plain voice user interfaces. A reason for this can be found in [6] which states that the use of DTMF increases the cognitive load of the users, since they not only have to focus on the instructions on the phone, but also on the buttons they have to press.

In this paper, we present the iterative design and evaluation of enhanced voice authentication user interfaces for password resets. In a pre-study, we found out that current deployments of voice authentication and voice user interfaces work very well from a technical point-of-view. Still, they can be enhanced to increase user satisfaction and ease-of-use. We present two adaptations of a commercially available system and their evaluation. Our results show that both adaptations increase naturalness and reduce the level of embarrassment. We additionally show that feedback in the form of audicons can be helpful and disturbing at the same time, depending on how it is used.

2 Pre-studies

Most voice-recognition based authentication systems work in at least two phases [3]: During enrollment, the user's voice is recorded and a voiceprint is created to which the user's voice input is compared during authentication. We evaluated a text-dependent commercial system in which the user has to repeat selected word pairs during enrollment and a randomly selected subset of those during authentication. We conducted nine stakeholder interviews and a usability analysis with 18 participants.

2.1 Stakeholder Interviews

We performed interviews in two major companies who employ voice authentication systems for password resets. We decided to interview system administrators as they are hubs of user-reported issues. It has to be noted that this also means that they are only aware of issues that are actually reported and that they may be positively biased to a certain extent as these systems save them a lot of work.

We interviewed nine system administrators (5/4 of each company) with an average age of 38 years (30-43 years). The interviewees reported that some users do not like to perform the password reset in open plan offices, mostly due to embarrassment issues (e.g. the use of word pairs that do not occur in daily conversations). Some users also requested to improve the naturalness of the voice user interface and to shorten the overall password reset process. However, most users are satisfied with how well the voice authentication password reset works in general and how easy it is to learn.

2.2 Lab Study

We recruited 18 participants (7 female) with an average age of 23 years (19-30 years). We invited the participants to our lab to evaluate a currently available commercial system. They were placed in an office-like environment and we recorded all participants with a camera while they completed their tasks. In the first task, participants were asked to enroll for the voice authentication service. They had to call the service and follow the instructions for enrollment. After their voice was registered successfully, we asked them to perform the voice authentication that is needed for password reset (they did not have to actually reset their password).

The results of the usability analysis show that users were positively surprised about how well the voice authentication system worked. As shown in [8], some users are negatively biased against biometrics, which does not correspond to the actual capabilities of those systems. The usability study also confirmed the issues revealed during the stakeholder interviews. Lack of naturalness, embarrassment and temporal demand were the main concerns. The detailed results are outlined in section 5.

3 System Adaptation

Based on the preceding interviews and the lab study, we adapted the commercial system in two different ways. We addressed aspects like the number of voice inputs, feedback, naturalness and embarrassment.

Number of Voice Inputs: Speech recognition is more reliable if higher numbers of speech samples are available. If the number is too low, users might not be recognized during authentication, while a high number leads to annoyance and frustration due to the longer duration. Thus, a tradeoff between usability and security needs to be taken into account. We reduced the number of voice inputs from four to three, after the consultation of the voice biometrics provider in order to maintain security.

Feedback: We added two types of feedback (in the form of audicons). One feedback was given after successful voice input, the other occurred when the user reached the next level of the authentication process. The goal was to decrease users' insecurity when longer pauses occurred (this is due to the so-called endpointing, during which the end of a voice input is identified, when the pause reaches a certain threshold).

Naturalness: In order to reduce the perceived duration of the password reset, we redesigned the overall conversation style. Therefore, we recorded a new dialog with a new speaker. Particular attention was paid to the right intonation of phrases, i.e. by adding linguistic stylistics. For example, the dialog appears more natural when using pronouns. Instead of saying "*The enrollment will take five minutes*", the system says "*It will take about five minutes*". Based on the context, the user will still be able to understand the meaning. We also introduced discourse particles into the dialog. These are words with no real semantics, which help to structure the statement (e.g.: "*Well, that's fine*") and helped us to make the overall conversation more natural.

Embarrassment: Most text-dependent authentication systems ask for randomly selected sequences of words to be repeated. Without context, this feels unnatural and leads to embarrassing situations in public or semi-public spaces (imagine someone repeating the word "*cuckoo*" many times incoherently). In the enhanced systems, the users recite fragments embedded in phrases. Furthermore, we focused on more common word pairs used in daily situations instead of word pairs that are rarely used.

3.1 Adaptation 1: Geography Facts

The main goal of this adaptation was to reduce perceived duration of the password reset. Users are not only authenticated by their voice, but they also learn geography facts. This places the authentication process in a more playful setting, leaving the users with the feeling of not wasting their time. Additionally, we used fragments that may also occur in daily conversations, making the overall dialog more natural.

Altogether, we used three different phrases: "*The Danube flows into the Black Sea*"; "*New Delhi is the capital of India*"; and "*Munich is the provincial capital of Bavaria*". Many other facts are possible with this approach as well. This is important to avoid replay attacks (where attackers record the phrase and replay). Figure 1 (left) depicts an example of the dialog for "geography facts".

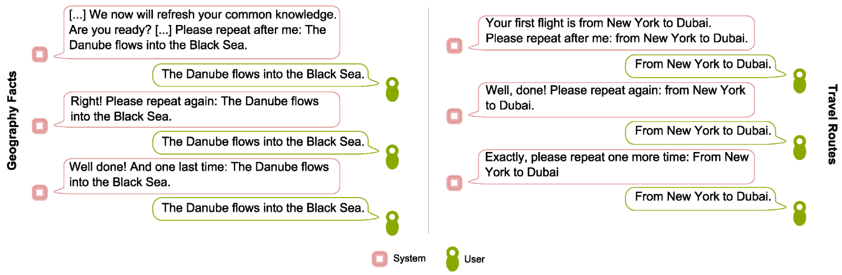


Fig. 1. Dialogs for the systems "geography facts" (left) and "travel routes" (right)

3.2 Adaptation 2: Travel Routes

The main goal of this adaptation was to reduce the level of embarrassment during the dialog with the voice user interface. We placed the communication into a travel context, where users imagine planning a travel route. Again, users had to recite conversational fragments like "*from Munich to Berlin*", "*from New York to Dubai*" or "*from Nuremberg to Hamburg*". This approach allows many different combinations. With only six cities, there are already 6x5 different combinations possible. Figure 1 (right) shows an excerpt from the dialog for "travel routes".

4 User Study

We performed an additional lab study with the adapted commercial system.

4.1 Study Design

We used a repeated measures factorial design. The dependent variables were *system* ("geography facts"; "travel routes") and *feedback* (with and without feedback) resulting in $2^2=4$ combinations. The order was counterbalanced to minimize order effects. For all combinations, we analyzed the subjective assessments of users. In particular, we were interested in how well the participants accepted the adapted systems. We also analyzed the influence of feedback on user satisfaction and the assessments of participants who had already participated in the pre-study.

4.2 Study Procedure

We conducted the study in an office-like setting to allow the participants to imagine the environment in which password resets are often done. The participants were informed about the general study procedure and were asked to provide demographic information. After that, the participants had to complete two tasks.

The first task consisted of the enrollment phase. In the second task, the participants tested "geographic facts" and "travel routes" with and without feedback. Additional questionnaires were handed out a) in case the participants tested a system that included feedback and b) when they tested a system for the second time. At the end, they were asked to answer a final questionnaire covering all systems (with and without feedback), including number of voice inputs, naturalness, speech, etcetera.

We recruited 24 participants (8 female). Their average age was 25 years (from 21 to 31). There were two groups of participants. 12 participants who already took part in the first lab study (group A) and 12 new participants (group B). This was desirable as it allowed for comparing the assessments of the different user groups.

5 Results

Duration: As shown in figure 2 (right), reducing the number of voice inputs had a positive influence on the perception of the overall duration. This is not surprising (since the actual duration was shortened). Nevertheless, it is interesting to see whether participants were satisfied with the number of voice inputs or whether they considered it as inappropriate. The differences between group A of the main study and the pre-study are statistically significant as shown by a Wilcoxon test (Z=-2,209, p=0.027). While 42% of group A rated the number of voice inputs inappropriate for the pre-study, only one person maintained this opinion. All other participants (75% of group

A) rated the duration as appropriate; almost half of which even found it as very appropriate. Regarding the main study, 67% of all participants considered the number of voice inputs as appropriate. The results are shown in figure 2 (left).

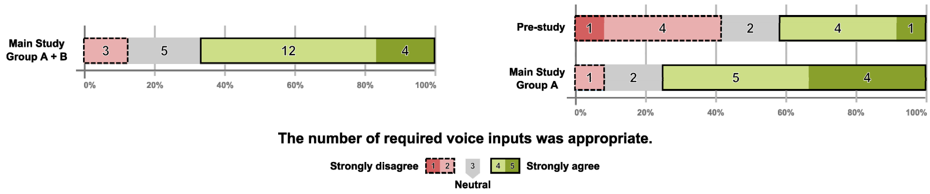


Fig. 2. User ratings of appropriateness of duration in the main study (left). Comparison of the results of group A to the results of the pre-study (right).

Feedback: There were two types of feedback (in the form of audicons), one after successful voice input, the other one when the user reached the next step. While more than 70% of participants considered feedback after successful voice input as important, more than 50% rated the latter form of feedback as obsolete (see figure 3).

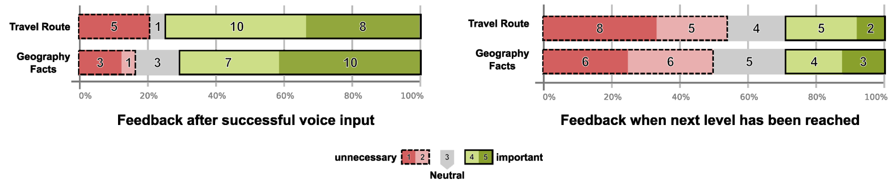


Fig. 3. Importance of the different types of feedback for "travel routes" and "geography facts"

Naturalness: Increased naturalness was one of the main aspects that our participants mentioned in the pre-study. Figure 4 (left) shows the assessment of all participants with respect to naturalness. 75% rated the naturalness as positive. On average, group A perceived the systems as more natural than group B. Regarding the language as well as the voice of the speaker, 92% of all participants found it pleasant, 64% of which even rated it as very pleasant. An overview can be seen in figure 4 (right).

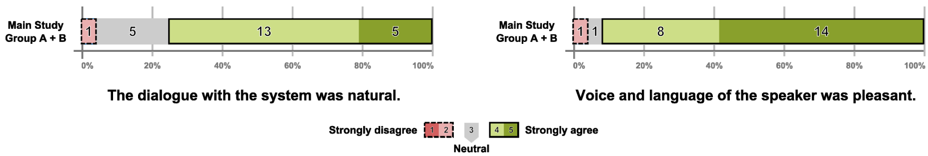


Fig. 4. Naturalness (left) and pleasantness of voice as well as language of the speaker (right)

Embarrassment: Amusement occurs with different notions: It can be positive, increasing pleasure, while a negative touch influences the level of embarrassment. This means that amusement, pleasure and embarrassment are tightly coupled.

One of the main goals of the adapted systems was to enhance naturalness and thus, decrease embarrassment. Figure 5 shows the embarrassment (left) and amusement (right) ratings of group A for each system in comparison to the pre-study. It can be seen that "travel routes" is rated least embarrassing and least amusing, while "geography facts" is also less embarrassing, but with a higher rating for amusement in comparison to "travel routes".

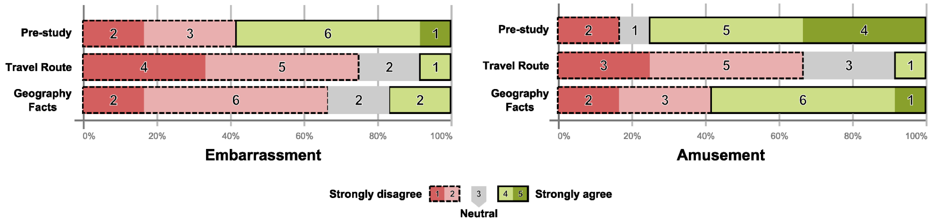


Fig. 5. Level of embarrassment (left) and level of amusement (right) for each system in comparison to the results of the pre-study

Geography Facts vs. Travel Routes: Finally, we asked the participants to compare the two adapted systems. For this, we developed a small webpage showing a simple square. Each corner of the square represented one system – "travel routes" and "geography facts" with feedback in the left corners; "travel routes" and "geography facts" without feedback on the right corners. We asked the participants to express their preferences by clicking inside the square. The results are shown in figure 6.

15 participants tended to systems with feedback and 9 participants preferred systems without feedback. Comparing the two systems, 10 participants preferred "geography facts" while 12 participants tended towards "travel routes". Interestingly, three participants selected almost the same area, independently from each other.

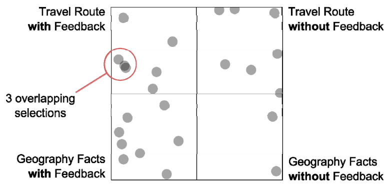


Fig. 6. User tendencies for "travel routes" / "geography facts" with and without feedback

6 Discussion and Conclusion

The results of the main study nicely show how slightly tweaking a few parameters (e.g. lower number of voice inputs) had a high influence on perceived embarrassment and naturalness when using voice user interfaces.

In particular, "travel routes" was the system that participants assessed as the least embarrassing, e.g. due to the use of neutral word pairs embedded in short phrases. In turn, it was also considered as less amusing than the old system. This is not necessarily a bad thing, since one has to consider the interplay between amusement and embarrassment. While the old system of the pre-study was rated as the most amusing, the embarrassment was almost as high. In this case, amusement has a negative notion and is not desirable. In turn, most participants found the system "geography facts" as amusing and only few rated it as embarrassing. The result highlights the entertaining factor of the system (and also the positive touch that the amusement holds). An interesting question related to this is the influence of entertainment on security awareness, meaning whether the users are still aware of the seriousness of the situation (e.g. the access to sensitive information) or whether this is clouded by the entertaining factor.

Having a closer look at the preferences, we can see that there are few participants with a strong preference (selections at the one of the four corners) for one of the four combinations. While only few participants were indecisive about their preference, most of them have a tendency towards one of the two systems. For example, one person may like the "travel routes" more, but is not averse to "geography facts" either. Therefore, one could think of enhancing one system with touches of the other (e.g. alternating between "geography facts" and "travel routes" for each level).

Regarding the use of feedback in the form of audicons, figure 6 shows that the opinions are diverse. While some liked having feedback, others preferred less feedback or even almost no feedback at all. Thus, feedback is mainly advisable at the end of a successful voice input, but should be avoided when the next step during the authentication process is reached. One may also think of allowing users to make personal feedback settings during enrollment.

There is still room for additional optimizations. For instance, a participant suggested adding a start signal for speaking to enhance feedback and to reduce insecurities during smaller pauses (which is due the so-called endpointing mentioned earlier). This is an interesting aspect and should be considered in future work. Additionally, a closer look has to be taken at the tradeoff between usability and security.

References

1. Adams, A., Sasse, M.A.: Users Are Not the Enemy. Communications of the ACM 42(12), 40–46 (1999)
2. Delogu, C., Di Carlo, A., Rotundi, P., Sartori, D.: Usability Evaluation of IVR Systems with DTMF and ASR. In: Proc. of ICSLP 1998 (1998)
3. Dugelay, J.L., Junqua, J.C., Kotropoulos, C., Kuhn, R., Perronnin, F., Pitas, I.: Recent Advances in Biometric Person Authentication. In: IEEE Transactions ICASSP 2002, vol. 4, pp. 4060–4063 (2002)
4. Forrester Consulting. Enhancing Authentication to Secure the Open Enterprise. Whitepaper commissioned by VerySign (Symantec) (2010)
5. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics 14(1) (2004)
6. Lerer, A., Ward, M., Amarasinghe, S.: Evaluation of IVR Data Collection UIs for Untrained Rural Users. In: Proc. of DEV 2010, pp. 2:1–2:8 (2010)
7. Mannan, M., Barrera, D., Brown, C., Lie, D., van Oorschot, P.: Mercury: Recovering Forgotten Passwords Using Personal Devices. In: Financial Cryptography and Data Security, pp. 315–330 (2012)
8. Pons, A.P., Polak, P.: Understanding User Perspectives on Biometric Technology. Communications of the ACM 51(9), 115–118 (2008)
9. Rabkin, A.: Personal Knowledge Questions for Fallback Authentication: Security Questions in the Era of Facebook. In: Proc. of SOUPS 2008, pp. 13–23 (2008)
10. Wood, H.M.: The Use of Passwords for Controlling Access to Remote Computer Systems and Services. In: Proc. of AFIP 1977, pp. 27–33 (1977)

Exploring the Use of Distributed Multiple Monitors within an Activity-Promoting Sit-and-Stand Office Workspace

Kathrin Probst[1], David Lindlbauer[1], Florian Perteneder[1], Michael Haller[1], Bernhard Schwartz[2], and Andreas Schrempf[2]

[1] Media Interaction Lab, University of Applied Sciences Upper Austria, Hagenberg/Austria
[2] Medical Technology Department, University of Applied Sciences Upper Austria, Linz/Austria
mi-lab@fh-hagenberg.at

Abstract. Nowadays sedentary behaviors such as prolonged sitting have become a predominant element of our lives. Particularly in the office environment, many people spend the majority of their working day seated in front of a computer. In this paper, we investigate the adoption of a physically active work process within an activity-promoting office workspace design that is composed of a sitting and a standing workstation. Making use of multiple distributed monitors, this environment introduces diversity into the office workflow through the facilitation of transitions between different work-related tasks, workstations, and work postures. We conducted a background study to get a better understanding of how people are performing their daily work within this novel workspace. Our findings identify different work patterns and basic approaches for physical activity integration, which indicate a number of challenges for software design. Based on the results of the study, we provide design implications and highlight new directions in the field of HCI design to support seamless alternation between different postures while working in such an environment.

Keywords: Multi-Monitor, Distributed Display Environment, User Interaction.

1 Introduction

Over the past centuries, our lives have become more and more dominated by sedentary behaviors such as prolonged sitting. Particularly in the work context (e.g., office jobs), the number of tasks focused around sitting at a computer has increased markedly in the recent years [21, 22], although a growing body of research has connected this sedentary lifestyle to a dramatic increase of health-related risk factors (e.g., for cardiovascular or musculoskeletal diseases) [7, 30]. In contrast to that, a physically active work process promotes the avoidance of prolonged sedentary behavior and related degenerative phenomena. Thus, the World Health Organization (WHO) has identified the workplace as important area for setting health-promoting preventive measures with great potential to solve the problem of rapidly increasing health care costs [28].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 476–493, 2013.

Despite the growing evidence for the benefits of such workplace health promotion programs, interventions have proven to be limited though, as most of them pose demands on workforce commitment (e.g., gym classes), require employees to leave their desks for exercising, or are unlikely to be sufficient in the context of otherwise sedentary lifestyles [5]. Rather, the accumulation of small amounts of low-intensity physical activity throughout a day provides health benefits that may even exceed those associated with more vigorous activities [19]. Based on this insight, recent approaches have focused on the motivation of reduced sitting, increased walking or standing activities (e.g., taking breaks, postural diversity, stand-up meetings) [11], and the design of novel workspace environments that provide opportunities for integrating light activities into the daily office workflow [8, 18, 20, 23].

Since computer technologies have become an integral part of today's office work, the integration of physical activity into the work routine however brings along a number of challenges that have been rarely addressed by the HCI community so far [23]. The facilitation of a physically active work process poses demands on user interface technology (e.g., ergonomics) and software (e.g., interface design, interaction metaphors, data management) that may be quite different from those addressed in the design of today's stationary work environments. Addressing these demands while simultaneously ensuring work efficiency therefore needs to be considered for the design of future systems within increasingly flexible and mobile work environments.

In this paper, we study a sit-and-stand workspace environment that is designed to promote physical activity in a computer workplace (see Fig.1), which is composed of a traditional desk workstation and a standing table workstation facilitating execution of work-related tasks in sitting or standing posture [23]. It is not clear though, how users will use this novel workplace and what issues they need to overcome in order to work in a both efficient and physically active way. To get a better understanding of how people are working in such an environment and how this may imply requirements for the design of future computer systems, we performed an exploratory background study. In the remainder of this work, we present the observations of this study and the resulting implications on user behavior and HCI design.

Fig. 1. A sit-and-stand workplace design integrating the concept of working in sitting and standing posture into the daily office workflow

2 Related Work

2.1 Workplace Health Promotion

Different structural elements have been proposed for the promotion of physical activity in the workplace, such as standing desks [8], stepping devices [20], or "walk-and-work" workstations [18], which are directly integrated into the office environment. Especially sit-and-stand workstations have shown to offer high potential for effective workplace health intervention through variations between different postures. Previous studies showed a generally positive attitude of workers towards such workspaces [29], and significant reduction of musculoskeletal discomfort [12] without any negative effects on work performance [14]. While workers equipped with a sit-and-stand workspace tend to be positive about having the possibility to choose between sitting and standing, previous research also indicates that people sometimes tend to avoid standing posture because of ergonomic (e.g., workplace size), hardware-related (e.g., cable reach), or software-related (e.g., data access) constraints.

Another recent ergonomic approach for physical activity promotion [23] proposes a new concept of working "in-motion" and takes into account the smooth integration of hardware- and software-related structural elements to provide opportunities for office workers to seamlessly change between different work environments. Based on this work, we use a similar environment and extend it by equipping the design with a multi-monitor setup for further investigation.

2.2 Multi-Monitor Usage

Multi-monitor systems have been studied in a series of previous work. Initial evidence indicates that users can complete tasks more quickly with multiple monitors [6], but typically experience inefficiencies or frustrations (e.g., losing the cursor, bezel problems, distal information access problems, window management problems, task management problems, configuration problems [24, 26, 27]) that can be addressed by alterations to basic interaction [16]. Further, a number of observations of multi-monitor users provide insights on the arrangement of digital information into focal and peripheral regions [9], window management [17] and other multi-monitor usage patterns. However, while much of the recent research involving multiple monitors has focused on compact setups where monitors are placed directly next to each other, a workplace design with two separate sitting and standing workstations introduces considerably larger distances than commonly used.

Since such environments have remained largely unconsidered by existing HCI research, we performed a background study to explore the usage in a real-world scenario, and provide insights into the user experience of working on a computer in sitting and standing posture. We analyze (1) the effects of such a distributed multiple monitor environment on user behavior, user interaction, usage patterns, and (2) how these may inform HCI design for supporting the concept of working while alternating between different work-related tasks, work postures, and corresponding workstations.

3 Concept and Method

In accordance to a recently proposed activity-promoting workplace design [23], the basic concept of the present study involved providing participants with two workstations (see Fig.1) for execution of work-related tasks in sitting or standing posture, which consisted of two height-adjustable desks[1] (Fig.2, left) and two active seats: a *swopper*[2] active chair (Fig.2, middle) that supports dynamic sitting through 360° tilting and vertical up-and-down flexibility, and a *muvman* chair[3] (Fig.2, right) that can be used as a leaning chair to support working in standing posture.

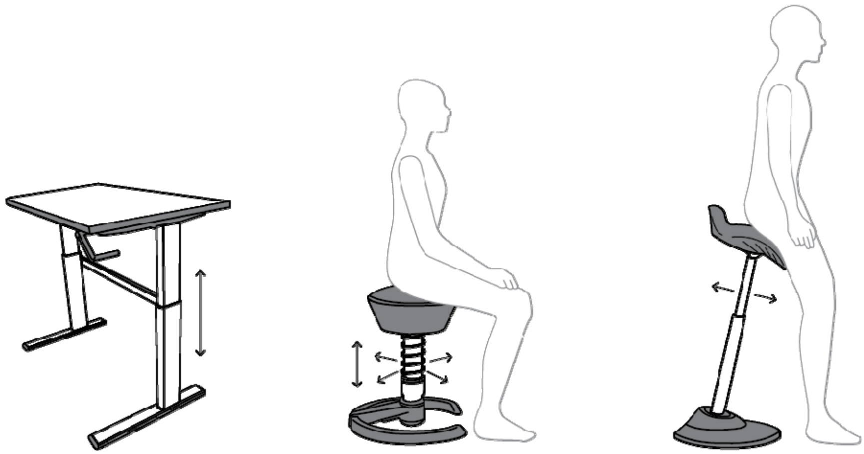


Fig. 2. As part of the study concept, participants were provided with two height-adjustable desks (left), a *swopper* active seat (middle), a *muvman* standing chair (right)

In the background study, we were interested in understanding several aspects of working within the proposed sit-and-stand workplace environment:

- How do participants organize the physical and hardware-related configuration of the workstations within the workspace, and how do people make use of the individual workstations within the environment (*workspace configuration*)?
- How do participants distribute their daily work time and work tasks among workstations within the workspace, and how is the usage influenced by the concept of working in different postures (*workstation usage*, *switching behavior*)?
- How well is this specific setup supported through current operating systems and software applications? Which problems and challenges arise (*software support*)?

[1] http://actiforce.eu/nc/products/steelforce
[2] http://www.swopper.de/en/swopper
[3] http://www.swopper.de/en/muvman

3.1 Participants

Five unpaid volunteers from the local university participated in a 4-week background study of their daily work activities within an activity-promoting workplace environment. All participants (4 male, 1 female; age 25-40) were experienced computer users, full-time employed as researchers in the field of computer science, ready to use such a setup to reduce sedentariness during the work routine. Four participants had no health problems, one participant reported on chronic back problems. All participants stated that they spent the majority of their work time (70-85%) in front of the PC, and had used multi-monitor setups for six months to three years. They replaced their existing work environments with a provided activity-promoting workstation setup.

3.2 Procedure and Setting

At the beginning of the study, all participants were provided with identical basic office furniture sets as described in the previous section. Participants were instructed to arrange these components within their office space to form a personal sit-and-stand workspace environment that consists of a sitting and standing workstation (see Fig.1).

Participants were introduced to the adjustable features of the furniture and *Ergonomics of Human System Interaction* (ISO 9241). Corresponding to their individual work-related requirements, each workstation desk should be equipped with display and input devices connected to a single workplace computer to provide equal access from each workstation. Further, participants were introduced to their operating systems' native multi-monitor functionality[4], and given free choice whether running their monitors in *duplicated* or *extended* mode throughout the course of the study.

3.3 Data Collection and Analysis

After setting up their personal workspace environments, participants' activities were observed for four subsequent weeks (20 work days). The individual workspaces were photographed, and specifications of the hardware equipment were collected. Participants reported on the time spent in sitting or standing posture, and overall work time at the end of each day. Throughout the study, six semi-structured interviews were conducted with every participant: after the first and third work day, and at the end of each work week. During the interviews, participants reported on their experience, and were assisted with suggestions for possible solutions to their reported problems. Interviews were audio-recorded (average duration 45 minutes) and affinity diagramming was used to analyze similarities, differences, and patterns in participants' behavior. In addition, a custom logging tool was installed on participants' computers, running in the background and collecting input and window events (similar to [17]). The data was used to gain insights on how often users switched between sitting and standing workstations, and which activities they performed in these locations.

---

[4] The systems' multi-monitor settings could be accessed and changed by pressing *Win+P*. `http://windows.microsoft.com/en-US/windows7/Connect-to-a-projector`

4 Results

Based on the data collected during interviews and logging (on average 330,000 input events per day for each participant), we drew information on participants experiences, behavior and (computing) activities during the conducted study.

In this section, we present our results according to aspects that we were primarily interested in: how did participants combine the provided furniture components with hardware-related structural elements (*workstation configuration*)*,* how did they integrate the concept of working in sitting and standing posture into their work routine (*workstation usage, switching behavior*), and to what extent this concept is supported by current software systems (*software support*).

4.1 Workspace Configuration

For all five participants, the arrangement of the provided furniture components resulted in a similar sit-and-stand workspace (see Fig.3). The tables were adjusted to participants' body height and arranged adjacent to each other, with a small gap of 10-15 cm –either at a slightly tilted angle of approximately 45° (four participants) or in straight line (one participant)–and were combined with the provided active seats to form the proposed sitting and standing workstations.

Fig. 3. Two participants' arrangements of their personal sit-and-stand workspace

Given the free choice how to combine the provided furniture with hardware components, participants further chose a similar configuration consisting of one computer (placed in a central position on the floor) with multiple connected monitors that were distributed across the tables. Four participants arranged a single monitor, keyboard, and mouse on each table (e.g., Fig.3, left). One participant had a dual-monitor setup on one table and a front-projected tabletop-display with stylus input on the other (Fig.3, right). All computers ran Microsoft Windows operating system (one with Windows 8, three with Windows 7, and one with Windows XP).

4.2 Workstation Usage

During the study period, participants reported an average daily work time of 7.4 hours (*SD* = 1.86), of which they spent the major part of 6.0 hours (*SD* = 1.97) within their personal office workspace. During that time, their main work tasks (Fig.4, left) included software development, reading/writing (e.g., reports, articles, presentations), research/browsing, data analysis (e.g., statistics), communication (e.g., e-mail, telephone), meetings, and other (e.g., design, graphics, measurement) activities.

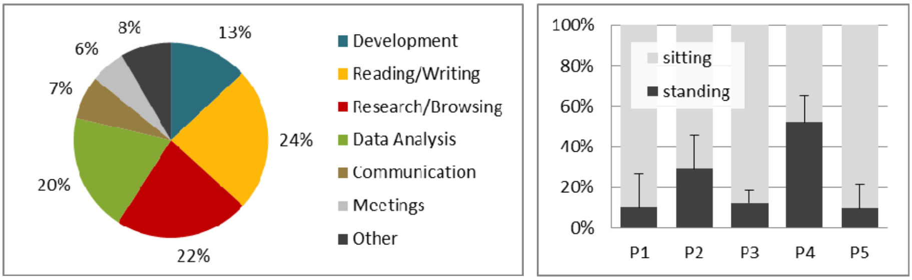
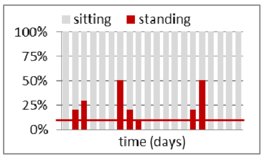


Fig. 4. Participants' reported daily work tasks (left) and sitting/standing times (right)

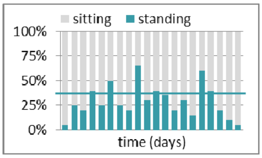
The average distribution of the reported sitting and standing times shows that all five participants made use of the opportunity to incorporate the concept of working in different postures into their regular office workflow. They reported spending on average 22.5% (*SD* = 21.1%) of their daily office time in standing posture (Fig.4, right). Given the relatively high variations, the participants' comments during the interviews provide further insights on a spectrum of different approaches for utilizing the novel sit-and-stand workspace. In-depth reports on the different aspects contributing to the above reported distribution are discussed in detail in the remainder of this chapter.

4.3 Participants' Overall Experience

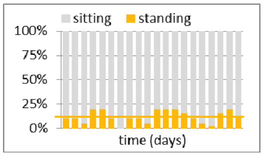
To provide a brief overview of participants' overall experience, we present their basic approaches, behaviors and usage patterns over the course of the study–along with the daily sitting and standing times (average indicated through a horizontal line) that are depicted for each participant individually.



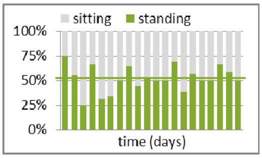
P1 used two monitors in *extended mode*, spent most of his work time using the primary screen at the sitting workstation, and switched to standing posture at rare intervals–virtually treating it as an independent unit that provided him with a scope for handling specific work tasks in isolation. Along with that, nevertheless, he experienced limitations during cross-monitor interaction (e.g., losing the mouse cursor during drag interactions) and during window switching (e.g., no taskbar on the secondary screen)–which did not encourage him to make use of the standing workstation on a more regular basis.



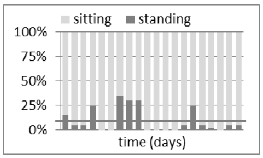
P2 used two monitors in *extended mode*, mainly working at the primary screen at the sitting workstation, and switching to the standing workstation multiple times throughout a day to break up sitting periods from time to time. By doing so, however, he experienced a range of problems with inconsistent window placement, during cross-monitor interaction (e.g., effort of dragging windows), and during window switching (e.g., no taskbar on the secondary screen). Therefore, after the second week of the study, he started to use the operating systems' native multi-monitor support–which enabled him to relocate the primary screen during workstation switches, and dissolved most previously existing problems.



P3 used two monitors in *duplicated mode*, spent most time at the sitting workstation, and tried to develop a routine to integrate working in standing posture into his workflow. After initially taking the opportunity to perform short e-mail or browsing tasks at the standing workstation, he finally took a rather opportunistic approach of occasionally alternating his work posture at logical points within a task. He appreciated the convenience of switching between the duplicated monitors without effort, but raised concerns on peripheral distraction and privacy. To avoid these problems, he sometimes used the native multi-monitor support, but considered this as additional effort to some extent.



P4 used two monitors in *duplicated mode*, spent approximately equal amounts of his work time in sitting and standing posture, and alternated his work posture on a regular basis throughout a day. Having chronic back problems, he felt that this was a possibility to relieve the pressure on the spine, and appreciated the flexibility to switch between the duplicated monitors whenever he wanted. Over the course of the study, some distinctive patterns started to evolve from his day-to-day work, as he tended to start the day with checking e-mails in standing posture, to work in sitting posture after lunch break, and to perform certain work tasks at one of the workstations (e.g., collaborative settings in standing posture, cognitive demanding tasks in sitting posture).



P5 used two monitors and a tabletop projection in *extended mode*, using the dual-monitor setup at the sitting workstation and the tabletop-setup at the standing workstation for debugging of pen-based input. During programming, he found that this worked well for him, as he perceived the high-frequent regular switches as a welcome change. Due to the distinct separation of coding and debugging views and the pen input, he did not have to take any additional effort for dragging windows or moving the mouse cursor. Over time, however, he did not have any opportunity to work in standing posture besides of this specific implementation task and ended up spending the majority of the time seated.

4.4 Switching Behavior

During a work day, the logged data shows that participants switched on average 6.27 times (*SD* = 8.47) between the sitting and standing workstations at intervals of 1.13 hours (*SD* = 1.12), with an average duration of 1.82 hours (*SD* = 1.88) in sitting and 0.44 hours (*SD* = 0.36) in standing posture. Comments from the interviews provide evidence for a variety of factors related to participants' physical and emotional *personal condition*, subject and progress of the current *work task*, or hardware- and software-related *workspace configuration* that were perceived to influence participants' switching behavior in a positive (encouraging) or negative (limiting) way. These reported influencing factors are discussed in detail in the remainder of this section and summarized in Table 1.

Table 1. Reported influencing factors with encouraging (+) or limiting (–) effects

Category	*Influencing Factor*	*P1*	*P2*	*P3*	*P4*	*P5*
Personal Condition	Physical Discomfort	+	+	+	+	+
	Personal Habits	–		–	+	
	Emotional Stress		–	–		
Work Task	New/Other Task	+	+		+	+
	Workstation Association	+	+			+
	Logical Breakpoint		+	+	+	
	Cognitive Demand		–	–	–	
	Task Involvement	–		–		
Workspace Configuration	Collaboration	+	+	+	+	+
	Accessibility		+	+	+	
	Physical Constraints			–	–	–
	Software Limitations	–	–	–	–	–

Personal Condition. One consistent finding applicable for all participants is that *physical discomfort* of prolonged sitting (e.g., back muscle tension) or standing (e.g., fatigue, aching feet) was the main trigger to switch from one workstation to another. Participants stated to appreciate the possibility to bring diversity into their daily office routine by alternating between sitting and standing posture. This was especially true for P4, who had chronic spinal problems and he felt that regular posture changes relieved his back pain fundamentally. Besides of that, four participants (P1-P4) noted their switching behavior being dependent on *personal habits* to a certain extent. P4 for example, started nearly every work day in standing posture as he was generally feeling energized in the morning, and P1/P3 stated to rarely work in standing posture as they were not used to this kind of bodily load. Two participants (P2, P3) reported limited number of postural switches (i.e., preference of sitting posture) in situations of high *emotional stress* when they were feeling anxious or impatient.

Work Task. Besides the triggers related to participants' personal condition, most of them (P1-P4) also stated that the start of a *new task* or switching the current task was commonly associated with postural changes. To find additional opportunities for switching their workstations on a frequent basis, two participants (P3, P5) further started to introduce direct *association of workstations* with particular tasks. P3 for example, was initially taking occasional breaks from his mainly seated day by performing short e-mail or browsing tasks at the standing workstation. P5 used the dual-monitor setup on the sitting workstation for his common work, and the projector-setup on the standing workstation for debugging activities. Along with that, however, they oftentimes ran into situations when their work tasks were not diverse enough to support a regular switching behavior. Given that, P5 tried to find additional tasks that he considered suitable to be performed at the standing workstation (e.g., reading on printed paper), while P3 changed his usage pattern after the first week of the study, thenceforth (similar to P2 and P4) rather changing his posture at logical *breakpoints* in-between the current task (e.g., at the end of a paragraph in a text). Limitations for switching behavior, on the other hand, were indicated by three participants (P2-P4) tending to avoid standing posture for tasks with relatively high *cognitive demand* (e.g., programming), and two participants (P1, P3) showing deep *task involvement* that made them simply "forget" about changing their posture when they were completely focused on the current task.

Workspace Configuration. Another interesting finding reveals that all participants pointed out the suitability of the standing workstation for *collaborative activities*. For typical settings in small groups of 2-4 people (e.g., discussions, showcasing), participants highlighted the affordances of the standing posture such as equal sight on the display surface and being at equal eye level. Closely related to that, the *accessibility* of the standing desk encouraged most participants (P2-P4) to use it as a first entry-point after leaving the workspace temporarily for external activities (e.g., meetings, breaks). On the other hand, some participants (P3-P5) noted situations where *physical constraints* (e.g., hardware components, notes on printed paper) turned out as limitations to their switching behavior, and all participants consistently reported that *software limitations* (e.g. relocation of on-screen content) had a considerable effect on their switching behavior–which is discussed in detail within the following section.

4.5 Software Support

The analysis of the collected logging data shows that participants actively used their workplace PCs (incoming mouse, keyboard, or pen input events) on average 4.98 hours per day. Throughout a work day, participants used 5-12 different software programs (Fig.5, left), which included programming, text editing/viewing (e.g., text processor, PDF viewer), web browser, spreadsheet, communication (e.g., e-mail client, instant messenger), media (e.g., graphics/video/audio editor), and other (e.g., file browser) applications.

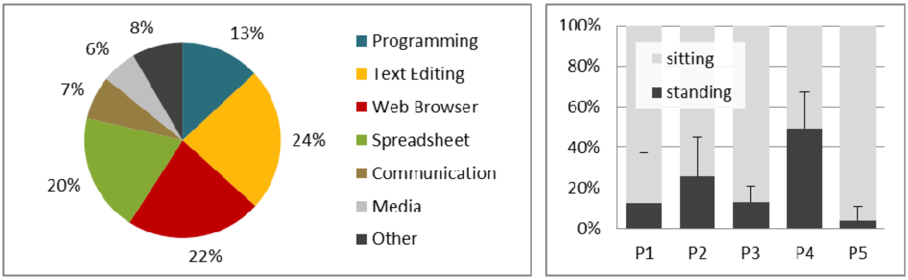


Fig. 5. Participants' logged software usage (left) and sitting/standing times (right)

The reported sitting/standing times (see Fig.4, right) are largely confirmed by the logging data (Fig.5, right), showing on average 20.61% (*SD* = 21.42%) of the input time being recorded at the standing workstation. In addition to participants' different motivations for workstation switching, the distribution of monitors within the workspace had a major effect on their behavior. With a relatively large gap (80-150cm) between monitors spread across the sitting and standing tables, all participants reported to perceive the workstations more or less as two separate units–feeling that they could use either the one or the other, but not both at the same time. As a result, two participants (P3, P4) decided to run their monitors in *duplicated* display mode, while the other three (P1, P2, P5) ran their setup in *extended* mode, and two participants (P2, P3) changed to a hybrid *relocated* display mode in the course of the study by manually switching the display configuration. Participants' experienced software support issues are discussed in the remainder of this section and summarized in Table 2.

Table 2. Reported software support issues and their occurrences per participant

Issue	*P1*	*P2*	*P3*	*P4*	*P5*
Primary/Secondary Screen Distinction	•	•			•
Window Placement Inconsistency	•	•			•
Cross-Monitor Interaction	•	•			
Content Relocation Effort	•	•	•		•
Peripheral Distraction			•	•	
Privacy			•		
Waste of Resources	•	•	•	•	•

Extended Screen. In contrast to regular multi-monitor setups, the display units in our participants' setups were non-adjacent [26]. Consequently, none of them exhibited patterns like using the display space as a whole (e.g., straddling a window across monitors), using the two monitors in direct support of each other (e.g., reading on one screen while writing a text on the other), or for displaying peripheral information (e.g., at-a-glance e-mail notifications) [9]. Instead, participants used the two screens in a rather separate fashion, thereby experiencing a variety of problems that had a major influence on their usage of the single workstations.

While the inherent *primary/secondary screen distinction* makes sense in a traditional extended desktop setting, this multi-monitor usage paradigm did not hold for the present setup. Two of the three extended screen users (P1, P2) reported window management limitations due to the taskbar, start menu, and task switching overview being displayed on the primary screen only. Consequently, while working on the secondary screen, they experienced difficulties to manage multiple windows without the visual feedback and one-click access functionality of these interface elements. As a result, P2 started to prepare well-defined work packages that did not require extensive window switching activities before switching to the secondary monitor at the standing workstation. P5 in contrast, reported to benefit from recently added multi-monitor support in the latest Windows version (Windows 8) providing him with the possibility to put a separate taskbar on each individual screen. A problem that however applied to all three extended screen users was the missing visual feedback of the quick-access task switcher (Alt+Tab[5]) being located on the primary screen by default.

Another issue reported by all extended screen users was the problem of *window placement inconsistency*. Most applications' inherent memory mechanisms [15] to open application windows on their last known location were holding only as long as people kept running a program at the same workstation (e.g., P5). Whenever participants were transitioning between workstations though, windows that continued to appear on the same (but not necessarily current) screen had a rather detrimental effect. P1/P2 for example, reported that applications were frequently launched on the screen that they were currently not working at, that pop-up windows (e.g., notifications, file dialogs) easily went unnoticed on the second screen, and P5 experienced similar problems whenever he performed debugging tasks while sitting in some exceptional cases.

Associated with the above mentioned issues, participants oftentimes ran into situations when they had to apply *cross-monitor interactions* to shift digital content (e.g., windows, folders, icons) from one screen to another – thereby traversing the distance between the monitors by moving the mouse at one workstation to reach to a screen area on the other. Since the physical distribution introduced visual discontinuities [27], two participants (P1, P2) reported difficulties in navigating the mouse cursor, or reading text on the distant screen. Losing the mouse cursor [24] was reported to frequently occur in situations when the cursor was "left behind" during workstation switches, or when "disappearing" out of sight along the screens edges.

Finally, the need to associate changes in posture with interactions for digital *content relocation* was considered as a major limitation. Two participants (P1, P2) stated that they perceived it as annoying to bring windows from one side to the other during workstation switches. For P3 and P4, this additional effort was one main reason why they had decided to run the monitors in their workspace in duplicated display mode: "*It's not the same as having two monitors on one desk. If the other monitor is far away and you're trying to see when you're trying to pass a window from one screen to the other, this is not so easy to do. (…) That's why I couldn't imagine working with the monitors in extended mode.*" (P3).

---

[5] `http://support.microsoft.com/kb/79869`

Duplicated Screen. Participants who used the setup in duplicated display mode (P3, P4) did not experience the same problems as extended screen users. By virtually having a single screen replicated on the extended monitor, they did not have to deal with discrepancies between primary and secondary screen, losing the mouse cursor, or the effort of moving digital objects from one screen to another. Having the exactly same view on all monitors, they were flexible to switch between the sitting and standing workstations at any desired point during task execution. Apart from that, duplicated screen users experienced other problems, though.

Having screen contents duplicated between workstation, participants reported feeling slight *distraction* due to on-screen interactions (e.g., mouse cursor movement, window pop-up) being visible on the second monitor within their peripheral vision. Closely related to this issue, one participant (P3) raised minor *privacy* concerns due to on-screen activities being observable for other people within the office space. While this was not a problem for P4, whose monitors were not visible from within the office space, the three extended screen users stated that privacy was a considerable reason why they had decided not to run their setups in duplicated mode.

A major concern however, was the *waste of resources* in terms of screen real estate and energy consumption, which was similarly indicated by the three extended screen users. By using only one of the two available monitors at a time, participants could not make use of a large part of their available screen space: "*If I could work on both monitors at the same time, if I could control them in a better way… that would be much better. To be flexible to change between the two workstations, I should be able to make it feel more like one. (…) In terms of digital communication, it could be smoother, more continuous–instead of totally separated.*" (P1).

Relocated Screen. To avoid the disadvantages of both duplicated and extended display mode in the present multi-monitor setting, P2 and P3 started using the Windows native keyboard shortcut Win+P for second screen projection after the first two weeks of the study. In this special usage scenario, postural changes were associated with pressing the keyboard shortcut to *relocate* the primary screen between sitting and standing workstation monitors–which resolved issues regarding primary/secondary screen, window placement, and cross-monitor interaction, and dissolved concerns about peripheral distraction, privacy, and energy consumption. Especially for P2, the new functionality changed his workstation switching behavior fundamentally, encouraging him to change his work posture at arbitrary points during a task. However, the major problem of losing a large part of available screen real estate remained unsolved.

Other remaining problems in this configuration were for example the monitor's latency to display content after enabling it with the shortcut. Further, it was not always clear which workstation would be enabled after pressing the keyboard shortcut, and which monitor would be turned on when participants returned to the workstation after prolonged inactivity (e.g., absence from the workspace due to external meetings). Furthermore, P2 felt like the black screen on the unused workstation was a minor barrier compared to an activated monitor, and P3 rarely made use of the screen relocation function because he considered this as additional effort to some extent.

5 Discussion and Implications

Based on the described observations, we draw the following implications for the support of environments similar to the studied sit-and-stand workspace design.

5.1 Basic Approaches for Physical Activity Integration

From the findings of the study, we identify two basic approaches that were followed to incorporate the concept of working different postures into the daily office work:

- **Task-Driven** approaches are based on the concept of associating a work task with a specific workstation, thus implicitly motivating people to take up different postures throughout a work day. P5 for example, associated debugging activities with the standing workstation, P4 preferred the execution of collaborative tasks in standing posture, and P3 initially handled e-mail activities at the standing table.
- **Self-Determined** approaches are in contrast based on the concept of performing any work tasks at any workstation, leaving the participants free to change their posture at all times. P3 for example, proceeded to alternate between the sitting and standing posture every once in a while, or P4 always changed his work posture whenever he felt the desire to break continuous sitting or standing periods.

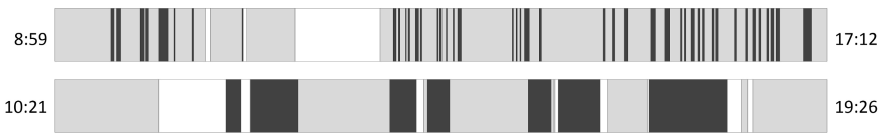


Fig. 6. Example distributions of computing events for task-driven (P5, top) and self-determined (P4, bottom) switching between sitting (gray) and standing (black) workstation.

Considering the experience of participants following a rather task-driven switching behavior, it becomes evident that this approach is strongly dependent on the daily work activities. While participants stated to perceive postural changes as rather natural side effect of their regular office work, it is on the other hand not suitable for tasks that are not performed on a regular basis. Self-determined switching behavior turned out as a flexible approach that leaves people free to carry out their work in any posture, but is on the other hand associated with a certain cognitive demand and might not work during phases of high involvement in the ongoing task. From the analysis of participants' switching patterns, we see that self-determined approaches resulted in significantly longer standing periods (Fig.6, bottom), while task-driven approaches resulted in a higher number of switches (Fig.6, top) during a work day.

Since both approaches have their pros and cons, and participants tended to not exhibit one of these two approaches exclusively, we think that it is essential to support a hybrid solution: while the task-driven approach has high potential to achieve an implicitly motivating effect, users should at the same time not be limited to exclusive association of task and workstation, but rather be able to switch in a self-determined manner whenever task-driven approaches are not applicable or undesirable.

Furthermore, while our logging data confirmed findings on equal data-entry performance in any posture [14], participants in our study oftentimes preferred the sitting workstation, mainly due to reasons of personal habit or convenience. Especially for participants with self-determined switching behavior, we think that subtle reminding mechanisms (e.g., ambient displays [10]) can assist the adoption of a sustainable behavior change. For a task-driven approach, this may not be needed as long as the daily work routine consists of a diverse range of tasks. This is where assignment mechanisms (e.g., intelligent window arrangement algorithms [13]) can provide assistance for users to achieve a balanced task distribution among individual workstations.

5.2 Supporting Distributed Multi-Monitor Usage

Considering the study results, we see that physical separation introduced with distribution of multiple monitors across the work environment fundamentally changed the way people perceived and handled them. Although the monitors were connected to one machine, enlarged distances coupled with differences in height and relative orientation to the user introduced visual discontinuities [27], which presented major barriers to associated multi-monitor interaction. In contrast to adjacent multi-monitor setups, where users slightly rotate their neck to cover the increased screen space [26], the physical gaps in the present setup required them to additionally change their orientation or even physical location to switch focus between screen content spread across different monitors. To make effective use of both workstations, we identify a variety of challenges for designing future systems to support the physical distribution through *exclusive*, *supportive*, or *peripheral* multi-monitor usage [9]:

- **Exclusive** usage implies the association of the physical distribution with corresponding *logical separation*–virtually treating the monitor(s) and input devices at each workstation as a single logical unit. The basic functionality of the Win+P display switching mechanism could be extended or automated to activate a workstation monitor whenever the corresponding mouse or keyboard are operated.

  Although exclusive monitor usage solves certain problems described earlier (e.g., window placement inconsistency, cross-monitor interaction), a major drawback remains. Participants expressed the desire to make use of the secondary monitor, since they felt like they were not using the full potential of their workstations. Concerns were raised that having two monitors but only having one in use was perceived as waste of resources (basically losing half of their working area).

- **Supportive** usage implies the association of the physical distribution with corresponding *visual separation,* thereby treating the displays at each workstation as visually isolated spaces. Digital barriers can serve as a basis for the restriction of mouse cursor movement [2], border snapping mechanisms, or consistent placement of windows on the current screen.

  The effort of moving digital objects between monitors was identified as major barrier for fluent transitions from one workstation to another, implying the need to support effortless window management mechanisms such as grouping [25], switching, or relocating [1, 3] digital content on the individual screens.

- **Peripheral** usage implies the adaption of the user interface and interactions, since participants avoided interactions with the secondary monitor due to non-optimal mouse support and poor visibility of digital content across the distance in extended display mode. In response to that, interface elements on the distant screen could be enlarged and simplified to achieve a good readability, or convenient access to these elements should be supported through novel (multi-modal) interaction styles.

Essentially, users need to be able to equally access both workstations and the switching process needs to be effortless and seamless to maintain their motivation for the adoption of a physically active work process. Some participants were favoring the workstation with the primary screen (thus not switching workstations) because of existing software limitations. Avoiding these limitations to provide users with a seamless switching experience is a very important factor for this kind of activity-promoting office environments, though. Therefore, it seems crucial to gain knowledge about the user's activity (e.g., the currently used workstation) and properties of the surrounding environment (e.g., presence of other people). Based on that knowledge, an intelligent office space could for example respond to a user's activity by bringing digital objects into his focus of attention automatically [4].

6 Conclusion and Future Work

In this paper, we presented a background study with five office workers, who were carrying out their daily work tasks within a sit-and-stand workspace environment that follows the concept of integrating light activity into the daily work routine by providing the opportunity to work in sitting and standing posture [23]. Prior research in the field of multi-monitor usage is extended through the discussion of user behavior associated with the concept of working "in-motion". We indicated a number of challenges originating from user requirements within such an environment and discussed effects of different influencing factors on participants' workstation switching behavior. We identified two basic approaches for physical activity integration, and concluded that the situation-related support of both task-driven and self-determined patterns has high potential to provide users with a seamless switching experience. In a distributed multi-monitor setup, however, the physical separation of the individual displays has to be handled in a different way than in regular adjacent multi-monitor setups. Observations from our study showed that the distribution of monitors fundamentally changed the way people perceived and used them, and therefore experienced a variety of problems associated with running the monitors in extended, duplicated, or relocated mode. We highlighted resulting implications for the HCI design of future systems to support the seamless alternation of postures while working in such an environment.

Based on the findings of this study, future work will focus on the development of novel user interfaces and interaction techniques that provide an optimal interaction for exclusive, supportive and peripheral usage of distributed multi-monitor setups. Sensing methods will be developed on multiple levels to gain knowledge about the user's activity within the office workspace.

Another interesting insight provided by the presented study was that especially the standing workstation can be a great addition for collaboration and communication within an office space, which will also be addressed through the development of specific software support for such collaborative settings. Based on the findings on users' workplace switching behavior, we will examine these patterns more closely including the possibility to use and trigger them to promote workplace switching. Finally, future research will include additional studies with a larger number of participants and more diverse hardware setups (including e.g. notebooks and other mobile devices), and the investigation of challenges arising with the utilization of these setups within activity-promoting office environments.

Acknowledgements. We would like to thank Dr. Sheelagh Carpendale and the InnoVis group at the University of Calgary for their ideas and advice. Funding for this research was provided by the Austrian Research Promotion Agency GmbH (FFG) within the Bridge project #2579917.

References

1. Baudisch, P., Cutrell, E., Robbins, D., Czerwinski, M., Tandler, P., Bederson, B., Zierlinger, A.: Drag-and-Pop and Drag-and-Pick: techniques for accessing remote screen content on touch- and pen-operated systems. In: Interact 2003, pp. 57–64. IOS Press, Amsterdam (2003)
2. Benko, H., Feiner, S.: Pointer Warping in Heterogeneous Multi-Monitor Environments. In: GI 2007, pp. 111–117. ACM Press, New York (2007)
3. Biehl, J., Bailey, B.: ARIS: An Interface for Application Relocation in an Interactive Space. In: GI 2004, pp. 107–116. Canadian Human-Computer Communications Society, London (2004)
4. Brumitt, B., Meyers, B., Krumm, J., Kern, A., Shafer, S.: EasyLiving: Technologies for Intelligent Environments. In: Thomas, P., Gellersen, H.-W. (eds.) HUC 2000. LNCS, vol. 1927, pp. 12–29. Springer, Heidelberg (2000)
5. Craft, L.L., Zderic, T., Gapstur, S., Vaniterson, E., Thomas, D., Siddique, J., Hamilton, M.: Evidence that women meeting physical activity guidelines do not sit less: An observational inclinometry study. International Journal of Behavioral Nutrition and Physical Activity 9, 122 (2012)
6. Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., Starkweather, G.: Toward Characterizing the Productivity Benefits of Very Large Displays. In: Interact 2003, pp. 9–16. IOS Press, Amsterdam (2003)
7. Freak-Poli, R., Wolfe, R., Peeters, A.: Risk of Cardiovascular Disease and Diabetes in a Working Population With Sedentary Occupations. Journal of Occupational and Environmental Medicine 52(11), 1132–1137 (2010)
8. Gilson, N., Suppini, A., Ryde, G., Brown, H., Brown, W.: Does the use of standing "hot" desks change sedentary work time in an open plan office? Preventive Medicine 54(1), 65–67 (2012)
9. Grudin, J.: Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In: CHI 2001, pp. 458–465. ACM Press, New York (2001)
10. Haller, M., Richter, C., Brandl, P., Gross, S., Schossleitner, G., Schrempf, A., Nii, H., Sugimoto, M., Inami, M.: Finding the right way for interrupting people improving their sitting posture. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 1–17. Springer, Heidelberg (2011)

11. Healy, G., Lawler, S., Thorp, A., Neuhaus, M., Robson, E., Owen, N., Dunstan, D.: Reducing prolonged sitting in the workplace. Human Factors and Ergonomics Society, Melbourne (2012)
12. Hedge, A., Ray, E.J.: Effects of an electric height-adjustable worksurface on self-assessed musculoskeletal discomfort and productivity in computer workers. In: HFES 2004, pp. 1091–1095. Mira Digital Publishing, Brentwood (2004)
13. Henderson, D., Card, S.: Rooms: The Use of Multiple Virtual Workspaces to Reduce Space Contention in a Window-Based Graphical User Interface. ACM Transactions on Graphics 5(3), 211–243 (1986)
14. Husemann, B., Von Mach, C., Borsotto, D., Zepf, K., Scharnbacher, J.: Comparisons of Musculoskeletal Complaints and Data Entry Between a Sitting and a Sit-Stand Workstation Paradigm. Human Factors 51(3), 310–320 (2009)
15. Hutchings, D., Stasko, J.: Consistency, Multiple Monitors, and Multiple Windows. In: CHI 2007, pp. 211–214. ACM Press, New York (2007)
16. Hutchings, D., Stasko, J., Czerwinski, M.: Distributed Display Environments. Interactions 23(6), 50–53 (2005)
17. Hutchings, D.R., Smith, G., Meyers, B., Czerwinski, M., Robertson, G.: Display Space Usage and Window Management Operation Comparisons between Single Monitor and Multiple Monitor Users. In: AVI 2004, pp. 32–39. ACM Press, New York (2004)
18. Levine, J., Miller, J.: The energy expenditure of using a '"walk-and-work'" desk for office workers with obesity. British Journal of Sports Medicine 41(9), 558–561 (2007)
19. Levine, J.A.: Non-Exercise Activity Thermogenesis (NEAT). Best Practice & Research: Clinical Endocrinology & Metabolism 14(4), 679–702 (2002)
20. McAlpine, D., Manohar, C., McCrady, S., Hensrud, D., Levine, J.: An office-place stepping device to promote workplace physical activity. British Journal of Sports Medicine 41(12), 903–907 (2007)
21. Miller, R., Brown, W.: Steps and sitting in a working population. International Journal of Behavioral Medicine 11(4), 219–224 (2004)
22. Owen, N., Bauman, A., Brown, W.: Too much sitting: a novel and important predictor of chronic disease risk? British Journal of Sports Medicine 43(2), 81–83 (2009)
23. Probst, K., Leitner, J., Perteneder, F., Haller, M., Schrempf, A., Glöckl, J.: Active Office: Towards an Activity-Promoting Office Workplace Design. In: CHI Ext. Abstracts 2012, pp. 2165–2170. ACM Press, New York (2012)
24. Robertson, G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., Tan, D.: The Large-Display User Experience. IEEE Comput. Graph. Appl. 25, 2–9 (2005)
25. Smith, G., Baudisch, P., Robertson, G., Czerwinski, M., Meyers, B., Robbins, D., Andrews, D.: GroupBar: The TaskBar Evolved. In: OZCHI 2003, pp. 34–43. CHISIG of Australia, Canberra (2003)
26. Swaminathan, K., Sato, S.: Interaction Design for Large Displays. Interactions 4(1), 15–24 (1997)
27. Tan, D., Czerwinski, M.: Effects of Visual Separation and Physical Discontinuities when Distributing Information across Multiple Displays. In: Interact 2003, pp. 252–255. IOS Press, Amsterdam (2003)
28. WHO: Preventing Noncommunicable Diseases in the Workplace through Diet and Physical Activity: WHO/World Economic Forum Report of a Joint Event (2008)
29. Wilks, S., Mortimer, M., Nylén, P.: The introduction of sit-stand worktables; aspects of attitudes, compliance and satisfaction. Applied Ergonomics 37(3), 359–365 (2006)
30. Wilmot, E., Edwardson, C., Achana, F., Davies, M., Gorely, T., Gray, L.J., Khunti, K., Yates, T., Biddle, S.: Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 55(11), 2895–2905 (2012)

Modeless Pointing with Low-Precision Wrist Movements

Theophanis Tsandilas[2,3], Emmanuel Dubois[1], and Mathieu Raynal[1]

[1] IRIT – Elipse, University of Toulouse, Toulouse, France
[2] INRIA, Orsay, France
[3] Univ Paris-Sud (LRI) & CNRS, Orsay, France
Theophanis.Tsandilas@inria.fr,
{Emmanuel.Dubois,Mathieu.Raynal}@irit.fr

Abstract. Wrist movements are physically constrained and take place within a small range around the hand's rest position. We explore pointing techniques that deal with the physical constraints of the wrist and extend the range of its input without making use of explicit mode-switching mechanisms. Taking into account elastic properties of the human joints, we investigate designs based on rate control. In addition to pure rate control, we examine a hybrid technique that combines position and rate-control and a technique that applies non-uniform position-control mappings. Our experimental results suggest that rate control is particularly effective under low-precision input and long target distances. Hybrid and non-uniform position-control mappings, on the other hand, result in higher precision and become more effective as input precision increases.

Keywords: Pointing techniques, constrained wrist movement, elastic devices, rate control, clutching.

1 Introduction

Technology becomes more and more ubiquitous, and a variety of handheld devices such as Wii Remotes start becoming widely available. Researchers and designers have been envisioning scenarios that move user interaction beyond desktop computers. A user interacts with a public display in a museum to get more information about an exhibit or with a wall display in an airport to learn more about a flight. Driven by such scenarios, research in Human-Computer Interaction has been exploring new pointing techniques that go beyond the use of a mouse. For example, Vogel and Balakrishnan [23] proposed techniques for freehand pointing in front of high-resolution displays. Other work [22] has explored the use of mobile phones as pointing devices for interaction with public displays. In such scenarios, user interaction does not rely on the presence of specialized input devices and sophisticated motion tracking systems. Pointing precision becomes an important issue. Factors that affect pointing precision include natural hand tremor and limited hand precision [2, 21] or the use of low-resolution input devices.

Several solutions have tried to improve pointing precision by proposing mechanisms that balance between absolute and relative pointing. "Clutching" is a common

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 494–511, 2013.

input recalibration mechanism that allows users to reposition the frame of reference of absolute pointing. Despite its simplicity, it requires explicit mode transitions between pointing and clutching. When using a mouse, clutching is activated naturally, by lifting the mouse and translating it over the table. However, when the input device is positioned in free space, mode switching relies on the use of a button [24] or a special hand gesture [23]. Such mechanisms may not be available in certain situations (e.g., when the user manipulates a physical object with no additional input channels), rely on error-prone hand-gesture recognition mechanisms or compete with the activation of other actions, such as the selection of a target.

This paper explores solutions that achieve pointing precision without requiring the user to switch between modes. We focus on movements of the wrist, which are important for the manipulation of handheld devices and can be easily detected by accelerometer and gyroscope sensors of common mobile devices. We investigate three solutions. Based on the observation that wrist movements occur within a limited range, beyond which natural self-centering forces develop, we investigate rate (velocity) control as a first solution. Inspired by the work of Casiez et al. [5], we explore a second technique that combines position and rate control to extend the input range of the hand. We claim that the wrist can be viewed as a hybrid device: isotonic when movements take place around its neutral position, and elastic when its position extends beyond a certain range. Finally, we examine a third technique influenced by previous research on multi-scale pointing [12] and non-uniform position-control mappings [1, 20]. We show that rate control is insensitive to input precision, which means that it is particularly valuable when input precision is low. We measure and control input precision by dividing the input range into discrete units of input. We use input precision as an independent variable to explore the design space and evaluate the three techniques.

2 Related Work

In indirect pointing, the pointing device is decoupled from the display. This separation results in two different spaces: (1) the *display* space, which is the space of the pointer's movement, and (2) the *motor* (or *control*) space, where the manipulation of the input device takes place. In *relative-pointing* devices, the mapping between the two spaces can change dynamically, for example, by using recalibration mechanisms that change the frame of reference of pointing on the display space. *Clutching* is the most common recalibration mechanism [13]. When using a mouse, clutching can be performed by lifting and moving the mouse off the table. When pointing occurs in free space, clutching requires the activation of an explicit mode-switching mechanism, for example, by pressing a button [24] or by changing the hand's posture [23].

The Control-Display (C-D) gain [11] maps the movement of the input device (motor space) to the movement of the display pointer (display space). It can be calculated as the ratio of the pointer velocity to the velocity of the input device. Casiez et al. [6] examined the impact of C-D gains on pointing performance with a high-resolution mouse. Their results indicate that pointing time follows an L curve as a function of C-D gain. When the gain is low, performance slows down because clutching becomes frequent. When it is high enough, the C-D has no effect on user performance, i.e., increasing its value has no cost. Problems due to the accuracy of hands and fingers appear in tiny movements of about 0.2 mm. When the hand moves in free space, however, problems appear in larger movements in the range of 3 – 5 mm [2]. In this case, the use of high C-D gains is problematic. For example, ray casting, which involves high C-D gains when used at a distance, is extremely sensitive to hand tremor and results in high error rates [18, 19, 23].

A pointing device can control either the position (*position control*) or the velocity (*rate control*) of the pointer. Position control is most often used with *isotonic* devices, e.g., a mouse, while rate control is used with *isometric* and *elastic* devices, e.g., joysticks. Rate control eliminates the need for clutching but relies on a self-centering mechanism, not present in isotonic devices. Zhai [24] compared all the possible mappings between input types (isotonic and isometric) and types of control (position and rate) and found that isotonic rate-control input was about 50% slower than isometric rate-control input and isotonic position-control input. More recently, Casiez and Vogel [5] examined the effect of the stiffness of elastic devices on the performance of rate control. They found that rate control performed well even for very low stiffness values, as long as a self-centering mechanism was present and velocity-control functions were carefully selected. They also found that pointing performance was only 15% slower when stiffness was zero, i.e., when pointer control was purely isotonic.

Research in Virtual Reality has proposed hybrid movement mappings to facilitate the manipulation of virtual objects, avoiding the use of high C-D gains, explicit interaction modes, and clutching mechanisms. The Go-Go technique introduced by Poupyrev et al. [20] extends the reach of a user's hand in the virtual world by applying non-uniform C-D gains. The technique uses a one-to-one movement mapping between the real and a virtual hand as long as the hand stays within a fixed area around the user. This ensures that users can manipulate nearby objects with precision. However, when the arm of the user extends beyond the proximity area, the C-D gain grows through a non-linear (parabolic) function. Variable C-D gains can help users reach remote objects without having to "clutch". A major drawback of the technique is that high gains develop towards extreme input positions, requiring users to bring remote objects closer so that they can effectively interact with them. Bowman and Hodges [3] explored variations of the Go-Go technique, including the *stretch go-go* technique, which controls the velocity of the virtual arm rather than its absolute position. Unfortunately, little evidence about the effectiveness of the above techniques exists. More recently, Appert et al. [1] introduced the *Ring lens*, a high-precision magnification lens that applies position control at two scales. The cursor moves with a low C-D gain within a Ring lens, supporting high-precision control in the magnified area. As the cursor reaches the border (ring) of the lens, the lens follows the movement of the cursor. The technique outperformed both speed-depended precision control and mode–switching precision control.

A few approaches have examined hybrid designs. Dominjon et al. [9] proposed an interaction technique for haptic devices that combines position and rate control.

The technique is based on the visualization of the input space as a three-dimensional bubble. Movement is isotonic and position controlled inside the bubble, but it becomes elastic and rate controlled as the cursor crosses the boundaries of the bubble to the outside. Dominjon et al. [9] implemented the bubble technique with a PHANTOM haptic device and evaluated it on 3D-model painting tasks. The technique was more efficient than both absolute positioning and clutching and received higher subjective rankings. RubberEdge [7] is a hybrid device that applies a similar approach to 2D pointing tasks. It resembles to a regular touch pad, but its boundaries are elastic. When the finger of the user moves from the central isotonic zone of the pad to its elastic boundaries, cursor movement becomes rate controlled, allowing the user to traverse long distances without clutching.

3 The Wrist as a Constrained Pointing Device

Wrist movements are constrained by the hand's joints. Joint constraints decrease the operating range of input, hindering user performance. This is a major problem when users need to interact with large visual spaces without losing in pointing precision. Clutching techniques as well as techniques that let users switch between relative and absolute pointing can increase the range of movement but require an additional input channel. Unfortunately, such mechanisms may not be available in certain situations, for example, when users interact with non-specialized devices or physical objects.

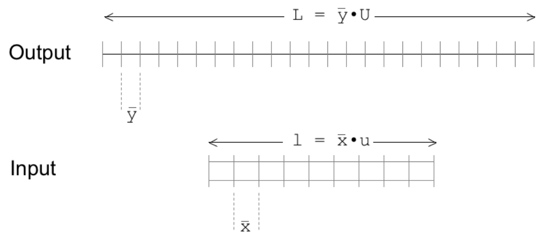


Fig. 1. Input and output unit ranges

Physical constraints may not pose a problem if movements of the hand are highly accurate, when pointing distances are relatively short or targets are large. Problems arise when hand movements are imprecise, distances are large, and targets are small. To better describe this problem we study user performance in relation to the number of discrete units of movement. Fig. 1 presents how the input and output space can be partitioned into units of movement. U is the range of the display measured in units, and $\bar{y}$ is the size (in mm, dots or pixels) of the minimum unit of output movement. Clearly, targets on the display have to be larger than $\bar{y}$ to be selectable. Chapuis and Dragicevic [8] found that the discretization of the output had practically no effect on pointing performance as long as targets were selectable. This implies that performance only depends on motor-space control and not on the visual representation of the cursor. In most real applications, the minimum output movement $\bar{y}$ is never larger than 1 pixel, which allows for smooth cursor transitions and the selection of tiny targets. In this case, U represents the display size in pixels. Similarly to output range, we can view input range as a discretized entity, where u is the range of detectable discrete input units, and $\overline{x}$ is the size of each unit in mm. The minimum input unit $\overline{x}$ represents the minimum distance that can be recognized by an input device and controlled by the user. When the number of output units is greater than the number of input units ($U > u$), the user cannot reach all the output units with an absolute mapping between input and output. Given an input-output configuration (u, U), where U is much greater than u ($U >> u$), our goal is to examine pointing techniques that maximize pointing performance. The output-to-input ratio:

$$r_{io} = \frac{U}{u} \tag{1}$$

determines the minimum number of clutches required to traverse the whole output range. An alternative definition of the *C-D* gain in terms of input and output units is:

$$CDU_{gain} = \frac{U_m}{u_m} \tag{2}$$

where U_m is the number of output units covered when the input device moves u_m units. Notice that, in contrast to the *C-D* gain, the above measure is not unit-free.

Research in biomechanics has proposed several models to describe the mechanics of joints and muscles. Lemay and Crago [16] used a simple mathematical model to express passive elastic torques M in human limbs, according to which the joint's stiffness is composed of both a linear and an exponential component. The exponential component dominates close to the limits of a limb's operating range, but movement is highly isotonic around the rest position. Lehman and Calhoun [15] studied extensions and flexions of the wrist. They found that wrist movements are isotonic in a range of 40 degrees to the left and the right of the rotational axis. Passive torques were less than 0.1 N·m within this movement range. Elastic torques started appearing rapidly as movement extended towards its extreme positions. Fig. 2 presents the angular range of wrist rotations and our hypothetical model of the wrist as a hybrid device.

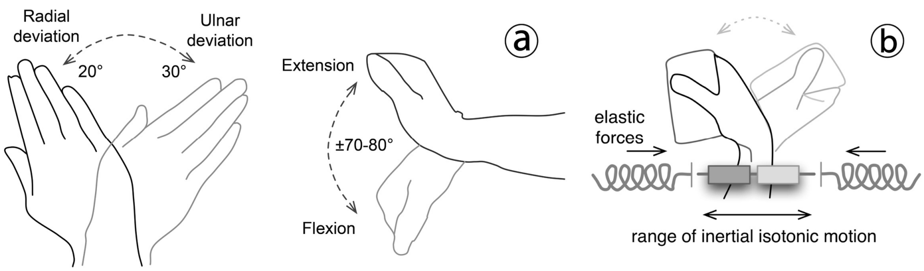


Fig. 2. The wrist. (a) The range of its angular movements. (b) Modeled as a hybrid of an isotonic and an elastic device.

The precision of such wrist movements can be particularly low. Rahman et al. [21] found that errors inflated considerably as users tried to control more than 12 levels of purely flexion-based tilts when interacting with a mobile phone.

4 Techniques

We examined three techniques that allow the user to point by moving the wrist without making use of explicit mode-switching mechanisms. Motivated by previous work on hybrid movement mappings [7, 9, 20], we reconsider the use of rate control in conjunction with wrist movements. Rate control depends on the existence of a self-centering mechanism. Previous results [5] suggest that the stiffness of an elastic device has no effect on the effectiveness of rate control. Rate control can work equally well with low stiffness values provided that a self-centering mechanism is present. Their results also suggest that when movement occurs within a small range, the removal of external self-centering mechanisms only slightly affects user performance. Given that the wrist has a natural resting position and movement takes place within a relatively small range around it, we expected that rate control would be a viable solution even if we did not externally reinforce the wrist's centering mechanism.

4.1 Pure Rate Control

Rate control is based on the application of a transfer function $f_{d \rightarrow v}$ that maps the displacement *d* from a neutral input position to velocity units. Previous work [5, 24] has made use of linear transfer functions. Our informal tests, however, showed that non-linear transfer functions result in better motor control. This can be explained by the fact that the neutral position of the wrist is not strict but expands within a certain range where self-centering is absent and velocity control is difficult.

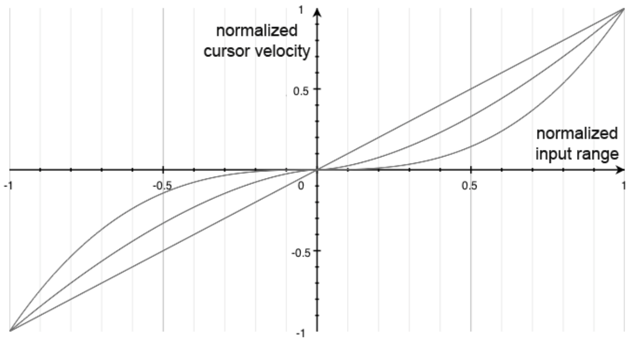


Fig. 3. Three different transfer functions: $v(x) = sign(x) \cdot |x|^b$, where $b \geq 1$

The form of polynomial functions (see Fig. 3) that we tested is the following:

$$v(x) = sign(x) \cdot |x|^b, b \geq 1 \quad (3)$$

where $x = 0$ is the hand's resting position. Both the input range and velocity are normalized in [-1, 1]. We observed that transfer functions that grow rapidly around the zero position, i.e., when *b* is low, are sensitive to input precision. Velocities must be low enough so that small targets can be easily selected. If the width of the smallest target is W_{min}, the minimum cursor velocity v_{min} must be as follows:

$$v_{\min} \leq \frac{W_{\min}}{T_{react}} \quad (4)$$

where T_{react} is the delay for the user to react and stop the movement when the cursor enters the area of the target. Based on previous experimental results [17], we estimated this delay to be around 150 ms.

4.2 Hybrid Control

We explored mixed-control designs, balancing between high precision afforded by position control and smooth long-distance movement afforded by rate control. Our approach is based on existing techniques [7, 9] but does not assume the availability of specialized elastic devices. As shown in Fig. 4, we divide the input range into three zones. The central zone is reserved for position control. The portion of the display that corresponds to movements within this zone is communicated to the user as a framed window. The side input zones let the user reposition this window by controlling its velocity. As a result, pointing takes place at two stages. First, the user turns the hand out of its central zone to bring the window around the target. Then, the hand returns towards the central zone to point to the target within the window's boundaries.

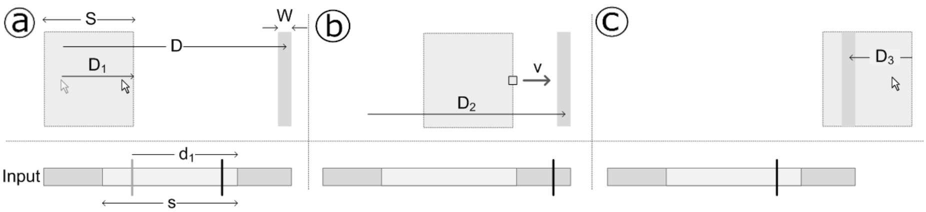


Fig. 4. Steps of a selection task with the hybrid technique when $D > D_1$. The vertical line in the input space shows the input position. (a) The cursor travels a distance D_1, which corresponds to a movement d_1 along the isotonic range of the input. (b) The window travels towards the target, which appears in a distance D_2 from the initial center of the window. The velocity v of the window increases as the input position extends towards its extreme sides. (c) The cursor moves to the target by covering a distance D_3.

We follow the approach of Dominjon et al. [9], who visually communicate the range of position control, rather than the approach of Casiez et al. [7], who use a single cursor representation. In our case, there is no haptic feedback to communicate transitions to the user, and hence, visual feedback is essential. We also found that breaking the pointing task into two different scales, i.e., rough pointing with rate control, and precise pointing with position control, was more appropriate than assuming that rate and position control constitute symmetric ways of pointing.

Clearly, the range of input units determines the maximum size of the window, but how to divide this range is not straightforward. Larger central zones increase the active width of position-controlled pointing but do not necessarily result in faster movements because (1) a shorter range is reserved for rate control, and (2) transitions from position to rate control, and inversely, require longer hand movements. The model proposed by Casiez et al. [7] assumes that movement has two parts. Movement first occurs in the isotonic area until the user's finger reaches the elastic boundary of the device. Then, movement continues in the elastic zone until the target is finally selected. In contrast, our model assumes that movement always returns to the central isotonic zone to complete the pointing task. Our early tests have shown that direct target selection with rate control is particularly hard when a hybrid design is used. When movement takes place in a single side (i.e., elastic) zone of the input, the cursor can only move towards a single direction. If the cursor overshoots the target, movement has to return to the isotonic zone. Fig. 4 demonstrates the steps of a selection task based on this model.

4.3 Non-uniform Position Control

We have designed a window-based technique that applies position control to move the window of high-precision pointing. As shown in Fig. 5, a movement of the hand out of its central zone translates the window to the left or to the right by using a high C-D gain. The position of the window freezes when the wrist starts moving towards its central zone. This mechanism allows the user to recalibrate the movements, keeping high-precision pointing around the wrist's rest position. During recalibration, the user may have to repeat multiple forward and backward movements before bringing a distant target within the window. More precisely, the number of such movements depends on the target's distance, the input resolution, the size of the central zone, and the selected C-D gains. This recalibration mechanism is analogous to regular clutching mechanisms but does not require the use of additional input channels. Also, it makes use of two distinct gains (CDU_{low}, CDU_{high}), which allows for minimizing the number of recalibration actions. The steps required by the third technique to point to distant targets are similar to the ones required by the hybrid technique except that now position control is used to position the window. The cursor moves to the boundaries of the window. Then, the window moves to the target, following a series of recalibration actions. Finally, the cursor points to the target.

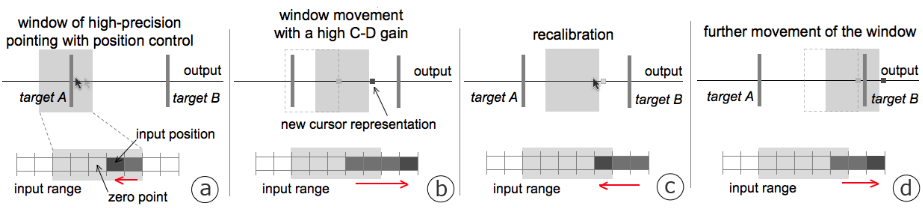


Fig. 5. Non-uniform position control. (a) Absolute pointing in the central zone of the input range with a low C-D gain. (b-d) Position recalibration with back and forth movements.

If w units of a total of u input units are reserved for absolute pointing, the maximum output distance d traversed without recalibration can be computed as follows:

$$d = w \cdot CDU_{low} + \frac{u - w}{2} \cdot CDU_{high} \tag{5}$$

To maximize this distance and, at the same time, ensure that all the pixels can be visited, we can select CDU_{low} to be equal to 1 pixel/unit and CDU_{high} to be equal to w:

$$d_{\max} = -\frac{w^2}{2} + \left(1 + \frac{u}{2}\right) \cdot w \tag{6}$$

The value w_o that maximizes this expression can be calculated as follows:

$$\left.\frac{\partial d}{\partial w}\right|_{w_o} = 0 \Leftrightarrow -w_o + \left(1 + \frac{u}{2}\right) = 0 \Leftrightarrow w_o = 1 + \frac{u}{2} \tag{7}$$

This equation shows that the distance is maximized when the zone of absolute pointing is approximately half of the input range. This estimation has been based on the highest possible value for CDU_{high}. As we have already discussed, previous results [6] show that C-D gains do not impact pointing performance as long as no clutching and input resolution problems arise. Besides, other results [8] show that the discretization of the cursor's movement does not hurt user performance. Yet, our own tests showed that high gains could hinder motor control. Our explanation is that non-uniform control requires users to switch from a low to a high gain, adapting accordingly their movement strategies. It seems that the higher the difference between the two gains, the higher becomes the cost of such movement adaptations. Therefore, we select CDU_{high} to be equal to w/a, where $a>1$. Then, based again on Equation 5, we find that the value w_o that maximizes the distance d is as follows:

$$w_o = a + \frac{u}{2} \tag{8}$$

In practice, the best values for a and w must be empirically selected.

5 Experiment

We conducted an experiment to compare the performance of the three techniques. We focused on low-precision wrist input with up to 241 input units.

5.1 Participants and Apparatus

Twelve volunteers (three women and nine men), 21 to 40 years old (the median age was 27) participated. Two participants used their left hand to perform the tasks. One participant had also participated in the pilot experiments (see below).

We used an Ascension Flock of Birds 6-DOF motion tracker to detect extensions and flexions of the wrist around a vertical axis. Participants were seated and moved a sensor cube (25.4mm × 25.4mm × 20.3mm) within a range of about 50 - 80 cm from the transmitter and a distance of about 70 – 80 cm from the monitor. We attached the sensor cube on a solid pen-like extension (see Fig. 6c), allowing participants to grab it more comfortably. We used a 22-inches monitor with a 1680 × 1050 resolution. We detected only rotations parallel to the ground and used a simple Kalman filter to remove noise. To control for input precision, we discretized the rotation values measured by the magnetic tracker by diving the effective rotational range into discrete input units. The experimental software was written in Java 1.6.

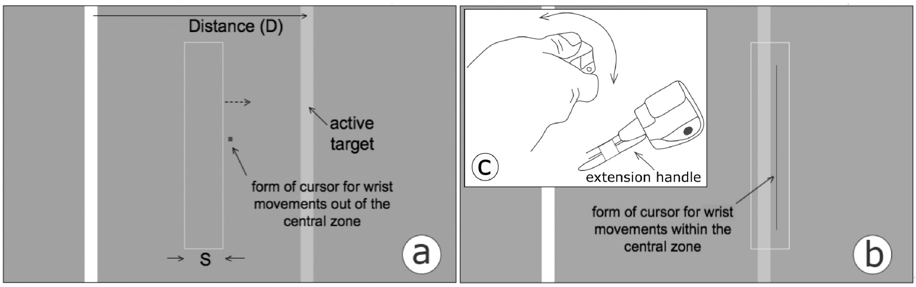


Fig. 6. Experimental task for the hybrid and the non-uniform position-control technique. (a) When the hand moves out of the central zone, the cursor becomes inactive and takes the form of a small rectangle. The short distance between this rectangle and the boundaries of the window provides feedback about input position out of the central zone. (b) A move of the hand back to the central zone activates the cursor, which takes its regular line form.

5.2 Task and Stimuli

Participants performed a series of reciprocal 1D pointing tasks by selecting two targets forward and then backward in succession. Targets were rendered as solid vertical bars and were selected by pressing a key with the non-dominant hand. Participants were required to successfully select a target before moving to the next. They were told to perform tasks "as fast as possible, trying to avoid errors".

For the rate technique, the cursor was rendered as a one-pixel-thick vertical line. For the hybrid and the non-uniform position-control technique, we used two cursor representations to communicate the two different levels of input control, as shown in Fig. 6. When the wrist of the user moved out of the central zone, the cursor turned into a small rectangle and passed control to the window, which started moving. Its position relative to the boundaries of the window slightly changed, providing direct feedback about the position of the wrist with respect to the central zone of input.

5.3 Optimizing the Techniques: Summary of Two Pilot Experiments

For each technique, we had to choose several parameters. What transfer function to use? How to split the input range? What *C-D* gains to choose? To reduce the design space, we conducted two pilot experiments. Six volunteers participated in each pilot.

Pilot 1. The first experiment tested pure rate control under three transfer functions (Fig. 7a) and two levels of low input precision: 21 and 61 input units. We selected the functions f_2 and f_3 to approximate an optimal transfer function for both levels of precision. We expected that the transfer function f_1 would be the slowest one, especially under the low input precision. Under a range of 21 units, the lowest velocity allowed by f_1 is $2 \cdot 0.1^{1.5}$ pixels/ms or 63 pixels/sec. For targets of 8 pixels and minimum response times of 150 ms, this value is higher than 53 pixels/sec, which is the velocity limit calculated by Equation 2.

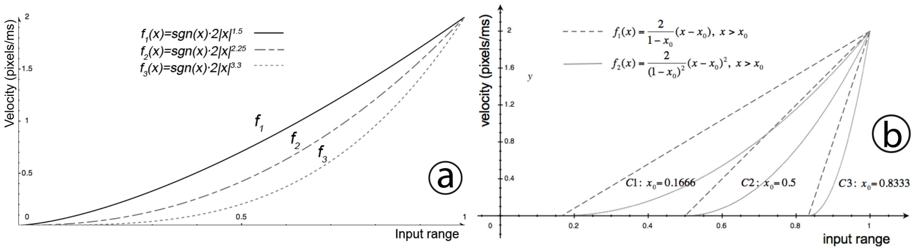


Fig. 7. Transfer functions tested by the two pilot studies, shown for the right half of the input range: (a) Pure rate control, and (b) Hybrid control

The transfer function has a significant effect on selection time ($F_{1.68,\ 8.38}$=15.47, p=0.002) and error ($\chi^2_{df=2,N=6}$=10.17, p=0.006). The results are summarized in Fig. 8. The error rate is considerably high for f_1, especially when input precision is low. The best function for errors is f_3. The results suggest that as long as an optimal transfer function is selected, there is no clear penalty for reducing the input precision: errors rates are not hurt and the cost in speed is minimal, less than 5%.

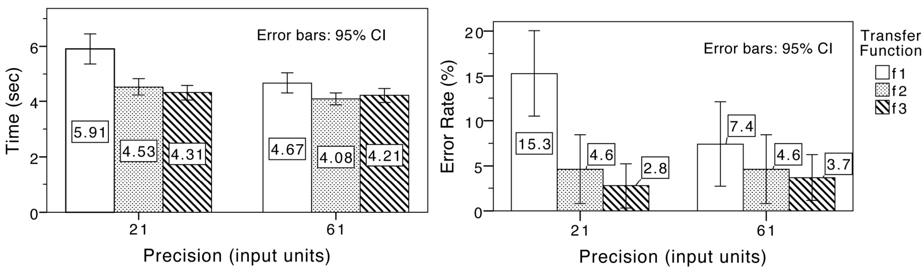


Fig. 8. Overall results for Pilot 1

Pilot 2. The second pilot explored issues related to the design of the hybrid technique. We compared three assignments of input range to isotonic and elastic zones. More specifically, we tested a single level of input precision of 61 input units under the three configurations shown in Table 1. Presentation order was counterbalanced among participants. The CDU gain for cursor's movement was set to 4 pixels/units. As a result, the size *S* of the area cursor for the three configurations was 44, 124, and 204 pixels, respectively. For each configuration, we empirically selected two transfer functions (see Fig. 7): a linear function f_1, and a parabolic function f_2.

Table 1. Input configurations tested by Pilot 2

Config.	Left Zone % (units)	Right Zone % (units)	Central Zone % (units)	*S* (pixels)
C1	40.98% (25)	40.98% (25)	18.03% (11)	44
C2	24.59% (15)	24.59% (15)	50.82% (31)	124
C3	8.20% (5)	8.20% (5)	83.61% (51)	204

Results are summarized in Fig. 9. An ANOVA repeated-measures analysis indicates a significant main effect of input configuration on selection time

($F_{1.67,8.33}$ =37.84, p<0.0001). Pairwise comparisons with Bonferroni's adjustment show that C1 is significantly slower than both C2 (p=0.001) and C3 (p=0.004). There is no significant difference between C2 and C3 (p=1.0). The mean error rate is also higher for C1 but with no consistency among participants. The pointing time is not significantly different for the two transfer functions ($F_{1,5}$=1.4, p=0.29), but the mean error is lower for f_1 under C2 and C3. In conclusion, we can select a configuration with f_1 as transfer function, where 50% to 84% of the input range is dedicated to the central zone.

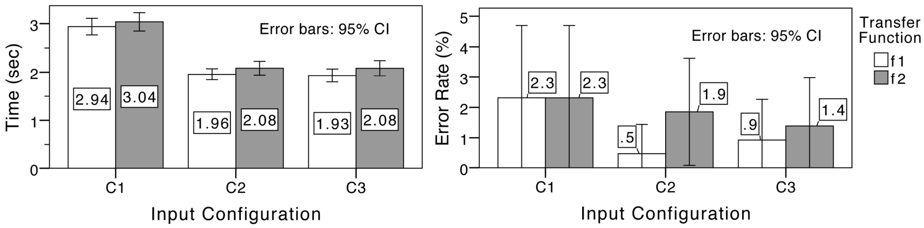


Fig. 9. Overall results for Pilot 2

5.4 Conditions

The experiment tested three pointing techniques: rate control (Rate), hybrid control (Hybrid), and non-uniform position control (Position). For each one, the experiment tested two levels of input precision: 61 (low) and 241 (high) discrete input units. These values are higher than the ones tested by Pilot 1 and 2, but they still correspond to a low input precision. We made sure that participants could visit every pixel on a window of 61 or 241 pixels with our configuration when using a simple position-control technique. We also verified that the Device's Human Resolution [2] was high enough so that participants could comfortably select the smallest 8-pixel targets. For each condition, we tried to optimize user performance by refining input control based on our theoretical analysis and the two pilot studies. Below, we justify how various parameters for each technique were selected.

Rate Control. For the low-precision condition, we used the transfer function $f_3(x)=sgn(x)\cdot 2|x|^{3.3}$. Pilot 1 showed that this function provided the best tradeoff between time and errors. However, it was unclear whether this function would be optimal under larger numbers of discrete input units. Informal tests with 241 units showed optimal results for functions ranging between f_2 and f_3. We chose $f(x) =sgn(x)\cdot 2|x|^{2.5}$.

Hybrid Control. The CDU gain for position-control was set to 1 pixel/unit. This gain enables the selection of every pixel and makes optimal use of the available input range. Table 2 shows how we divided the input range for the two levels of input precision. For both, we used a linear transfer function (see f_1 in Fig. 7b). Notice that we avoided using large central zones, as they would result in longer sub-movements for the final target selection.

Table 2. Input configurations tested by the experiment

Precision (units)	Left Zone % (units)	Right Zone % (units)	Central Zone % (units)	S (pixels)
61	19.67% (12)	19.67% (12)	60.66% (37)	37
241	24.90% (60)	24.90% (60)	50.21% (121)	121

Position Control. Equation 8 calculates the size (in input units) of the zone of absolute pointing that minimizes the number of recalibration actions. This zone is approximately half of the input range provided that $a << u$. Yet, a minimal number of recalibration actions do not necessarily result in optimal performance. We simplified our analysis by keeping the design of the technique as similar as possible to the design of Hybrid. We used the input configuration shown in Table 2 and set CDU_{low} to 1 pixel/unit. Finally, we empirically selected CDU_{high} to 20 pixels/unit for the low input precision (61 units) and to 8 pixels/unit for the high input precision (241 pixels).

5.5 Design and Procedure

A full-factorial repeated-measures within-participants design was used. The presentation order of techniques (Rate, Hybrid, and Position) and input precision (21 and 241 units) within each technique were counterbalanced among participants. For each combination of technique and input precision, participants completed three blocks of trials. We tested three target distances *D* (250, 500, and 1000 pixels) and three target widths (8, 16, and 32 pixels). Each block consisted of nine randomly sorted trials, which correspond to the nine combinations of target distances and widths. For each technique, participants also completed a total of nine practice trials. To reduce fatigue, we allowed participants to take brief breaks between blocks and techniques. Experimental sessions lasted approximately 40 to 50 minutes.

5.6 Measures and Hypotheses

We measured selection time as the total time spent to complete a reciprocal pointing task divided by two. We also measured error rates. We expected that the performance of Position and Hybrid would improve as input precision increases. Given the results of Pilot 1, however, our hypothesis was that input precision would have no effect on the performance of Rate. We also predicted that Position would be less effective under the low-precision condition due to the recalibration problem. We expected that Hybrid would provide the best balance between pointing precision and speed.

5.7 Results

We eliminated a total of 25 outliers (1.2% of total measurements) for values three standard deviations away from the within-cell mean. The main results are presented in Fig. 10. An ANOVA repeated-measures analysis indicates a significant main effect of input precision on time ($F_{1,11}=91.23$, $p=1.2\cdot10^{-6}$). This effect, however, is different for each technique. More specifically, there is a significant interaction between input precision and technique ($F_{1.74,19.16}=37.42$, $p=4.8\cdot10^{-7}$). Pairwise comparisons with

Bonferroni's adjustment show that its effect is significant for Hybrid ($p=2.6\cdot10^{-5}$), significant for Position ($p=6.0\cdot10^{-7}$), but no significant for Rate ($p=0.876$). Consistent with our findings from Pilot 1, the results suggest that input resolution has minimal or practically no effect on the performance of pure rate control as long as an optimal transfer function is selected.

The effect of the technique on time is significant ($F_{1.87,20.55}=7.39$, $p=0.0044$). Yet, pairwise comparisons show that differences between the three techniques are only significant under low precision. Position is significantly slower than both Rate ($p=2.2\cdot10^{-4}$) and Hybrid ($p=0.013$), but no significant difference is found between the two latter techniques ($p=0.11$). Although rate control performs well in terms of pointing time, it results in high error rates. The error rate is approximately 5% for this technique, in contrast to a 1.1 - 1.7% error rate observed for Hybrid and Position. Freedman's non-parametric test shows that the effect of the technique on errors is statistically significant ($\chi^2_{df=2,N=12}=9.14$, $p=0.01$). Overall, our results indicate an advantage of Hybrid over Position for 61 input units and an advantage of both these techniques over Rate for 241 input units.

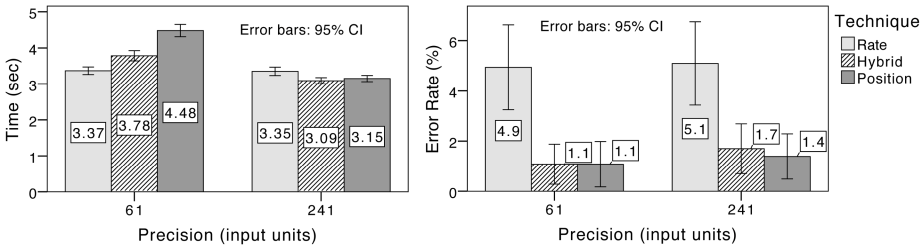


Fig. 10. Overall results of the 3rd experimental study

Typical error rates in Fitts' law experiments are in the range of 3 - 4%. The error rates that we observe for Hybrid and Position are considerably lower, even though the instructions given to the participants were consistent all over the experimental session and rather neutral towards precision. Our explanation is that the two window-based techniques reduce the likelihood of errors by dividing the pointing task into two sequential subtasks. Rapid ballistic movement, which is rough and imprecise, takes place during the first subtask. Errors are possible, but the user must correct them before continuing with the final pointing movement. As the participants were given visual feedback about the success of the first movement in the form of an area cursor, errors during this step were eliminated. In addition, the second subtask was relatively slow and precise, as it took place in a limited area around the target.

As expected, the effect on pointing time is significant for both target width ($F_{1.73,19.03}=236.77$, $p=9.1\cdot10^{-14}$) and distance ($F_{1.90,20.84}=460.99$, $p=8.3\cdot10^{-18}$). We also find a significant interaction effect between technique and distance ($F_{3.62,39.87}=18.04$, $p=3.5\cdot10^{-8}$). As shown in Fig. 11, the performance of Position degrades considerably in long distances, as the cost of recalibration actions becomes higher. A pairwise analysis using Bonferroni's adjustment shows that Position is significantly slower than both Rate ($p=2.7\cdot10^{-4}$) and Hybrid ($p=4.4\cdot10^{-4}$) for distances of 1000 pixels and significantly slower than Rate ($p=0.006$) for distances of 250 pixels.

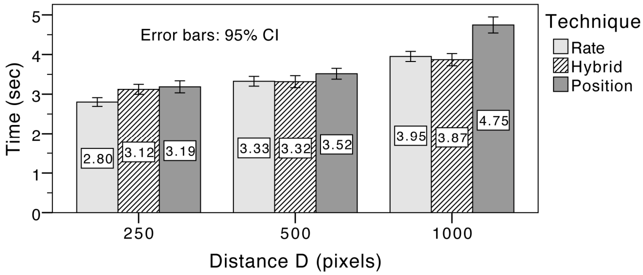


Fig. 11. Pointing time across distances

5.8 Predictive Models

We can attempt to predict user performance for other levels of input precision. For Rate, we use the regular formulation of Fitts' law (see Eq. 9 in Fig. 12), which is independent of the number of input units. For Hybrid and Position, we use Eq. 10. The parameter s ($0 \leq s \leq 1$) represents the portion of the input range dedicated to the movement in the central zone. S is the size of the movable window, which can be calculated as $S = CDU_{gain} \cdot s \cdot u$. To produce Eq. 10, we followed an analysis similar to that of Cao et al. [4]. Due to limited space, we do not provide the details here.

	$MT = a + b \cdot \log_2\left(\frac{D}{W}+1\right)$			Eq. 9
Rate	***a***	***b***	R^2	***Std. Err.***
	.189	.631*	.987	.091

	$MT = a + b_0 \frac{s}{D} + b_1 \log_2\left(\frac{D}{S}+1\right) + b_2 \log_2\left(\frac{D}{W}+1\right)$					Eq. 10
	a	b_0	b_1	b_2	R^2	***Std. Err.***
Hybrid	-1.332*	580.7*	.373*	.567*	.938	.190
Position	-2.278*	562.7*	.833*	.548*	.913	.361

*Coefficients significantly different than zero (p<0.05)

Fig. 12. Models tested for the three techniques

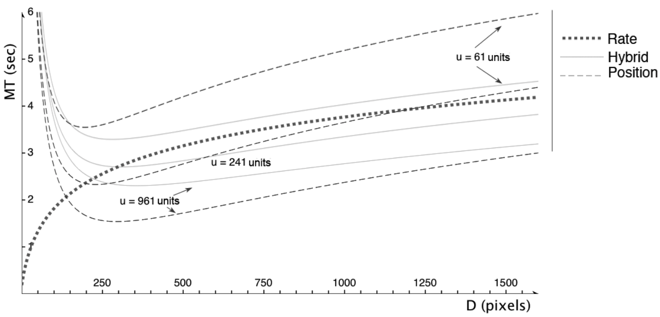


Fig. 13. Extrapolated movement times for the three pointing techniques and three levels of input precision (*u = 61, 241,* and *961 units*). Target width is constant (*W = 20 pixels*).

Based on the above models, we have extrapolated movement time with up to 961 input units and distances as high as 1500 pixels. Fig. 13 gives an overview of the trends. Position exhibits a poor performance under low input precision, but outperforms Rate and Hybrid as precision improves. The performance of Position is particularly sensitive to both input precision and target distances. Variations in the performance of Hybrid are less radical. The technique seems to be more appropriate in the range of 61 of 241 units. However, it can be a good alternative for higher levels of precision, particularly when pointing involves large distances. Clearly, prediction results for the position-control and the hybrid technique do not apply to short target distances. As long as movement takes place only within this window, movement will be purely position controlled and faster than the time predicted by the above models.

6 Limitations and Future Directions

We examined levels of input precision from 21 to 241 units and targets distances from 250 to 1000 pixels. Output configurations can vary greatly beyond the range of values reported here. For example, wall displays can support very high resolutions. In such environments, pointing precisely and moving quickly in space are equally important, so future work must verify how our techniques behave in such settings.

Some previous work [10, 14, 23] has made use of pointer acceleration to effectively balance between precision and speed. To simplify our analysis, we did not consider pointer acceleration here. If used with our techniques, acceleration can be applied at two levels: (1) to the window of precise pointing, and (2) to the cursor within this window. The former solution could increase the range of the window's movement and hence reduce the number of recalibration actions. The latter solution, however, is problematic, as it can cause the de-calibration of the wrist movements away of the joint's neutral position.

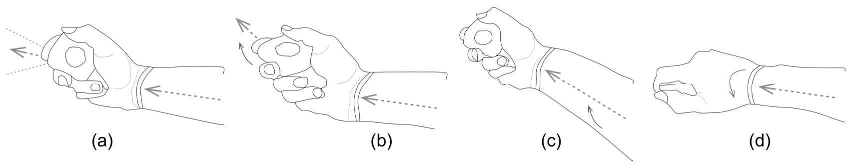


Fig. 14. Interaction mixing movements of various joints. A bracelet attached around the wrist allows for detecting its absolute position and measuring the rotations of the hand relative to the forearm. (a) Slight wrist rotations within a central zone. (b) Wrist rotations out of the central zone. (c) Movements of the whole arm. (d) Rotations of the forearm.

To extend our approach to two-dimensional tasks, we must first consider that hand movements are not perfectly symmetric. Wrist flexions and extensions have a wider angular range than ulnar and radial deviations (Fig. 2). Techniques must be adapted to accommodate these asymmetries. In addition, natural hand movement combines rotations of the wrist and movements of other joints (Fig. 14), which could be mapped to different levels of a pointing task, e.g., by using different CD gains. Exploring techniques that combine various joints is an interesting future direction. Another future goal is to test non-linear mappings between input and output [21].

7 Conclusions

The paper explored pointing with low-precision input when the input device is controlled by movements of the wrist. We examined techniques that achieve pointing precision without making use of explicit mode-switching mechanisms. Wrist movements are physically constrained and take place within a small range around the hand's resting position. Based on this observation, we explored rate control as a possible solution. Our results show that rate control is effective under low levels of input precision. Interestingly, we found that as long as optimal transfer functions are selected, the performance of rate control remains practically constant for a range of input precision from 21 to 241 units. This suggests that rate control could be particularly valuable for low-precision input devices (e.g., low-fidelity camera-based capture of hand movements) and users with hand-tremor problems.

In addition to rate control, we examined two techniques that split the pointing task into two scales. Pointing at a macro scale is performed with an area cursor, which is visually communicated to the user as a framed window. Pointing at a micro scale is performed with a regular point cursor, which moves within the window. The first technique combines position and rate-control and derives from previous work [7, 9]. The second technique uses position control to move both the area and the point cursor and has been based on previous work on two-scale pointing [12] and early techniques in Virtual Reality that applied non-uniform position-control mappings [20]. According to our experimental results, these two techniques are more precise than pure rate control and their performance improves as input precision increases. The hybrid technique was particularly effective in balancing between precision and speed in the range of 61 to 241-input units. It outperformed the non-uniform position-control technique in long target distances over 1000 pixels. Nevertheless, our analysis suggests that the performance of the latter is faster as input precision improves. Based on these results, we derived predictive models of user performance, but future work must compare the techniques on different task scales. We are particularly interested in exploring their application to high-resolution displays where distances can be as high as 10k to 20k pixels. We also plan to study the proposed techniques on the movement of other human joints. For example, the thumb moves within a limited range. Enabling it to precisely point on large output surfaces in isolation of or in combination with hand movements is a challenging problem and certainly worth of future investigation.

References

1. Appert, C., Chapuis, O., Pietriga, E.: High-Precision Magnification Lenses. In: Proc. ACM CHI, pp. 273–282 (2010)
2. Bérard, F., Wang, G., Cooperstock, J.R.: On the Limits of the Human Motor Control Precision: The Search for a Device's Human Resolution. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 107–122. Springer, Heidelberg (2011)
3. Bowman, D.A., Hodges, L.F.: An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: Proc. Symposium on Interactive 3D Graphics, pp. 35–38 (1997)
4. Cao, X., Li, J.J., Balakrishnan, R.: Peephole Pointing: Modeling Acquisition of Dynamically Revealed Targets. In: Proc. ACM CHI, pp. 1699–1708 (2008)

5. Casiez, G., Vogel, D.: The Effect of Spring Stiffness and Control Gain with an Elastic Rate Control Pointing Device. In: Proc. ACM CHI, pp. 1709–1718 (2008)
6. Casiez, G., Vogel, D., Balakrishnan, R., Cockburn, A.: The Impact of Control-Display Gain on User Performance in Pointing Tasks. In: Human-Computer Interaction, 2008, vol. 23(3), pp. 215–250 (2008)
7. Casiez, G., Vogel, D., Pan, Q., Chaillou, C.: RubberEdge: Reducing Clutching by Combining Position and Rate Control with Elastic Feedback. In: Proc. ACM CHI, pp. 129–138 (2007)
8. Chapuis, O., Dragicevic, P.: Effects of Motor Scale, Visual Scale and Quantization on Small Target Acquisition Difficulty. In: ACM ToCHI, 2011, vol. 18(3), pp. 13:1–13:32 (2011)
9. Dominjon, L., Lécuyer, A., Burkhardt, J.-M., Richir, S.: A Comparison of Three Techniques to Interact in Large Virtual Environments Using Haptic Devices with Limited Workspace. In: Nishita, T., Peng, Q., Seidel, H.-P. (eds.) CGI 2006. LNCS, vol. 4035, pp. 288–299. Springer, Heidelberg (2006)
10. Frees, S., Kessler, G.D., Kay, E.: Prism Interaction for Enhancing Control in Immersive Virtual Environments. In: ACM TOCHI, vol. 14(1) (2007)
11. Gibbs, C.B.: Controller Design: Interactions of Controlling Limbs, Time-Lags and Gain in Positional and Velocity Systems. Ergonomics 5(2), 385–402 (1962)
12. Guiard, Y., Beaudouin-Lafon, M., Mottet, D.: Navigation as Multiscale Pointing: Extending Fitts' Model to Very High Precision Tasks. In: Proc. ACM CHI, pp. 450–457 (1999)
13. Hinckley, K., Pausch, R., Goblel, J.C., Kassell, N.: A Survey of Design Issues in Spatial Input. In: Proc. ACM UIST, pp. 213–222 (1994)
14. König, W.A., Gerken, J., Dierdorf, S., Reiterer, H.: Adaptive Pointing - Design and Evaluation of a Precision Enhancing Technique for Absolute Pointing Devices. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 658–671. Springer, Heidelberg (2009)
15. Lehman, S.L., Calhoun, B.M.: An Identified Model for Human Wrist Movements. Experimental Brain Research 81(1), 199–208 (1990)
16. Lemay, M.A., Crago, P.E.: A Dynamic Model for Simulating Movements of the Elbow, Forearm, and Wrist. Journal of Biomechanics 29(10), 1319–1330 (1996)
17. Maury, S., Athènes, S., Chatty, S.: Rhythmic Menus: Toward Interaction Based on Rhythm. In: Proc. ACM CHI, Extended Absracts, pp. 254–255 (1999)
18. Myers, B.A., Bhatnagar, R., Nichols, J., Peck, C.H., Kong, D., Miller, R., Long, A.C.: Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. In: Proc. ACM CHI, pp. 33–40 (2002)
19. Oh, J.-Y., Stuerzlinger, W.: Laser Pointers as Collaborative Pointing Devices. In: Proc. Graphics Interface, pp. 141–149 (2002)
20. Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa, T.: The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In: Proc. ACM UIST, pp. 79–80 (1996)
21. Rahman, M., Gustafson, S., Irani, P., Subramanian, S.: Tilt Techniques: Investigating the Dexterity of Wrist-Based Input. In: Proc. ACM CHI, pp. 1943–1952 (2009)
22. Raynal, M., Gauffre, G., Bach, C., Schmitt, B., Dubois, E.: Tactile Camera Vs. Tangible Camera: Taking Advantage of Small Physical Artefacts to Navigate into Large Data Collections. In: Proc. NordiCHI, pp. 373–382. ACM Press (2010)
23. Vogel, D., Balakrishnan, R.: Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays. In: Proc. ACM CHI, pp. 33–42 (2005)
24. Zhai, S.: Human Performance in Six Degree of Freedom Input Control, Ph.D. Dissertation, University of Toronto (1995)

Wands Are Magic: A Comparison of Devices Used in 3D Pointing Interfaces

Martin Henschke, Tom Gedeon, Richard Jones, Sabrina Caldwell, and Dingyun Zhu

College of Engineering and Computer Science,
Australian National University, Acton ACT, Australia
{martin.henschke,tom.gedeon,richard.jones,sabrina.caldwell,
dingyun.zhu}@anu.edu.au

Abstract. In our pilot study with 12 participants, we compared three interfaces, 3D mouse, glove and wand in a 3D naturalistic environment. The latter two were controlled by the same absolute pointing method and so are essentially identical except for the selection mechanism, grasp action versus button. We found that the mouse performed worst in terms of both time and errors which is reasonable for a relative pointing device in an absolute pointing setting, with the wand both outperforming and favored by users to the glove. We conclude that the presence of a held object in a pointing interface changes the user's perception of the system and magically leads to a different experience.

Keywords: "magic wand", "3D mouse", "hand gesture", "fatigue", "user satisfaction".

1 Introduction

Tasks that are typically accomplished by human beings can be separated into two broad categories: tasks achieved through physical manipulation of objects, such as chopping wood with an axe or writing with a pen and paper, and communication with other people, commonly done both using words and a series of communicative gestures. When not operating physical tools, gestures are usually a form of communication and often intertwined with spoken language[3]. Although operating computers can be considered a form of communication, typical day-to-day interactions with a standard desktop or laptop computer fall into the physical object category. Interfaces that use motion or body capture allow users to perform these sorts of tasks without requiring a device to be held, with interfaces implementing gestures or methods of interaction found in both categories. The challenge of making these interfaces sufficiently ubiquitous remains formidable, and in the meantime it remains unclear if such interfaces are superior or preferred by users to an object-based interface, in a comparable setting. The ergonomic factors relating to these devices, namely how much physical strain or fatigue they cause with continued use, remain an issue in interface design[9], and merit further study.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 512–519, 2013.

Our pilot study used a series of pointing interaction methods in a 3D interface, comparing operating an interface while holding a physical object to performing interactions in the absence of one. We wish to determine if there are any changes in the mode of operation, the severity and locality of discomfort caused by their operation and the preferences of the users.

2 Background

The MIT Media Room[1] is a seminal example of an object-free interface, in which users would use a combination of spoken language and pointing gestures at a large projected display to create and modify the state of various shapes in a scene. A similar system, using a 3D spatial environment and gesture interface captured interactions with a specially designed glove and camera system[8]. Kjeldsen and Kender (1996) presented a camera-based pointing system applied directly to performing mouse tasks on a standard desktop computer[7]. Another similar system provided a means of absolute pointing through capture of the user's entire arm and projecting a line from the user's shoulder to their fingertip[5]. The system used multiple cameras to make any form of device such as a glove or sensor unnecessary.

Research into the ergonomics or feasibility of such systems remains limited, though study into developing ergonomic gestures has been conducted[9]. An assessment of performance of gesture interfaces in VR environments reported the interfaces were almost four times slower than using a traditional mouse system, and fatigue was a common complaint[2]. A study comparing Wii remote and Kinect interfaces found a hands free interface performed tasks faster and was preferred by users to the Wii interface, which used in-built accelerometers rather than arm movements for tracking[4]. Another study, looking specifically at the ergonomic considerations of interfaces with and without physical objects indicated that performing a task with a 'virtual object' was more difficult and fatigue-inducing than an equivalent one performed with a real object. The trial found users extended their fingers earlier and farther with the virtual object, making it the more fatiguing of the two interactions[6].

3 The System

We have developed a 3D pointing interface, in which users are able to select, grab and drag objects in a virtual space by pointing and performing interactions with the given device. The interface makes use of an absolute pointing system, in which both the user's position in the work area and the screen's position and dimensions are used to determine where user is pointing. The system works by having the user physically point at the display surface with their hand. From here, capturing user input involves detecting the position of the user's forearm and hand to construct a 3D vector, which is then projected forward to the 2D plane on which both the Kinect and the display surface rest. This provides a 2D point on the plane where the user is pointing at any given point in time. The system allows for a relatively flexible definition of pointing, with different arm orientations and holding the device in different ways possible.

Our pilot makes use of a total of three different interaction methods. The first is a wired glove interface as an approximation of operating a user interface without physically manipulating a device. It operates using an Essential Realities P5 Wired Glove, weighing 82g for tracing finger movements and a Microsoft Kinect that tracks general body movements. The second is a Wand interface, operating with Kinect in the same way as the Wired Glove interface, but with a hand-held controller in place of a glove. The controller is from a Nintendo Wii, and is identical to a standard Wii Remote controller, except that it has a slimmer, longer and more cylindrical profile, weighing 78g. The controller is only used for button presses as position information is acquired from the Kinect. An OnmiMotion Air Mouse, weighing 65g, is used for our third device, using a relative pointing system in contrast to the other two devices. This gyroscopic mouse is similar to a standard desktop mouse in appearance, but moves the cursor through rotating the body of the mouse rather than laser tracking.

4 Experiment

The general research question of this paper is *does the absence of a physical object of manipulation impact how the user interacts with an interface?* We define a physical object in this context as one that is held rather than worn, and is manipulated in-hand. We are looking at three particular aspects of this question: user performance, induced fatigue and preference.

A 3D visualization was conducted for the experimental trials, designed to have the appearance of being performed in a Gothic château, with the tasks themselves resembling ones being accomplished by casting magic spells. This allowed users to treat each device as operating according to their own expectations, and adopt an approach that felt contextually natural in operating the devices. A total of three separate kinds of tasks were performed:

- A selection task, in which the user was required to select a ghost on the screen by pointing at it and performing a selection. For the glove, this was achieved by quickly tapping the index finger down and back up again, and for the wand this was done by pressing the 'A' button on the shaft of the device. For the mouse, the left mouse button was used. The position and size of the targets was randomized. Nine selections were performed per trial.
- A select-drag task, in which the user was required to select a key, randomly placed near the bottom of the display, then 'grab' the object and drag it over a lock. In the case of the glove, the grab was performed by forming a fist, and in the case of the wand it was by pulling a trigger beneath the device. Six drag tasks were performed per trial.
- A select-drag-drop task, in which the user was required to select a firefly and drag it to a cage, then release it. The target area remained the same for each task, but the positions of the selectable objects were randomized. Six of these tasks were performed per trial.

Fig. 1. Selecting a firefly in the select-drag-drop trial using the P5 Wired Glove

The trials presented the tasks consecutively with a brief break between each task explaining what needed to be done before it was begun. At the end of each sequence of tasks, the participant was given a 5-10 minute break to rest their arm before continuing with the next interface device.

During the task completion phase of the experiments, the mouse cursor is hidden from participants, as our trial is designed for a naturalistic environment without relative location cues, and could be used in a natural environment. Selection gestures are accompanied with a small burst of stars at the area pointed to, so users know where their gesture was directed.

Over the course of the trials, the system keeps track of every button press and gesture, the position of the pointing location and a time signature. Video from the Kinect was captured, to observe the movements each user makes when operating each interface. At the end of each trial, the user was also queried on the discomfort felt in their arms, and asked where specifically the discomfort, if any, was located and how severe it was on a scale of 0 to 10. At the end of each trial, the user was also asked to fill out a short questionnaire about the device they had just used, using a combination of Likert-scale questions and short written answers.

4.1 Results

The trial was run with 12 participants, 8 male, and 4 female, between the ages of 21 and 36, with a mean age of 25. All participants were computer science students, and reported regular computer use. Of those, 9 reported familiarity with the Kinect. Other gesture capture technologies such as the Nintendo Wii were also familiar to participants, with 1 participant having never used any motion capturing devices.

The relationship between the number of incorrect selections made and the time taken to complete the trial demonstrates a moderate to strong trend for most users, so these values are used as a measure of how accurately and easily each participant was able to use each device.

Table 1. Mean and Standard Deviations for time and errors in completing selection tasks for each device (time in seconds)

	Glove		*Mouse*		*Wand*	
	Time	Errors	Time	Errors	Time	Errors
μ	136.8s	50.3	149.0s	71.5	79.7s	45.8
σ	62.1	18.4	81.3	33.8	33.8	20.7

Performance between the glove interface and wand were roughly equivalent in terms of accuracy, but users on average took substantially longer in performing operations with the glove. The mouse was on average the poorest performing device, but also the device with the greatest level of variance. The wand can be seen here as having generally the best and most consistent performance, though all devices are demonstrated as being able to perform quickly and with relatively few errors.

During the trial, a variety of arm positions were adopted. The method of control and orientation of the arm was consistent enough between participants to be separated into three categories, ordered from lowest induced fatigue to highest:

- The shoulder at rest, with the elbow kept at the side and the forearm extended pointing at the display. Holding the device in this manner, users typically made wrist and small forearm movements to control the interface.
- The shoulder partially at rest, the elbow out from the side and bent, with the forearm raised and the wrist bent to face the device to the display. Forearm rotation and wrist movements were the primary method of interaction.
- The arm fully extended with the elbow locked. The user moves the entire arm at the shoulder in this position.

While users typically kept a consistent arm orientation for the trials, it was not uncommon for them to change positions. This was most commonly moving from an outstretched arm to a relaxed or partially relaxed arm, likely to combat fatigue, or from a relaxed or partially relaxed arm to an outstretched arm, usually due to issues with accuracy. Figure 3 shows the prevalence of various orientations with each device, by counting the number of times and the duration of the trial during which the stance appeared for all participants.

On average, the glove interface was reported to be the most fatiguing by participants (μ = 4, σ = 1.9), followed by the wand (μ = 2.4, σ = 2.2) with the mouse being the least stressful to use (μ = 1.3, σ = 2.5).

Fig. 2. From left to right: shoulder at rest, partially extended forearm, fully extended arm

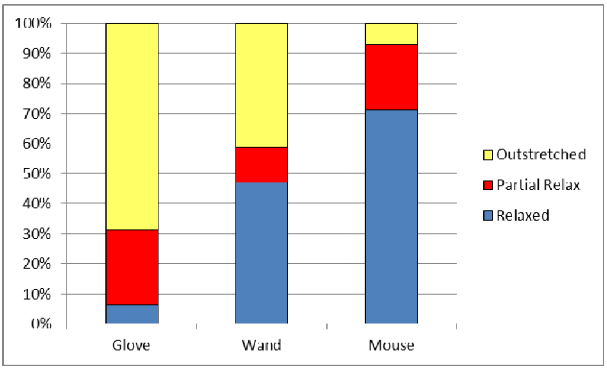


Fig. 3. Prevalence of different arm positions for all participants using the interface

Immediate fatigue induced from use, and the extent of continued fatigue from use for each device was also reported in the questionnaire. The results are shown in Table 2, with no significant difference between the wand and mouse ($p > 0.05$), but more pronounced variation between the glove and all other devices. This confirms the glove as being the most fatigue-inducing device to operate.

Table 2. Distribution-free analysis of questionnaire results pertaining to immediate and continually induced fatige from each device

	Immediate Fatigue	*Continuous Fatigue*
Non-parametric test	χ^2 =7.023, p=0.03	χ^2 =11.286, p=0.004
Pair-wise Glove vs. Wand	p = 0.008	p = 0.015
Pair-wise Glove vs. Mouse	p = 0.04	p = 0.005
Pair-wise Wand vs. Mouse	p = 0.55	p = 0.232

In the questionnaire, users were asked to grade each interface system on a seven-level Likert scale, how natural, intuitive, learnable, reactive, accurate and generally easy to use the interface was perceived to be. Of those, only two yielded a statistically significant result in non-parametric analysis. In both instances, pair-wise analysis revealed no significant difference between the glove and mouse. The wand however was found to be more intuitive than the glove (p=0.016) and the mouse (p=0.018). The wand was also found to be easier and quicker to learn than the glove (p=0.024). The fact that all other results lacked a clear correlation suggested personal preference played a large part in these reports.

Users were also asked to rank each interface in terms of preference, and provide comments as to those rankings. The results of this are shown in Figure 3, indicating a general preference for the wand interface. Comments justifying this preference referenced the wands appearance or comfortable grip or being easier to press buttons than on the mouse or with finger gestures with the glove. When the mouse was preferred over the wand, it was typically due to being the less strenuous to operate, with complaints of the mouse being its unpredictable behavior and difficulty in reaching buttons. Preferences for the glove were reported as being the more natural of the interfaces, while complaints included a poor responsiveness and fatiguing to use.

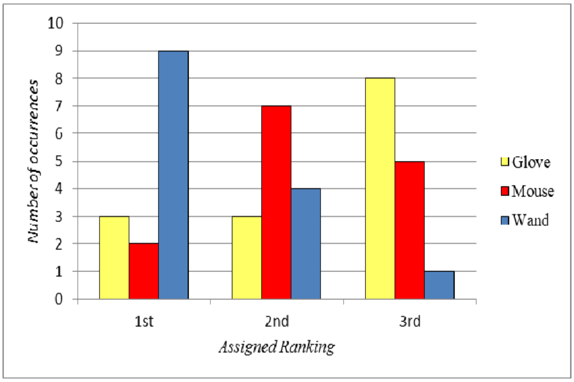


Fig. 4. Sum of rankings given by each user to the three interface devices

5 Discussion and Future Work

Our expectations were that users would find differences in induced fatigue and preference between the devices, but pronounced differences in performance and speed were not expected. The results suggest the presence of an object held in the user's hand has a profound impact on how users view the system, and the amount of effort users expended.

The results from observations suggest that interfaces with a physical object held in hand tended to encourage users to use a less direct form of pointing, performing pointing with the upper arm fully or partially at rest, and the elbow at the side, while having no object encouraged users to keep their arms fully extended. This fits the natural analogue of attempting to point to an object with your hand, in which the user raises it and looks down their arm towards their index finger, where the wand was interpreted much more generally and had more varied usage patterns. A generally larger number of changes in posture were seen with wand users; they would find configurations they liked rather than what they felt was necessary. The wand was observed to also be subject to in-hand manipulation; users were observed turning or rotating the device within the palm of their hand rather than performing the equivalent rotation with their wrist or forearm. This rather counter-intuitively seems to be accompanied with a better performance index for the wand. Less physical movement by the user is required overall to move the pointer to the desired area, but as the region of selection is far smaller, inaccuracy was expected to be higher. This may be attributed to the fact that selection and grabbing was reported to be much easier with the wand and mouse when compared to the glove. Users on average seemed to exert more energy attempting to perform selections and grabs, as the glove required a strong, deliberate motion to select an object on screen compared to the other two. That a strong correlation can be seen between increasingly extended arm, trial length and fatigue matches expectations, and remains an important consideration in interface design.

The poor results for the mouse can largely be attributed to the absence of a visible cursor and the use of rotation as opposed to lateral motion as with a desktop mouse. Overshooting targets and quickly clicking to inform the user where the cursor was occurred regularly in the trials.

The variation in user preference may reflect the background of participants. Comments such as ‘familiarity with the Wii may have been useful’ appeared in the questionnaire, despite the fact the system functions quite differently to the Wii, but the method in which users operated the interface may have reflected that experience. Preferences for any given device appears to be heavily influenced by those users found to perform most effectively, though there were some instances where this was supplanted by familiarity with that device, particularly with the mouse.

This trial has indicated substantial nuances in the problem that merit further investigation. Controlling and examining each of the findings in this paper in more detail will be the focus of later experiments. Running trials for extended periods of time and encompassing more complex interactions will be performed in future work.

References

1. Bolt, R.A.: “Put-that-there”: Voice and gesture at the graphics interface. SIGGRAPH Comput. Graph. 14, 262–270 (1980)
2. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the usability of gesture interfaces in virtual reality environments. In: Proc. of the 2005 Latin American Conference on Human-Computer Interaction, pp. 100–108. ACM, Cuernavaca (2005)
3. Cassell, J.: A Framework For Gesture Generation and Interpretation. Computer Vision and Machine Interaction, pp. 191–125 (1998)
4. Francese, R., Passero, I., Tortora, G.: Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In: Proc. of the International Working Conference on Advanced Visual Interfaces, pp. 116–123. ACM, Capri (2012)
5. Fukumoto, M., Suenaga, Y., Mase, K.: “Finger-Pointer”: Pointing interface by image processing. Computers & Graphics 18, 633–642 (1994)
6. Kim, Y., Lee, G.A., Jo, D., Yang, U., Kim, G., Park, J.: Analysis on virtual interaction-induced fatigue and difficulty in manipulation for interactive 3D gaming console. In: Consumer Electronics (ICCE), pp. 269–270 (2011)
7. Kjeldsen, R., Kender, J.: Toward the use of gesture in traditional user interfaces. In: Automatic Face and Gesture Recognition, pp. 151–156 (1996)
8. Maggioni, C.: A novel gestural input device for virtual reality. In: Virtual Reality Annual International Symposium, pp. 118–124 (1993)
9. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 409–420. Springer, Heidelberg (2004)

Assisting the Driver with Distance Estimation: Usability Evaluation of Graphical Presentation Alternatives for Local Traffic Events

Angela Mahr, Sandro Castronovo, Rafael Math, and Christian Müller

German Research Center for Artificial Intelligence,
Intelligent User Interfaces Department, Campus D 3_2,
66123 Saarbrücken, Germany
{angela.mahr,sandro.castronovo,rafael.math,
christian.mueller}@dfki.de

Abstract. When integrating numerous in-car information and assistance systems, a consistent way of spatial distance presentation for drivers is required. A common practice is to use discrete textual information (e.g. 500 meters) in combination with a graphical bar representing relative spatial information. Hitherto there exists no design consistency with respect to bars for distance illustration. Contemporary solutions differ in terms of movement direction (upward vs. downward), composition type (decreasing vs. increasing), or alignment (horizontal vs. vertical). We conducted a driving simulator experiment to investigate user preferences, perceived location, and eye gaze data for a meaningful subset of bars in a dynamic scenario. When approaching a traffic event (road works), one out of four vertical bar alternatives indicated the current distance. Subsequently, the associated horizontal bar type (decreasing or increasing) visualized the driver's spatial progress within the road works section. Our results indicate, that drivers prefer upwards-moving approach bars and progress bars decreasing from left to right. Eye-tracking data supports usage of decreasing bars instead of increasing bars. Accordingly, we elaborated an initial version of design guidelines for bars representing relative spatial information for local events. On this basis we implemented approach and progress indicators, which were adopted for numerous use-cases in a large field operational test for Vehicle-2-X Communication.

Keywords: Distance, Assistance, Design, Usability Evaluation, In-car, Driving.

1 Introduction

During the last decades, awareness regarding intuitive driver interfaces has constantly increased. With a rising number of in-vehicle information systems (IVIS) and advanced driver assistant systems (ADAS) being integrated, coherent presentation of information is crucial for drivers to understand contents immediately. Although multimodal presentations open new options, common graphical displays will not become dispensable for automotive applications. Due to visual driver distraction while

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 520–537, 2013.

glancing at the display, it is necessary to reduce glancing times to a minimum by intuitive presentations. This is one fundamental design principle for interaction with IVIS [1].

One of the frequently required functionalities of in-car displays is spatial distance presentation, e.g. in today's navigation systems when a maneuver point comes up (e.g. turn right in 300m). In cases of events that can be pinpointed to a location, we use the term spatial *approach* visualization. Additionally, modern ADAS require visual presentation of spatial validity, e.g. visualizing the remaining section where a traffic sign is valid. In this case of traffic events having an areal extend, we name it spatial *progress* visualization. Although in general HCI progress is well researched [2–7], the automotive domain lacks both a common concept, as well as design guidelines for implementations of spatial approach and progress. We found a general tendency of numerical distance presentation (e.g. 300 meters) in combination with a graphical progress bar representing relative spatial information when investigating hitherto existing distance presentation in diverse system types. For presentation of spatial information, there exist many plausible variants, and unfortunately no design consensus could be observed. Nevertheless, three major design factors could be derived: First, there exist two general orientations of the bar (horizontal vs. vertical), second, the bar can be filled or emptied with some kind of filling (increase vs. decrease), and third, the movement direction can be manipulated (up vs. down or left vs. right respectively). Moreover, color-coding is sometimes used for urgency emphasis.

In this paper we examine these different design factors with the help of a user study. Our goal was to find parameter combinations, which are preferred by drivers, perceived fastest, and minimize glancing times. Accordingly, a suitable driving task was designed and implemented via the openDS driving simulator [8] and an eye tracker was used to measure glance durations while driving. Based on our results, we suggest general design guidelines for spatial approach and progress in the automotive domain. These guidelines have already been applied to a Human-Machine Interface used in a large field operational test with over 20 novel ADAS (sim^{TD}).

2 The sim^{TD} HMI

The sim^{TD} project[1] comprises one of the largest field operational tests for Vehicle-2-X Communication Networks (V2X) and its applications. In this domain, cars communicate with other entities using different wireless communication technologies where `X` denotes the communication partner. These are typically other cars, road infrastructure like signal lights, or traffic management authorities. sim^{TD} investigates established V2X use-cases in real traffic scenarios around the Hessian city of Frankfurt/Main, Germany with a fleet of 120 cars. One of the goals of the project is to evaluate the effect of V2X-based applications on traffic and driver. It is currently the largest field operational test in this domain. Drivers are not only advised "experts" to the sim^{TD} system but also novices. Especially for this type of drivers it is of utmost

[1] http://www.simtd.de/

importance that the Human-Machine Interface (HMI) of the system is intuitive, non-distractive and features a consistent integration of all use-cases. It consists of a main area where the most relevant information is displayed and an additional area in the upper part containing six slots for additional information of parallel applications. Using two virtual buttons, the driver can also access a map and an option screen (see Figure 1). The Automotive group at "double-blinded" is responsible for developing and evaluating the HMI concept in sim^{TD} prior to the field operational test.

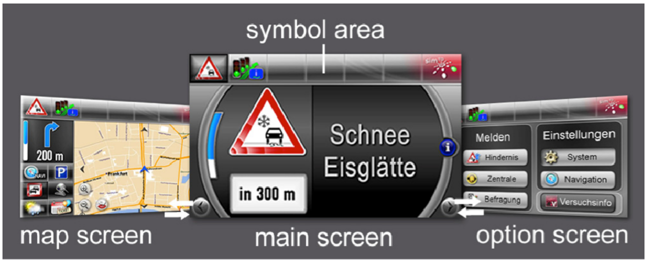


Fig. 1. The sim^{TD} HMI depicting three virtual screens and symbol area

The V2X use-cases evaluated can be classified into three categories: Safety, traffic efficiency, and value added services. Many events belonging to these use-cases are location-based, i.e. they pinpoint to a certain geographical position. Moreover, some events are valid for a certain road segment, e.g. road works. Table 1 lists a shortened selection of use-cases, which are investigated in sim^{TD}. Altogether, there are over 25 use-cases tested within the project. We refer the interested reader to the full list of selected functions in [9]. Especially warning functions, e.g. congestion warning, need to be displayed way before the actual event. Those approaches to upcoming events should be presented to drivers in a consistent way for all use-cases. Besides text-only information, the use of graphical information via progress bars can be found in several navigation systems. We analyze different realizations and have a look what can be learned from general HCI research in the following section.

Table 1. Some selected sim^{TD} use-cases involving approach and/or progress information

Category	Use-case
Safety	Lost Cargo Warning
Safety	Congestion Ahead Warning
Safety	Animals on the Road Warning
Safety	Approaching Emergency Vehicle Warning
Safety	Road Weather Warning
Efficiency	Traffic Sign Assistant
Efficiency	Road Works Information System

3 Background

Since the beginning of HCI research, design and display of progress is considered important and useful. For example, Myers in [7] concludes that especially for long-lasting tasks users prefer this kind of progress display. In traditional HCI, the design of progress was established even before graphical user interfaces were common. It was implemented using ASCII characters (cf. Figure 2a).

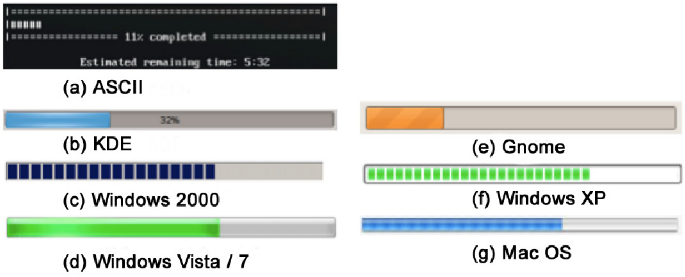
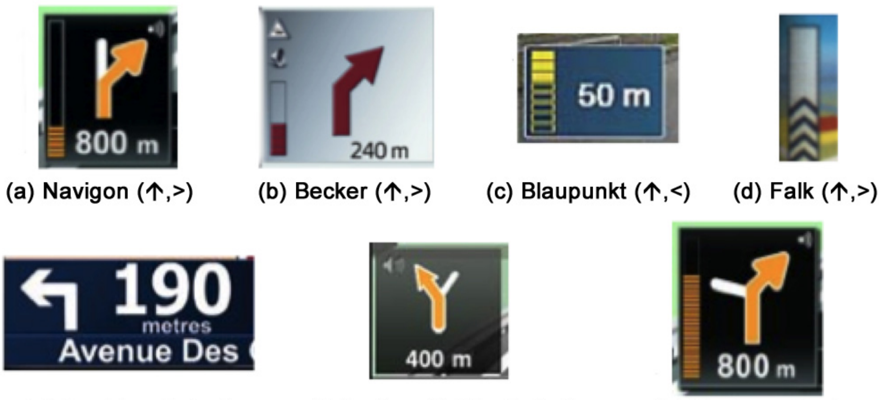


Fig. 2. Design of progress in different Operating Systems (b)-(g) and in ASCII code (a)

Nowadays, horizontal filling bars (from left to right) constitute a common implementation on Desktop Operating Systems and mobile devices. With Windows 2000 and Windows XP, segmented bars were invented but replaced by continuous bars in Windows Vista/7 (Figure 2, (c), (f), (d)). Modern versions of all major Desktops come with an animation of the bars' fill color either "pulsating" from left to right (Windows) or "ribbed" from right to left, e.g. in Mac, see Figure 2, (g). One might think this change was invented because of advances in technology such as increased computing power or in order to make the graphical user interface more appealing. But interestingly, according to [3], this type of bars result in statistically significant shorter perceived duration times when compared to their static counterparts. Another approach is hiding the actual progress from the user and displaying another one, as investigated in [4]. Harrison et al. applied non-linear functions for progress bars and found that users have a strong preference for speed-ups towards the end. On the other side, according to their results, users tend to tolerate slow-downs at the beginning better than towards the end. For tasks with dynamic completion, like defragmenting a hard-drive, such a non-linear progress bar is considered useful. Such a progress bar, on the other hand, is not appropriate for linear tasks. Other approaches try to exploit colors or different kinds of modalities. For example in [2], six combinations of colors are considered: blue/red for progress bar foreground color and cyan/orange/gray for background color. Furthermore, [6] encode progress into a series of vibrotactile pulses while [5] use auditory cues. To conclude, so far, researchers above all tried to reduce the subjective perception of waiting time, either by visual means, or by applying non-linear functions to progress bars.

However, these strategies should only be applied, when progress of the actual task is hidden from the user. In the automotive domain this decoupling between actual and displayed spatial information is neither given nor desired. Quite contrary, it is essential that both coincide as exactly as possible, e.g. during a spatial approach to a maneuver point. Besides that, transferred to our domain, Harrison et al. suggest in [4] to represent linear tasks by linear progress. Also, the use of animations in the automotive context might induce visual distraction.

Unlike in general HCI research, consistency with respect to design of spatial approach could not be observed between several manufacturers of navigation systems. Figure 3 depicts variants from different vendors. They vary in direction (decreasing vs. increasing) and movement (upwards vs. downwards). Even a single manufacturer employs increasing and decreasing bars in different models, see (f) and (g) in Figure 3. Falk (d) realizes segmented bars as arrows to indicate the filling direction. Several vendors provide no visual directional cue at all (NavigonMobile (f), TomTom (e)). This listing is not meant to be complete, but is supposed to reveal existing inconsistency found in different navigation systems as opposed to desktop progress bar design.



(a) Navigon (↑,>) (b) Becker (↑,>) (c) Blaupunkt (↑,<) (d) Falk (↑,>)

(e) TomTom (☒,☒) (f) NavigonMobile (☒,☒) (g) Navigon (↓,<)

Fig. 3. Design of spatial approach in different navigation systems. Legend: '↑'=Upwards; '↓'=Downwards; '>'=Increasing; '<'=Decreasing, '☒'=Feature not present.

Despite these differences, we could however observe several common criteria: All observed designs of spatial approach are realized as straight, vertical, and segmented bars. Minimum and maximum bar states could be inferred by a frame or similar means. Additional textual information (e.g. 500 meters) was presented in every design. We decided to reflect these shared criteria by setting them as preconditions in the design of experimental conditions. Accordingly, we used textual information in combination with a graphical bar representing relative spatial information in each condition. However, the observed inconsistencies in existing systems and a lack of literature on the issue in the automotive context lead to the conclusion that still many questions are unanswered when it comes to design of spatial approach and progress:

1. Do users find graphical approach and progress bars useful?
2. Which progress bar type is preferred (increasing vs. decreasing, upwards vs. downwards, use of colors, relevance of frame)?
3. Do different progress bar variants induce different visual distraction from the driving task (accessed by eye-tracking data or driving performance)?

The aim of the experiment presented here, was to answer the questions above and derive a universal guideline for designing consistent spatial approach and progress for in-vehicle information systems. It is important to notice that dynamic bars cannot be evaluated on static screenshots as it is for example impossible to infer movement direction or rate of changes. Furthermore, watching simply animated bars is not comparable with a driver (visually) engaged in a driving task using bar animations as relative spatial information for local events. On these grounds, we decided to conduct a driving simulator experiment with a mock spatial approach and progress scenario.

4 Experiment

4.1 Participants

24 participants participated in this experiment, 12 men and 12 women. Their age varied between 19 years and 59 years (mean = 37.0, SD = 15.1). All of them possessed a valid driver's license for at least two years. Half of the participants owned a navigation system and indicated to use it frequently, 2 participants indicated to use a navigation system occasionally, and 10 participants have only rarely used a navigation system before. Accordingly, no absolute novices took part. Participants were paid 10 Euros for approximately 1 hour of time.

4.2 Apparatus

The experiment setup consisted of a Logitech Momo steering wheel attached to a desk and two pedals on the floor, resembling the ergonomic condition of a driver in a car (cf. Figure 4). During the experiment, the participant's task was to steer and brake with these devices according to driving requirements. A computer flat screen (size 20 inch, resolution 1440 x 1050 pixel) presenting the driving simulation was set up on a desk with a distance of about 80 centimeters to the participant's head.

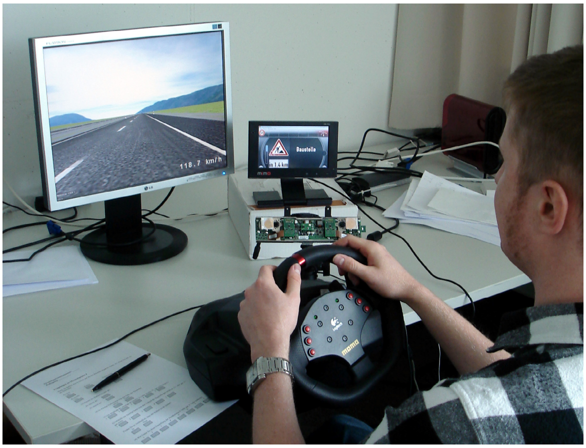


Fig. 4. Experimental Setup

White digits on the lower right side of the screen displayed the current vehicle speed. Their color turned into red, whenever drivers diverged from the currently prescribed maximum speed for more than 10 kilometers per hour. We used an open source driving simulation solution, which has recently been presented in [8]. For this experiment, several highway tracks were generated with each track including a road works section towards its end (see also Driving Task and Tracks section). In-vehicle information about an upcoming or current traffic situation was presented at the right hand side of the driver on a seven-inch MiMo screen. A Tobii eye tracker was installed below the small information screen for recording respective gazes onto this screen. A detailed description of the information screen and the presented bar variant combinations will be provided in the following section.

4.3 Information Screen and Presented Bar Types

Figure 5 depicts an example of all items displayed on the information screen.

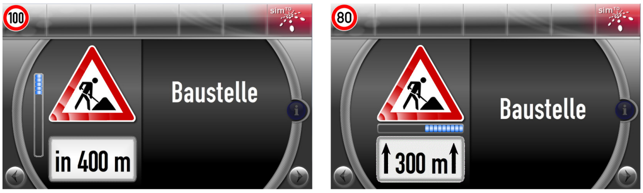


Fig. 5. Two screenshots of displays for the road works use-case. Distance is indicated in numerical form (in 400m/ 300m) accompanied by a graphical bar representing relative spatial information. The left side depicts a screenshot for approaching road works and the right side shows a presentation while driving within a road works section.

Besides the current speed limit in the upper left corner, the type of the upcoming traffic event (road works) is presented. Most important, numerical distance indicators combined with spatial bars were presented in all cases along the general lines of [10]. For approaching road works, a vertical bar was used and while driving through a section of road works, a horizontal bar indicated the current progress. Hence, the alignment of a bar itself would provide information whether a driver is in front of or within a road works section. Looking only at a screenshot, one cannot infer how a bar is animated. For example, it is not possible to decide, whether the bar depicted in Figure 5 (left) is decreasing upwards or increasing downwards. This is one reason why it is important to use dynamically presented bars in an actual driving task to enable a comparison between the different bar types.

In Figure 6 we present an overview of the two experimental factors composition type and moving direction leading to four experimental conditions. Three snapshots illustrate the animation of each moving bar (from the leftmost to the rightmost). For each experimental condition a combination of an approach bar (either increase or decrease; either upwards or downwards) for approaching the road works was successively combined with its consistent progress bar (increase or decrease; all moving from left to right) within the road works. This resulted in four main experimental conditions (increasing upwards: *inc_up;* decreasing upwards*; dec_up,* increasing downwards: *inc_dw;* decreasing downwards: *dec_dw;* cf. Figure 6).

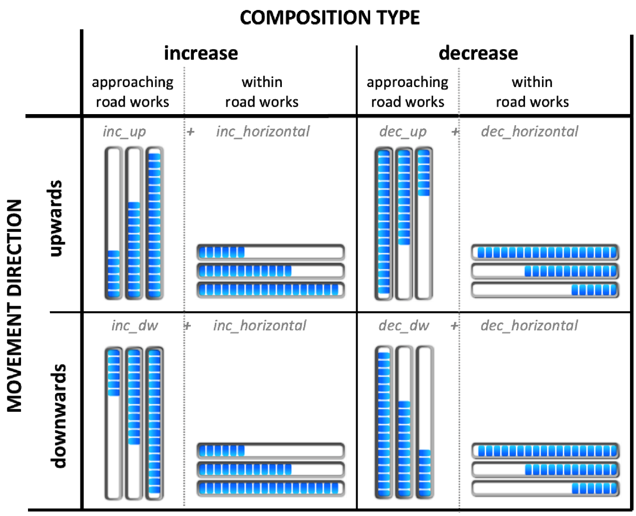


Fig. 6. Three snapshots illustrate the animation of each bar variant. Overview of four main experimental conditions: For each track an approach bar (either increase or decrease; either upwards or downwards) was combined with its consistent progress bar (increase or decrease; all moving from left to right).

As an informal pretest among eight colleagues had revealed, three bar types were favored or supported, whereas one variant (downward increase: *inc_dw*) was not appreciated by anybody. Furthermore, horizontal progress bars moving leftwards were evaluated as highly unsuitable. For reason of design completeness, we nevertheless kept the *inc_dw* condition, but we omitted the leftwards progress bars – also analogous to desktop solutions.

As participants might behave differently when for the very first time encountering a bar in the road works scenario, we considered that counterbalancing presentation order of conditions would not to be sufficient with respect to our research questions. Therefore, we decided to use the rather unpopular variant *inc_dw* as introductory condition. A second version of the *inc_dw* condition lacking the boundary line was generated and added as a fifth condition. The gray framing line is used for all four main bar variant conditions and indicates absolute minimum and maximum of bar states. Therefore, it might be a helpful reference for interpretation – especially for inexperienced users who do not know the bar range yet. To test this assumption, condition *inc_dw* and *inc_dw_fl* (fl = frameless) were alternately used between participants for either the introductory track (1) or the last track (5). Accordingly, the *inc_dw* condition was experienced only half by novice participants and half by experienced participants. This kept the *inc_dw* condition comparable with the remaining three major experimental conditions (*inc_up, dec_up, dec_dw*), which were only assigned to middle tracks (2-4). Order of five equivalent physical driving tracks (with slightly differing distances, see also Driving Task and Tracks section) remained constant. The assignment of the three middle experimental conditions to the fixed tracks (2-4) was counterbalanced between participants to control for learning or sequence effects as well.

For each bar type (except the *inc_dw_fl* condition), an animated color-coded version of the approach bars was additionally presented to participants (see Figure 7). We investigated, whether participants would prefer color-coding representing proximity and accordingly urgency over plain-colored bars.

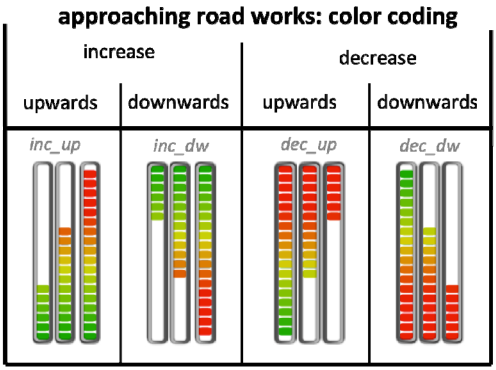


Fig. 7. Additional color-coded versions of approach bars presented to participants

4.4 Design

Two factors, movement direction (upwards, downwards) and composition type (increasing, decreasing) were manipulated within participants. A reference frame line as visual reference for minimum and maximum bar states was used for all four conditions. The *inc_dw* condition was additionally presented in a frameless version (*inc_dw_fl)*. Half of the participants started with a *inc_dw* track and finished with a *inc_dw_fl* track, whereas for the remaining half of the participants the order was reversed.

4.5 Driving Task and Tracks

Participants were instructed to drive as they would in a real world scenario. They should conform to German traffic rules (speed limits, taking the right lane whenever possible, not exceeding the current lane). Besides, they should finish each track as fast as possible. Moreover, they were informed that the small screen would provide helpful information about the upcoming traffic situations. Participants had to perform five equivalent two-lane motorway tracks without any other vehicles present. An example containing concrete distances between the events is depicted in Figure 8. All experimental tracks were designed according to the same scheme, but each track was created with different distances. This should prevent learning effects about the exact event positions and times.

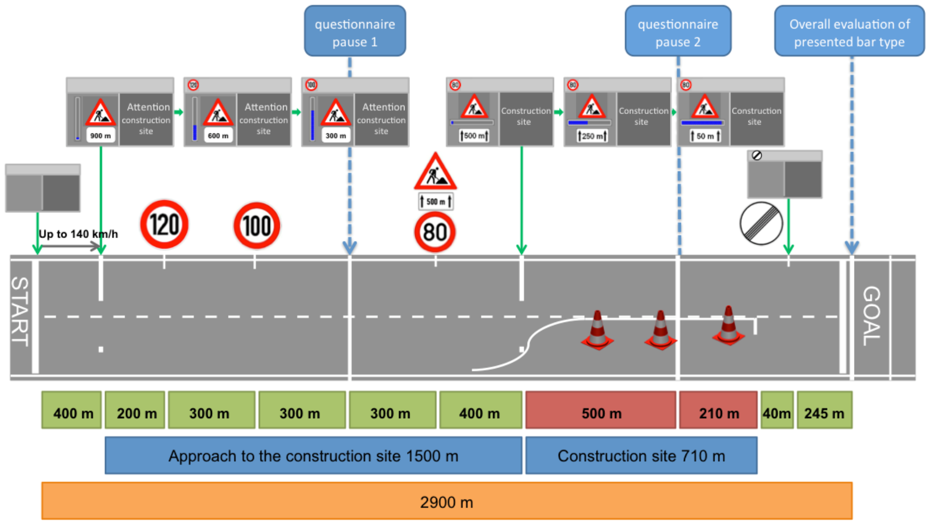


Fig. 8. Schematic overview of an experimental track containing speed limits, road works, and two pauses for questionnaires

After an acceleration phase without speed limit (acceleration to 140 km/h), a limit of 120 kilometers per hour requiring an immediate brake reaction was suddenly presented both on a road sign and on the small information screen. This speed limit was presented due to approaching road works. After a while, further speed limits showing 100 and 80 kilometers per hour were presented successively. When arriving at the road works it became obvious that the right lane was barred and participants had to change to the left lane. After passing the road works section at a speed of 80 kilometers per hour, the speed limit is revoked and participants should accelerate again. Shortly after, the track was finished. During each track, drivers were given notice of the road works on the small screen: Before arriving at the road works (from the first speed limit onwards), the remaining distance was shown by a numerical sign in combination with a vertical approach bar. Once the single lane within the construction work was entered, numbers combined with a horizontal progress bar displayed the remaining distance until the end of the road works segment. Having described participants, apparatus, experimental conditions, and the driving task, the next section will provide a description of the overall experimental procedure.

4.6 Procedure

After having been welcomed, participants filled out a first demographic questionnaire. The experimenter explained the driving task to the participant, and had her get familiar with the setup in a first practice drive on a straight road in the simulator. Subsequently, the participant completed a first baseline drive containing merely diverse speed limits but no road works or obstacles. Afterwards, the main experimental part was started, without the participant knowing anything about the upcoming road works in the first drive. The expectancy level certainly changed for the subsequent tracks, as both road works and speed limits reoccurred in each case. Nevertheless, exact locations were varied for each track, in order to prevent time-related expectancies and ignoring of the presented distance information. Within two breaks during each track and at the end of each track, participants had to fill out a questionnaire evaluating the current bar variant. Between plain-colored bars, color-coded bar variants were presented and evaluated as well. Following all experimental tracks, a second baseline drive had to be completed. After a final questionnaire participants were paid and said goodbye.

4.7 Dependent Variables

Most importantly, we assessed user preferences regarding bar presentation in general and intuitiveness of the presented bar alternatives. We were moreover interested, whether users interpret the meaning of some bars more consistently than the meaning of others. We assessed this by simply asking participants where they located themselves and the road works on the respective bar type during a short (driving) pause. Furthermore, we evaluated participants' gaze behavior onto the information screen, as this might vary for the different bar types. Low gaze distraction is on the one hand itself favorable for the driving context, and on the other hand it might serve as an implicit measure of presentation intuitiveness. Even though our driving task was rather simple, we recorded driving performance to control for any unforeseen disturbances or effects. For all dependent variables we intended to test whether for each factor main effects and interactions could be found, finally leading to a recommendation, which bar types seem(s) most suitable.

5 Results

As pointed out earlier, we wanted to generate a dynamic situation in a driving context that allows above all a valid subjective evaluation of different graphical distance estimation support variants. Possible differences in driving performance or glance behavior should be investigated as further dependent variables. Accordingly, we start with a rather detailed analysis of user preferences measured at different times during the experiment with the help of tailored questionnaires.

5.1 User Preferences and Perceptions

For a start, we present several results from the concluding questionnaire at the very end of the experiment. We asked participants their general opinion about the usage of progress bars. 75 % of the participants appreciate progress bars as very useful or rather useful addition to numerical distance indicators, whereas only 25 % of the participants rated them average, little or not useful at all. Hence, for the majority of drivers using bar indicators is valuable. Equivalent answers were given for using a differing alignment (vertical vs. horizontal) for approach and progress information. This leads to the assumption that different alignments predefined for different types of spatial information were assessed as a valuable premise. We state that this supports differentiation between approach and progress at a glance. In terms of use-cases, participants prefer progress bars for traffic events like traffic jams (mean = 4.1), or obstacles (mean = 4.1), whereas these bars are less appreciated for events like speed limits (mean = 2.8). An alternative design draft containing small symbols indicating car and hazard position received average ratings (mean = 3.3), whereas modifying the small segments into arrows indicating the movement direction of the bar (analogous to the Falk screenshot in Figure 3d) received comparably positive ratings (mean = 3.7).

As a second step, we elaborate on the results for the vertical approach bar alternatives. During a short pause within each approach phase (see Figure 8), participants should mark on a printed screenshot of the respective bar where they located themselves, as well as the beginning of the road works. For upwards-moving bars we initially considered the road works to be represented by the upper frame line. For the downwards-moving bars, we initially considered road works to be represented at the bottom frame line. For all conditions, the driver's position was presumably represented within the frame line at the changeover from bar content to empty. Only if a participant indicated both positions accordingly her answer was coded as "correct". Participants could choose not to answer if they were unsure, but this was not counted as correct either. As becomes obvious in Figure 9, participants were quite sure about their answers for the upwards-moving bars leading to a high percentage of correct answers. Regardless of our initial hypothesis of "correct" positions, results reveal that participants immediately access concordant mental models for this type of visualization. On the contrary, major dissent occurred for the downwards-moving bars. We found a high proportion of at least partially incorrect answers and non-answers. This reveals an inconsistent interpretation (between subjects) and uncertainty for bars moving downwards. A Chi-Square test confirmed the improved performance levels for two upward compared with two downward approach bars, $\chi^2(1) = 60.33$, $p < .05$.

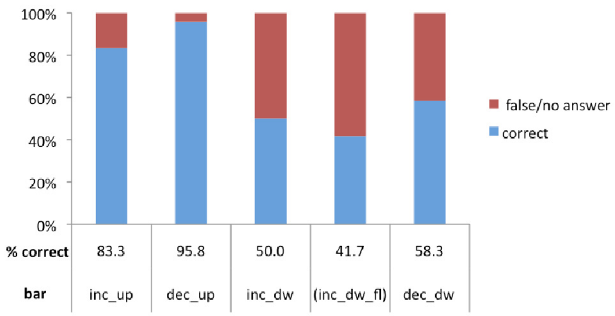


Fig. 9. Correct (red) versus at least partly incorrect (false or no answer; blue) answer percentage to the question: "Where are your car and the beginning of the road works located on this bar?". Results indicate consistency of bar assessment with xy.

Furthermore, participants were asked both during the test tracks as well as at the very end of the experiment, how intuitive they rated each graphical distance presentation from 1 – not intuitive to 5 – very intuitive. Figure 10 shows the results for the later ratings using a stacked bar graph [2].

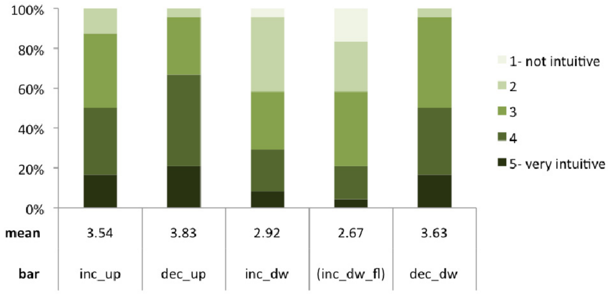


Fig. 10. Stacked bar graph illustrating the percentage of participants rating each approach bar from 1 – not intuitive to 5 – very intuitive. Darker bars indicate higher intuitiveness ratings.

A Repeated Measures MANOVA (i.e., we used the multivariate approach to the repeated measures analysis; see, e.g. [11, 12] with the factors movement direction (upwards vs. downwards) and composition type (increase vs. decrease) was conducted for the four main bar types. Movement direction had a significant influence on ratings with upwards moving bars being rated higher than the downward ones, $F(1,23) = 12.49$, $p < .01$. However, neither composition type, $F(1,23) = 2.34$, $p = .14$ nor the interaction revealed significant effects, $F(1,23) < 1$, *ns*. Another comparison targets the factor reference frame line, which has been altered between bar *inc_dw* and *inc_dw_fl*. Even though the order (first vs. last track) was counterbalanced, ratings did not show any significant difference $t(23) = -.77$, *ns*. However, relatively low ratings for a bar increasing downwards (*inc_dw*) might have concealed further negative influence of a missing reference frame (floor effect). Anyway, we would suggest using a frame line, as it does not worsen results. According to the results above, upwards-moving bars (*inc_up, dec_up*) score higher than downwards-moving bars. For the two upwards bars, we furthermore wanted to investigate, whether additional color-coding (see Figure 7) would influence subjective assessment. Accordingly, we conducted a Repeated Measures MANOVA on bar ratings, with the factors composition type (*inc_up* vs. *dec_up*) and color-coding (with vs. without). A significant main effect of color-coding, $F(1,23) = 9.31$, $p < .01$, indicates that participants clearly prefer color coding. Neither the other main effect of composition, $F(1,23) = 2.46$, $p = .13$, nor the interaction, $F < 1$, *ns.*, led to significant results.

Beyond the four vertical approach bars, we present results for the corresponding two bars for progress within the road works (*decreasing horizontal* or *increasing horizontal*). During another short pause within the road works section (see Figure 8), participants should mark on a printed screenshot of the respective bar where they

[2] The later overall ratings did not differ compared with the earlier ratings, $F(1,20) = 2.64$, *ns*. The interaction between point of time and bar variants also missed the level of significance, $F(4,17) = 2.42$, $p = .09$. Accordingly, the pattern of ratings does not change significantly over time and it is sufficient to only present the later ratings in detail.

perceived themselves, as well as beginning and end of the road works. For both bar types we considered the beginning of the road works to be represented by the very left frame line and the end at the very right. Consistent with the approach bars, the driver was presumably represented within the frame line at the changeover from bar content to empty. If participants were unsure about their assessment, they could also give no answer. Results were coded as “correct” if all position indications were as we had expected. The percentage of correct answers about the car’s and the road works’ position was relatively high for the decreasing (83%) as well as for the increasing horizontal bar variant (88%). For the frameless variant (horizontally increasing), rather inconsistent interpretations were achieved (63%). To conclude, both framed progress bars were interpreted quite consistent with our initial assumption. These findings do at least not contradict our initial assumption that preselecting progress bars moving rightwards is feasible.

When having a look at the subjective ratings about the intuitiveness of progress bar types (cf. Figure 11), a significant tendency in favor of the decreasing progress bar (*dec_ri*), compared with the increasing progress bar (*inc_ri*) can be identified, $t = 2.15$, $p < .05$. A comparison of both increasing bar types revealed significantly better ratings for the version with frame line (*inc_ri*) compared with the frameless bar (*inc_ri_fl*), $t(21) = 2.57$, $p < .05$.

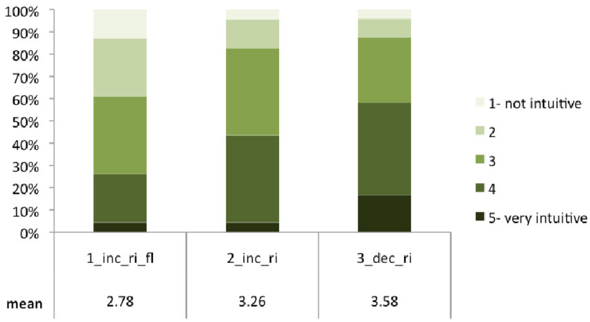


Fig. 11. Stacked bar graph illustrating the percentage of participants rating each progress bar from 1 – not intuitive to 5 – very intuitive. Darker bars indicate higher intuitiveness ratings.

5.2 Gaze Behavior

Besides gathering subjective ratings, also participants’ gaze behavior onto the small information screen was recorded and analyzed, as glance behavior can serve as a valuable indicator for visual distraction from the road. We recorded average glance duration, maximum glance duration, number of glances during each track, and the proportion of time that participants glanced onto the small screen while driving. Note that results only labeled according to the five vertical bar variants (see Table 2) but they also include the horizontal bar increasing or decreasing consistently in each track.

Table 2. Average fixation duration, number of glances, maximum glance duration, and percentage of time looking at the information screen for the five tracks

Condition	*inc_up*	*dec_up*	*inc_dw*	*inc_dw_fl*	*dec_dw*	BL 1	BL 2
Average glance duration [ms]	552	533	507	568	510	420	346
Number of glances	**22.4**	**19.8**	**21.0**	23.5	**18.6**	8	7
Maximum glance duration [ms]	1282	1236	1289	1286	1209	792	567
Gaze % of driving time	**10.4**	**8.7**	**9.3**	10.5	**8.5**	3.9	3.5

First of all, we compared the initial and final baseline drives with an average of all four experimental conditions for all dependent variables. In the baseline drives merely the current speed limits were presented on the small screen according to the road signs in the simulator. Therefore, we expected lower values of gaze measurement for baseline tracks. This was confirmed for all collected parameters, $F(1,13) > 12.79$, $p < .001$, for 14 out of 24 participants. Consistent with the implication of this finding, the remaining 10 participants did not glance onto the small screen at all in at least one of the baseline conditions, and therefore had to be excluded from the comparison above.

Most important we wanted to have a closer look at the two experimental factors movement direction and composition type. Analogous to the self-reported metrics in user preferences section, we entered all gaze metrics (average glance duration, maximum glance duration, number of glances, gaze percentage) of the four main bar alternatives into a two-way Repeated Measures MANOVA. We did not receive any significant effects for the average glance duration, $F(1,23) < 1.57$, *n.s.*, and for the maximum glance duration, $F(1,23) < 1$, *n.s.* This indicates that the duration of single glances does not vary between our variants. However, most interestingly the number of glances per track, $F(1,23) = 14.05$, $p < .001$, as well as the overall percentage of experimental time spent with looking at the small information screen, $F(1,23) = 7.25$, $p < .05$, showed a consistent main effect of composition type: Decreasing bars lead to less glances and less overall glancing time on the small information screen (cf. Figure 12). Neither significant main effects for movement direction, nor significant interactions were found. To sum up, participants rated the bars differently rather with respect to movement direction, but for gaze behavior the composition type seems to be more decisive.

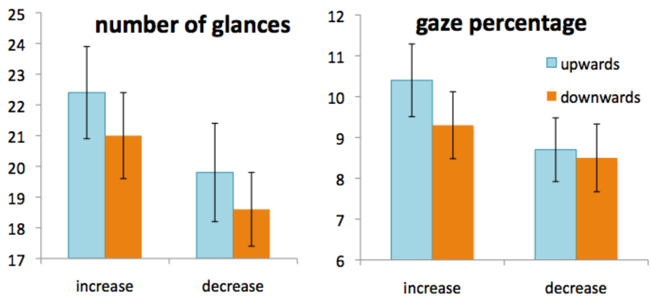


Fig. 12. Number of glances and percentage of gazes onto the small information screen during a track. Decreasing bars lead to significantly lower visual distraction than increasing bars.

5.3 Driving Performance

Even though the driving task was very simple and all the tracks were designed to be similar, we investigated driving performance to control whether any unforeseen disturbance or confound had influenced driving behavior. Reaction times upon speed limits, deviation from the current speed limit, and deviation from a predefined ideal line (drive in the middle of the respective lane, fast lane change when necessary) were logged. Thereby, Neither reaction times upon speed limits, nor deviation of the ideal line, nor deviation of the current speed maximum revealed any significant differences, all F‘s$(1,23) < 1$, *ns*. In summary, as expectable in this scenario no performance differences could be observed for the presented bar alternatives in the driving simulation scenario.

6 Discussion, Conclusion, and Future Work

Our results imply that participants appreciate the distinction of vertical approach bars and horizontal progress bars. For approach bars, upwards-moving bars are preferred over downwards-moving ones, whereas composition type ratings do not differ significantly. However, in gaze-based metrics decreasing bars induced less visual distraction from the primary driving task. This might occur because decreasing bar segments themselves contain all necessary spatial information without requiring a comparison with a further reference (e.g. frame line). Furthermore, for horizontal progress bars the decreasing bar compatibly received higher ratings as well. Moreover, color-coding leads to higher intuitiveness ratings, but obviously designers need to consider the trade-off between potential benefit and increased visual complexity in each individual case.

According to our findings, we suggest to use an approach bar decreasing in an upward direction. Spatial progress within a traffic event should consistently be presented via a horizontal bar decreasing to the right (see also Figure 13).

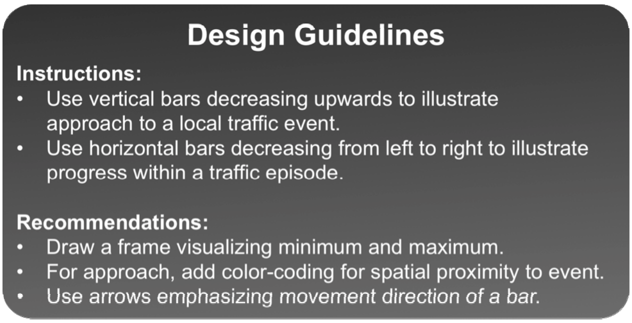


Fig. 13. Design guidelines for graphical bars representing spatial information when approaching local traffic events or for current progress within a traffic event section

On account of these guidelines, we designed graphical bars visualizing spatial approach and progress for the sim^{TD} system that has been introduced earlier. Figure 14 depicts example screenshots containing a final solution for approach and progress bars. In the field operational test the bars are consistently used for all appropriate use-cases, e.g. obstacle warnings, traffic jam ahead warning, road signs.

Fig. 14. Official sim^{TD} screenshots of road works information containing a color-coded approach bar (left) and a single-colored horizontal progress bar (right)

Our results are a first step towards a consistent common solution for assisting the driver with graphical distance and progress information. Agreed design guidelines for human interface designers in the automotive domain would help users to immediately comprehend the presented spatial information in diverse in-vehicle information systems at a glance and with minimal cognitive effort. However, further research is needed to quantify benefits from the most suitable bar version compared with no graphical distance indicator. Even though we found out that the majority of drivers prefer bars in our study, we still lack knowledge, whether a well-designed bar should be prescribed for all appropriate systems. For this comparison of driver distraction we would recommend a more realistic driving environment providing more ecologic validity. Accordingly, differences in gaze behavior and driving performance would become more obvious – preparing the ground for investigation of additional moderators like gender or prior experience.

Acknowledgements. This work was funded within the project sim^{TD} by the German Federal Ministries of Economics and Technology as well as Education and Research, and supported by the Federal Ministry of Transport, Building and Urban Development. EIT ITC Labs supported OpenDS via the activity Apps for Your Car (AIMS 13131) in the context of the project GetHomeSafe within the scope of the Seventh Framework Program of the European Commission.

References

1. "Driver Focus-Telematics Working Group": Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems (2006), `http://www.autoalliance.org/`
2. Hamada, K., Yoshida, K., Ohnishi, K., Koppen, M.: Color Effect on Subjective Perception of Progress Bar Speed. In: Third IEEE International Conference on Intelligent Networking and Collaborative Systems, pp. 863–866. Institute of Electrical and Electronics Engineers (IEEE), Fukuoka (2011)

3. Harrison, C., Yeo, Z., Hudson, S.E.: Faster Progress Bars: Manipulating Perceived Duration with Visual Augmentations. In: Proceedings of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1545–1548. ACM, New York (2010)
4. Harrison, C., Amento, B., Kuznetsov, S.: Rethinking the progress bar. In: Proceedings of the 20th Annual ACM Symposium on User interface Software and Technology, pp. 115–118. ACM, New York (2007)
5. Hsu, S., Le Prado, C., Natkin, S., Liard, C.: New Type of Auditory Progress Bar: Exploration, Design and Evaluation. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 868–877. Springer, Heidelberg (2007)
6. Brewster, S.A., King, A.: The design and evaluation of a vibrotactile progress bar. In: First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, pp. 499–500. Institute of Electrical and Electronics Engineers (IEEE), Pisa (2005)
7. Myers, B.A.: The importance of percent-done progress indicators for computer-human interfaces. In: Borman, L., Curtis, B. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 11–17. ACM, New York (1985)
8. Math, R., Mahr, A., Moniri, M.M., Müller, C.: OpenDS: A new open-source driving simulator for research. In: Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Appilcations, Portsmouth, NH, USA, pp. 7–8., http://www.auto-ui.org/12/
9. Paßmann, C., Schaaf, G., Naab, K.: simTD Project result: Selection of Functions, Deliverable D11.2 (2009), http://www.simtd.de
10. "the International Organization for Standardization": ISO/TR 16352: Road vehicles - Ergonomic aspects of in-vehicle presentation for transport information and control systems - Warning systems (2005)
11. Dien, J., Santuzzi, A.M.: Application of repeated-measures ANOVA to high-density ERP datasets: A review and tutorial. In: Handy, T. (ed.) Event-Related Potentials: a Methods Handbook, pp. 57–82. MIT Press, Cambridge (2005)
12. O'Brien, R., Kaiser, M.: MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin 97, 316–333 (1985)

Culturally Independent Gestures for In-Car Interactions

Sebastian Loehmann[1], Martin Knobel[2], Melanie Lamara[2], and Andreas Butz[1]

[1] University of Munich (LMU), HCI Group, Amalienstr. 17, 80333 Munich, Germany
[2] BMW Group Research and Technology, Hanauer Str. 46, 80993 Munich, Germany
{sebastian.loehmann,andreas.butz}@ifi.lmu.de,
{martin.knobel,melanie.lamara}@bmw.de

Abstract. In this paper we report on our ongoing work to introduce freehand gestures in cars as an alternative input modality. Contactless gestures have hardly been successful in cars so far, but have received attention in other contexts recently. We propose a way to achieve a better acceptance by both drivers and car manufacturers. Using a four-step process, we developed a small set of culturally independent and therefore easy-to-learn gestures, which can be used universally across different devices. We built a first prototype using distance sensors to detect the stop gesture in front of several devices. We conducted a user study during actual driving situations, testing the pragmatic and hedonic quality of the approach as well as its attractiveness. The results show a high acceptance of our approach and confirm the potential of freehand gestures as an alternative input modality in the car.

Keywords: freehand gestures, automotive user interface, culturally independent.

1 Introduction

Gestures have always been an important part of interpersonal communication. They are closely connected to speech, giving a deeper insight into speakers' thoughts [7]. When communication partners speak diverse languages, gestures help them to increase their mutual understanding. Furthermore, gestures are essential in situations in which other communication channels cannot be used. Examples are the gestures performed by divers or the sign language created for the deaf.

With the introduction of Microsoft Kinect[1] and the Leap Motion Controller[2], freehand gestures have moved into the focus of HCI research again. These sensors can be seen as an enabling technology bringing gestures into researchers' labs and people's living rooms by its easy handling and low cost. Being a novel input modality, gesture-controlled interfaces seem to stimulate people in a positive way and to evoke feelings like joy and pleasure. Another reason for the success of this type of interaction is its ease of use. Since gestures are monitored through the human's proprioception and can in many cases even be performed blindly, they need less visual attention compared to haptic controls such as buttons.

[1] http://www.microsoft.com/en-us/kinectforwindows/
[2] https://www.leapmotion.com/

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 538–545, 2013.

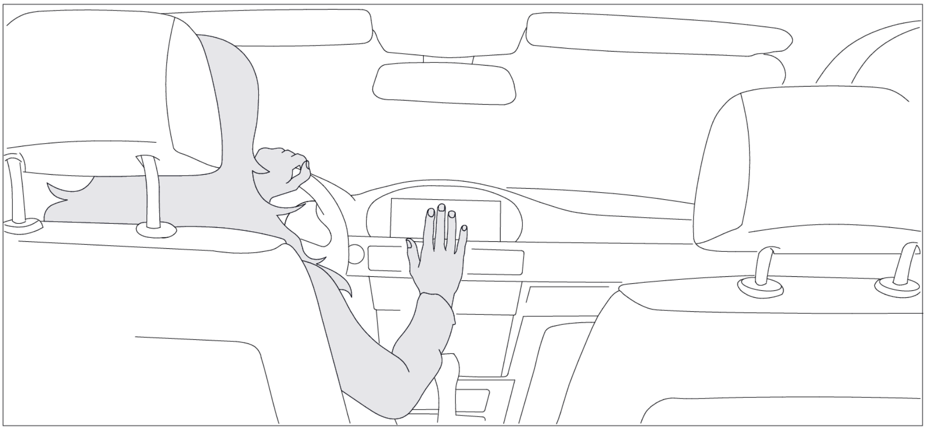


Fig. 1. Using freehand gestures as an input modality in cars

It has been shown that gestural input (see Fig. 1) significantly reduces drivers' visual distraction compared to haptic input [4]. This implies a potential increase in traffic safety by helping drivers to keep their eyes on the road [2], which makes gestures interesting again for car manufacturers. As a result of the increased interest, a first gestural interface has been introduced. The contactless tailgate opener [3] allows passengers to open the trunk by performing a foot gesture below the rear bumper. However, there is currently no interior gestural interface used in practice.

Taking into account lessons learnt from previous work on gestural interfaces and the limitations of the automotive context, we propose a novel approach for in-car gestures. Our goal is to increase their acceptance by manufacturers and drivers, which is often a challenge when moving interactive systems from the lab into practice.

In a pre-study, we identified a set of culturally independent gestures. The stop gesture was then successfully tested, providing early insights into drivers' acceptance of the approach. We built a prototype using infrared distance sensors and conducted a study in real driving situations, comparing the use of haptic controls and the freehand gesture in a quantitative and qualitative way as supposed in [11].

2 Related Work

In the past, several efforts (e.g. [1, 13]) have been made to bring gestures into the car. In spite of positive user feedback during studies, they were not commercially successful. Some drawbacks were caused by hand tracking based on infrared-cameras. Despite the complex software and expensive hardware, it is sensitive to varying lighting conditions [1], e.g. when driving out of a tunnel into direct sunlight. Furthermore, gestures are performed in a restricted interaction space above the gearshift [1], preventing direct interaction when controlling the dashboard's different devices. Another drawback seemed to be the rather large set of 22 gestures [13]. The need of remembering and recalling the correct movements might lead to a decreased acceptance.

More recent work introduced the Kinect to the automotive context. The combination of RGB- and depth-camera based tracking allows for a more robust gesture tracking and "is tolerant to the changing lighting conditions" [9]. In [10] the Kinect was mounted on the car's ceiling, successfully tracking gestures in the gearshift area. According to earlier work, direct interaction is not supported: all hand movements are tracked in the same spot but are used to control different applications. In [9] the sensor was placed on the dashboard and gestures were executed in front of the windshield. However, hand movements visible from the outside can cause misunderstandings and thus dangerous traffic situations. None of the systems described above were tested under realistic driving circumstances.

Regarding the creation of a gesture set, [12] stressed the importance of including end users into the design process. Nevertheless, [10] proposed to "balance user needs with technological requirements and limitations".

3 A New Approach

Taking lessons learnt from related work and the goal of using culturally independent gestures into account, our concept implemented the following principles:

Reduced Distraction. We utilize a general advantage of freehand gestures: As the movement needs to be less precise than finding and using a haptic control element, the interaction requires less hand-eye coordination and reduces visual distraction [4].

Direct Interaction. In prior automotive gestural systems, gestures were performed in a fixed interaction space [1, 9, 10, 13]. In contrast, we allow movements close to the intended device, enabling direct interaction. Additionally, visibility of the gestures to other traffic participants [9] can hereby be prevented, reducing the risk for possible misunderstandings.

Small Set of well-known Gestures. In [10] cultural differences were ignored to simplify the definition of the gesture set. We propose a process to identify few well-known and culturally independent gestures to support intuitiveness and achieve a low mental load [11]. According to [12], we included input by potential users but at the same time chose gestures to fit the specific automotive context as proposed by [10].

Universal Use of Gestures. We propose the application of gestures usable universally across different systems (e.g. radio or navigation). We believe that this will result in a good learnability, leading to instant usage and a higher acceptance by drivers.

4 Finding Universal Gestures

In order to find a small set of culturally independent gestures, we followed a four-step process. First, we collected 34 gestures in the corresponding literature (e.g. [7, 8]). They were chosen because of their explicit but also culturally independent meaning. All gestures are permanently used during interpersonal communication [7, 8]. Thus, they should be easy to understand and to remember.

Fig. 2. Reduced Gesture Set. Upper row from left: Slow Down, Clap and Fist. Lower row from left: Go Away, Stroke and Stop.

To reduce the initial set, we conducted a study involving possible end users (five participants, two female) in the second step. For each gesture, they received detailed descriptions, performed the corresponding movement and were interviewed about their subjective meaning. Using this information, we eliminated ambiguous gestures, resulting in a set of 18 gestures, each of them having a unique interpretation.

In the third step, four automotive experts (one female) were asked to rate the remaining 18 gestures. The goal was to identify redundancies concerning similar movement profiles. We produced 72 video clips showing each gesture from four different angles. During this expert review, twelve gestures were removed. An example is the similarity of the greeting gesture (whole hand, fingers close together) to the stop gesture (whole hand, fingers spread, see Fig. 3 on the right). The remaining set consists of six gestures, each unique in meaning and movement profile (see Fig. 2).

In the fourth and final step, the remaining gestures were rated by 91 participants (45 female) via an online survey. The results provided deeper insights into explicit meaning, implicated emotions as well as the compatibility of each gesture to the automotive context. To collect first evidence towards the acceptance of the identified gestures, we tested one of them in an early prototype. Based on the results of the online questionnaire, we chose to implement the stop-gesture (see Fig. 2, lower right).

5 User Study

We conducted a study to collect first evidence concerning the acceptance of the described approach. The experiment took place under real driving conditions. Our first goal was to determine if a device (e.g. radio) could be controlled by a gesture without accidentally triggering a spatially close second device (e.g. air conditioning). Second, we wanted to test whether participants preferred haptic or gestural input for the given tasks. Combining both quantitative and qualitative results, the goal was to verify whether a high acceptance for the use of the gesture was achieved by the approach.

5.1 Prototype

To detect the hand in front of interactive in-car devices, Sharp distance sensors emitting infrared light were installed on the dashboard of a car (see Fig. 3 left). The task was to interact with three different devices: The universal meaning of the stop-gesture was used to mute the radio, to turn off the ventilation of the air conditioning and to stop the navigation system. A stop-gesture was recognized, if the two sensors above and below a specific device were covered at the same time (see Fig. 3 right). All sensors were powered by an Arduino board, which was connected to a laptop. The system was able to recognize the stop-gesture for all three devices. Though, because of technical constraints, the corresponding car functions could not be triggered directly. Instead, the supervisor controlled the setup from the backseat. Whenever a gesture was recognized by the system, he triggered the corresponding functions via a laptop connected to the car's internal bus.

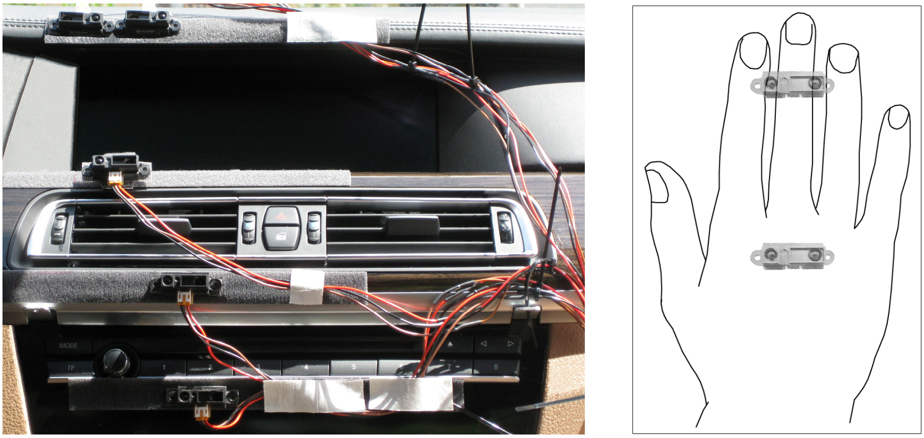


Fig. 3. Distance Sensors on the Dashboard of a car (left) detecting a Stop-Gesture (right)

5.2 Study Design

Participants were instructed to drive a predefined route in city traffic covering 9 kilometers, taking approximately 15 minutes. While driving, the three tasks (muting the radio, turning off the ventilation and stopping the navigation) were executed either by gestural or haptic input. The study followed a within-subjects design. Thus, each task was performed twice alternating the input modality. To avoid order effects, the sequence of the input modalities was counterbalanced. In a third round, both modalities were possible, giving participants the choice for every task. After each of the first two rounds, an AttrakDiff2 questionnaire [6] had to be completed. After round three, a questionnaire concerning the comparison of both modalities was provided.

5.3 Method

A German version of AttrakDiff2 [6] was used to evaluate the acceptance of both, gestural and haptic input. To evaluate the general appeal, we added the Attractiveness (ATT) scale from AttrakDiff1 [5]. In Hassenzahl's model [6] ATT consists of pragmatic quality (PQ), which is comparable to usability's "quality of use" [6] and hedonic quality (HQ), indicating the potential to fulfill pleasure in use and ownership. HQ is subdivided into stimulation (HQS, e.g. personal development) and identification (HQI, e.g. expressing oneself to others) [6]. The AttrakDiff2 including ATT comprises 28 seven-point semantic differential items, in each case seven for PQ (e.g. technical-human), HQI (e.g. isolating-connective.), HQS (e.g. conservative-innovative) and ATT (e.g. repelling-appealing).

Cronbach's alpha[3] for haptic interaction ranges between satisfying .66 (HQI) and .87 (ATT), both HQ values for gestures are not acceptable. After two HQI items (unprofessional-professional, tacky-stylish) were excluded based on an item analysis, Cronbach's alpha changes to .69 for both interaction types. Despite the inacceptable internal consistency of HQS for gestures (.45), we decided to include this scale exceptionally because composite reliability of .68 for both interaction types is acceptable.

Discriminant validity analysis show that ATT intercorrelates significantly with PQ and HQI between r=.58, p<0.01 and r=.76, p<0.001 for haptics and ranges between r=.60, p<0.01 and r=.78, p<0.001 for gestures, supporting Hassenzahl's model. For gesture interaction an additional significant correlation between PQ and HQI (r=.66, p<0.01) was found, whereas the averaged intercorrelation of HQI is lower. Since the composite correlation of PQ and HQI (r=.44) is higher than between HQI and HQS (r=.25) - both subscales of the higher order hedonic quality scale - we decided to analyze only the main factors ATT, PQ and HQ. As a side note: this study evaluates two interaction types for three functions instead of a general product. Thus, the limitations of the questionnaire might be a consequence of the narrowed interaction context.

5.4 Results

The 20 participants (two female) had an average age of 37 years, ranging from 26 to 57. All of them drive at least once a week, 85% of them more than 5000 kilometers a year. Results of AttrakDiff2 show that gestural interaction strengthens attractiveness and hedonic quality significantly compared to haptic interaction (see CIs in Table 1). Supporting previous results, there are no significant inter-correlations between both interaction forms. Findings suggest that well-known gestures like stop (see Fig. 3) are a desired interaction form in the automotive context. No significant Pragmatic differences show that both interaction types fulfill the intended functional requirements and are thus equally accepted. The comparison of the *technical-human* item between haptics (M=-1.35, SD=1.04) and gestures (M=1.15, SD=1.23) shows a significant difference (t=-7.43, df=19, p<0.001, d=1.66), indicating a gesture driven natural interaction.

---

[3] Cronbach's alpha is a psychometric criterion for test-quality and determines the internal consistency of a scale, respectively the correlation between the particular items.

The unpredictable-predictable item differs significantly (t=4.53, df=19, p<0.001, d=1.01), supporting haptics (M=2.20, SD=0.62) over gestures (M=0.75, SD=1.25). One reason is the participants' lack of experience in using gestures. All in all AttrakDiff2 results show that the gesture is accepted and favored in these use cases.

Table 1. Mean intensity (standard deviation, 95% confidence interval) Attractiveness (ATT) and AttrakDiff2 main factors Hedonic quality (HQ without unprofessional-professional and tacky-stylish) and Pragmatic quality (PQ)

N=20	**Haptics (95% CI)**			**Gestures (95% CI)**		
	M (*SD*)	**Lower**	**Upper**	**M (*SD*)**	**Lower**	**Upper**
ATT (7)	0.69 (*0.70*)	0.36	1.01	1.44 (*0.76*)	1.08	1.79
PQ (7)	1.02 (*0.71*)	0.69	1.36	1.30 (*0.83*)	0.91	1.69
HQ (12)	-0.48 (*0.57*)	-0.75	-0.21	1.27 (*0.38*)	1.09	1.45

To conclude the study, participants filled out a questionnaire. Being asked which input modality they liked better, 14 favored the gesture, 2 haptics and 4 both. This indicates the acceptance of the stop-gesture and reconfirms the AttrakDiff results. 16 stated that gestural input helped to focus on traffic (1 haptics, 3 both), indicating a potential reduction of visual distraction. Nobody said that gesturing influenced traffic safety negatively. All participants considered gestures to be a valuable alternative input modality and a helpful feature in the car. In round three, participants were able to choose between input modalities. Concerning the navigation task, gestural input was chosen in 30 out of 40 cases (ventilation 37 out of 40 and radio 24 out of 40).

Concerning the hardware, we discovered no influence of the lighting conditions throughout the study. Gesture detection was robust as no intention to perform a gesture was unsuccessful. Participants were continuously interacted with the correct device, meaning that the spatially close functions were easily discriminable.

6 Conclusions and Future Work

During real driving situations, we tested a first implementation of a new approach towards using freehand gestures as an alternative input modality in the automotive context. We tracked interactions of the driver's hand with devices in the dashboard. The stop-gesture was exemplarily used because of its universal use across devices and its cultural independency. Participants of the study performed the stop-gesture to turn off certain functions (e.g. mute the radio).

After exploring both haptic and gestural input, most drivers chose the gesture because of less presumed visual distraction and ease of use. Results of the AttrakDiff2 showed significantly higher attractiveness and hedonic qualities of gestural input. Furthermore, according to qualitative feedback, gesturing did not influence traffic safety negatively, but in a rather positive way. We can conclude that we found a high acceptance of our approach to gestural input in the case of a stop-gesture. This can be seen as an indication for the potential of choosing culturally independent freehand gestures to be an alternative input modality in automobiles in the near future. The results of this pilot-study motivate us to further develop our suggested approach.

As a next step, the improvement of the prototype will be most important. Due to the Wizard-of-Oz technique, there was a negatively rated delay of the functional feedback (e.g. the actual muting of the radio) after performing a gesture. This can be avoided by controlling the car's features directly via the internal bus so that no active supervisor is needed. Furthermore, we are considering the use of additional feedback. Some participants stated that the functional feedback (e.g. the sound of the radio) alone is not sufficient and for instance an audio signal could be helpful. We also noticed the request for more gestures. One driver spontaneously tried to turn the ventilation back on by waving the hand from the fan towards the body. Besides the fact that this is a positive observation concerning the acceptance of gestures, we will add functionality and further appropriate gestures to the prototype. For the implementation, we will need to improve tracking by using more flexible sensors like Kinect.

We are aware that the study setup with the use of one gesture is only a first step towards a comprehensive system. However, we are confident that the suggested approach can have a positive influence on the driver's visual distraction in the future.

Acknowledgments. We thank Simone Rodestock for her important contributions.

References

1. Akyol, S., Canzler, U., Bengler, K., Hahn, W.: Gesture control for use in automobiles. In: Proc. IAPR 2000, pp. 28–30 (2000)
2. Alpern, M., Minardo, K.: Developing a car gesture interface for use as a secondary task. In: Proc. CHI 2003 Ext. Abstracts, pp. 932–933 (2003)
3. BMW: Contactless Tailgate Opening (2013), `http://www.bmw.com/com/en/newvehicles/3series/sedan/2011/showroom/comfort/contactless_tailgate_opening.html` (accessed May 15, 2013)
4. Geiger, M., Zobl, M., Bengler, K., Lang, M.: Intermodal Differences in Distraction Effects while Controlling Automotive User Interfaces. In: Proc. HCII 2001, pp. 263–267 (2001)
5. Hassenzahl, M.: The Effect of Perceived Hedonic Quality on Product Appealingness. International Journal of Human-Computer Interaction 13(4), 481–499 (2001)
6. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction 19(4), 319–349 (2004)
7. McNeill, D.: Hand and mind: What gestures reveal about thought. University of Chicago Press, Chicago (1995)
8. Morris, D.: Gestures: Their Origins and Distribution. Jonathan Cape, London (1979)
9. Rahman, M., Saboune, J., El Saddik, A.: Motion-path Based in Car Gesture Control of the Multimedia Devices. In: Proc. DIVANet 2011 (2011)
10. Riener, A.: Hand and Finger Gestures in Vehicular Applications. Computer 45(4) (2012)
11. Wachs, J.P., Kölsch, M., Stern, H., Edan, Y.: Vision-Bases Hand Gesture Applications. Communications of the ACM 54(2), 60–71 (2011)
12. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092 (2009)
13. Zobl, M., Geiger, M., Schuller, B., Lang, M., Rigoll, G.: A real-time system for hand gesture controlled operation of in-car devices. In: Proc. ICME 2003, pp. 541–544 (2003)

Don't Text While Driving: The Effect of Smartphone Text Messaging on Road Safety during Simulated Driving

Kaspar Lyngsie[1,2], Martin S. Pedersen[1,3], Jan Stage[1], and Kim F. Vestergaard[1,4]

[1] Aalborg University, Department of Computer Science, 9220 Aalborg, Denmark
[2] Gamblify, Lyngbyvej 21, 2100 Copenhagen, Denmark
[3] Telenor, Skelagervej 9, 9000 Aalborg, Denmark
[4] Danske Bank, Edwin Rahrsvej 40, 8220 Brabrand, Denmark
kml@gamblify.com, novacdk@gmail.com,
jans@cs.aau.dk, kive@danskebank.dk

Abstract. Text messaging on smartphones uses a full soft keyboard instead of the numeric buttons on traditional mobile phones. While being more intuitive, the lack of tactile feedback from physical buttons increases the need for user focus, which may compromise safety in certain settings. This paper reports from an empirical study of the effect of text messaging on road safety. We compared the use of a traditional mobile phone and a smartphone for writing text messages during simulated driving. The results confirm that driver performance when texting decreases considerably as there are significant increases in reaction time, car-following distance, lane violation, number of crash/near-crash incidents, perceived task load and the amount of time the driver is looking away from the road. The results also show that smartphones makes this even worse; on key performance parameters they increase the threat from text messaging while driving. These results suggest that drivers should never text while driving, especially not with a smartphone.

Keywords: Driving, mobile phone, smartphone, text messaging, road safety, driver distraction, simulated driving experiment.

1 Introduction

Smartphone adoption is accelerating. In 2010, about 20% of the mobile phones sold were smartphones, and for 24-35 year olds 31% [16, 30]. A recent market analysis expected smartphone adoption to surpass that of traditional mobile phones by the end of 2011 [13]. The new smartphones provide more functionality, and touch-based interaction on a larger screen facilitates improvement of many functions.

Text messaging is a widely used form of communication on mobile phones. Early systems used the Multitap technique, where a letter is entered by pressing a numerical button a number of times. Later, Multitap was replaced by the predictive text entry system known as T9, which also relies on the numerical buttons but reduces the number of key taps and, thereby, allows text writing at the double speed [11]. On modern smartphones, text messages are written on a full QWERTY keyboard on the touch

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 546–563, 2013.

screen. This is obviously more intuitive for writing text messages than Multitap and T9. Yet the physical buttons on traditional mobile phones offer a tactile feedback that is lacking on the touch screen of a smartphone. This lack of tactile feedback may be a serious problem in special situations.

The number of drivers that are text messaging while driving has increased tremendously. A recent survey showed that as many as 51% of all young drivers admitted to have used text messaging during driving [15]. Up to one quarter of all car crashes are estimated to be a result of drivers engaging in distracting activities [19, 24]. Use of mobile phones while driving has a range of negative effects, including distracting the driver [26, 28, 36].

While the consequences of using mobile phones for calling and having conversations have been studied thoroughly, there are much fewer studies of the hazards of text messaging. Two recent studies found that text messaging while driving increases the risk of being in a crash or near-crash by 23 times [7, 16]. Unfortunately, the few studies that have been made are based on traditional mobile phones with tactile feedback [19, 23]. So far, it has not been documented how use of touch-based smartphones for texting while driving influence road safety.

Fig. 1. The keypad layout on a Nokia 3210

This paper reports from an empirical study of the effect of text messaging on road safety. The aim was to inquire into the safety hazards when texting on a smartphone while driving compared to a more traditional mobile telephone. In the following section, we describe text entry methods on mobile phones. Then we provide an overview of related work. The next section presents the method of our experiment that was carried out in a driving simulator. Then we present the results where we compare the two types of phones used for texting while driving. Next, the results are discussed and compared to related work. Finally, we provide the conclusion.

2 Text Entry Methods

The first text messaging systems, originally denoted as Short Message System or SMS, used the numerical keys on the mobile phone for text entry. The letters of the alphabet were distributed with more than one letter on each physical button, see Figure 1. The layout is a grid of 12 buttons, where the alphabet is distributed on buttons 2-9.

The Multitap technique was the first widely used implementation of text entry with this layout. With the Multitap technique, the user taps one or more times on a button to get a letter. To write the word ‘hey’, you need to press twice on button 4 to get ‘h’, twice on 3 to get ‘e’ and three times on 9 to get ‘y’. There is a timeout period of usually 0.5-2 seconds after each key press, which is used to delimit letter selections on the same button. So to write the word ‘hi’, you would need to press two times on 4 to get ‘h’, wait for the timeout period and then press three times on 4 to get ‘I’. The number of words that can be typed per minute with Multitap is about 5-10 wpm depending on experience [20].

In the late 1990s the “text on nine keys” (T9) predictive text entry technique was developed by Tegic Communications. It was based on the same keyboard layout as Multitap, but it used dictionaries to predict the desired text entry, based on the buttons being pressed. The T9 technique needs only one press on each button to predict what word is being written. To write the word ‘hey’, you only need to press the buttons 4, 3 and 9 to have the software suggest the word ‘hey’ and potentially other typical words that could be composed by tapping the same three buttons. The most frequent used word will be predicted first. T9 was quickly adopted as the primary form of text entry on products from all the major mobile phone manufacturing companies [11]. The number of words that can be typed per minute with T9 is about 7-25 wpm depending on experience [20], which is about twice the speed as with Multitap.

With the introduction of modern smartphones, the use of a mobile phone has exceeded that of simple telephony and text messaging, and the 9-button interface has been replaced by virtual full-size keyboards on touch screens. Smartphones are now also used for more advanced features such as document editing and mail composition and research are continuously attempting to increase the efficiency of the limited space available for text entry.

Fig. 2. Writing text on an Apple iPhone

Opposed to the common feature phones, most smartphones require larger screens in order to accommodate for the use of more advanced appliances such as email composition and text edition. This allows the smartphones such as Apple iPhone and many HTC Android-based phones to remove the physical keyboard all together and instead use an on-screen, full-size QWERTY keyboard which is stimulated through touch.

Writing text on a typical on-screen keyboard such as the Apple iPhone, see Figure 2, is done by touching the keyboard character on the screen. This solution also use dictionaries to predict the intended word, but does so by considering the surrounding characters for each key press [25]. A study on the Apple iPhone touch-based smartphone concluded the text entry speed to be 15.9 wpm [5].

3 Related Work

There is a broad variety of literature on mobile phone use while driving. Some of the literature emphasize the potential problems, e.g. that visual processing is affected by checking to see who is calling, motor control is affected when dialling a number, auditory distraction when attention is given to the conversation and not the sound of the traffic and higher-order (cognitive) processing when focusing on the conversation and not the act of driving [22]. There are also discussions where notable accidents with multiple casualties have been attributed to texting while driving [14].

Statistics show that teenagers are clearly the age group with the highest risk of being involved in accidents. Studies show that for every mile driven in the United States, teenagers are four times more likely to be involved in a car crash [2]. Studies on text-messaging habits in the last decade show an increase from 12 million to 135 billion text messages sent every month, where teenagers clearly being the most active age group, sending and receiving an average of 3,000 messages per month [6, 10]. Surveys conducted among teenagers conclude that half of all students admit to having texted while driving [19].

Table 1. Overview of related publications on mobile phone use while driving either in real world traffic or a car simulator

	Mobile Phone Task			
	Dialling	**Talking**		**Texting**
		Hand-held	**Hands-free**	
Real world	7, 32	7, 8, 31	8, 27	7
Simu-lator	28, 32	9, 28, 35	3, 4, 35	34

A significant number of empirical studies have inquired into the effects of mobile phone use on road safety, see Table 1. An early study, based on simulated driving, tried to establish the dangers when dialling and talking on a mobile phone [28].

A study of overall driving performance based on a car simulator concluded that a person talking on a mobile phone was more prone to being involved in a crash or near-crash event than a drunk driver [35]. A recent study based on a simulator concluded that the use of a mobile phone was much more distracting than a conversation with a passenger [9]. This confirms an early study in real world traffic where it was concluded that there is a much smaller risk when engaging in conversations with the passenger than talking on a mobile phone [8].

Many countries around the world have banned the use of hand-held mobile phones while driving. Yet an early study in a simulator concluded that a hands-free mobile phone was also dangerous to use while driving [3, 4]. This is confirmed by more recent studies in real traffic establishing that the risks of using a hands-free mobile phone are the same as using a hand-held, e.g. [27, 32]. The reason is that most of the distraction of talking on the phone is not related to visual processing or motor control, but to the cognitive workload.

Despite the dangers of hand-held and hands-free mobile phone usage, only very few countries have banned both, where more countries have customized laws depending on age group of the drivers or the location of where the driving is conducted [1]. Much debate has been raised about the effectiveness of prohibition, as the task of actually enforcing these laws has proven to be very difficult [17].

Studies conducted in real-world scenarios are harder to conduct, but have shown results that are similar to the car simulator studies. One study used accident analysis, where drivers who had been involved in car accidents were questioned about their mobile phone use prior to the accident, and their statements were compared to data from their mobile phone providers. The results showed that almost a quarter of these individuals had used their mobile phone in the 10 minutes preceding the crash [31].

In recent real world studies, the dangers associated with mobile phone usage has been quantified in more detail, where actions such as dialling a mobile phone increases the risk of accidents by 6 times and talking and listening double the risk [7].

Only few studies have been conducted on the subject of text messaging while driving. A notable exception is a study from Virginia Tech in 2009 where they observed drivers for more than 6 million miles of driving. It was concluded that texting while driving increased the risk of being involved in an accident by 23 times [7]. Most surprisingly, there is only a single study of text messaging on a smartphone, even though a market share analysis expects touch-based smartphone adoption to surpass that of feature phones by the end of 2011 [13]. In this single study, they inquired into texting on a Blackberry and an iPhone during simulated driving. Their focus was on the reasons why the number of texting car crashes is different and on differences between frequent and less frequent texters [34]. Thus they did not compare smartphones and traditional mobile telephones.

There are also a number of studies of the effects of interaction with other types of touch-based devices with while driving. One example is a study of GPS navigator use during real driving where they showed that the attention on visual output decreased driver performance significantly [21]. Another example is a study of the distraction caused by iPod use during simulated driving where they concluded that the effect was similar to that of operating a mobile phone [33].

4 Method

This section describes the experiment we conducted to study the effect on road safety when using a traditional mobile phone and a smartphone for writing text messages during simulated driving.

4.1 Setting

The study was conducted in a car simulator that was set up in the usability laboratory at HCI-Lab, Department of Computer Science, Aalborg University.

4.2 Apparatus

The car simulator was based on a computer with three screens, see Figure 3. The screen used for the front view of the road ahead was a 37” screen, and the side-view windows were two 32” screens. The screens were controlled by a powerful computer with extra graphics cards and a device for splitting the main display to the three screens.

Fig. 3. The physical setup of the simulator

The physical setup was centred around a passenger and a driver’s seat. A standard driving kit was used to control the simulator. The kit included a steering wheel, pedal board with gas, brake and clutch pedals and a manual gear shift. Tactile transducers called “bass shakers” were added to the driver’s seat to simulate engine vibration.

The software for the simulator was created using the Unity 3D game engine and the virtual world was modelled using CityScape. The car physics was based on an open source tutorial that was extended with support for the driving kit. A new module was developed to simulate traffic, and the city was fitted with intersections and traffic lights. A logger module recorded reaction times, velocity, distance from the middle of the lane etc.

4.3 Participants

The group of test subjects consisted of 28 people, where 18 were male and 10 were female. Among them, 10 normally used a T9 mobile phone and 18 normally used a touch-based smartphone. The participants all held valid driver's licenses and were in the age of 20-32 years old and had a variety of occupations.

4.4 Procedure

The subjects were instructed about the experiment. Then they made a test-drive in the car simulator in a city environment to get acquainted with the controls and behaviour of the car. No measurements were made during this drive.

Two driving scenarios were used: city and freeway. The freeway was a straight road with three lanes, a higher speed limit and traffic going at various speeds. The city had curved roads, intersections, traffic lights, a hilly terrain and traffic.

The experiment was conducted as between-subjects, where half of the subjects drove on the freeway, and the other half drove in the city. The participants were balanced over these two scenarios according to gender and mobile phone experience.

In each scenario, the subjects started in a baseline condition, which was a 5-minute drive without interacting with a mobile phone. Then they drove in the same scenario using one of the two mobile phones. Finally, they drove in the other scenario using the other mobile phone. The mobile phones were a touch-based smartphone and a tactile T9 phone. The uses of the two mobile phones were balanced in each scenario between test subjects in order to avoid a learning effect.

In the two conditions with mobile phones, the test subjects were instructed to enter up to five different text messages with a time limit of 5 minutes. If they did not complete all five messages during the 5-minute drive, the remaining messages were registered as in-complete. They were not required to complete the 5 messages. During the drive, the test monitor would read out the first sentence they had to write. When they had completed the sentence, they stated that, and the test monitor would read the next message and so on until they had written the five messages or the five minutes had passed. All text messages were randomly selected sentences of the same length and complexity [25] which were distributed among the subjects using latin squares. An example of a sentence is: “we are coming back tomorrow morning”.

In all three conditions, the test subjects had to follow a car driving in front of them. The car in front was programmed to brake at random intervals and it would continue to break until the test subject either pressed the brake pedal or collided with it. This is a commonly used way to measure the test subjects' reaction time in simulated driving [11].

After each condition, the test subjects filled in NASA Task Load Index (TLX)-scales, which is a NASA developed method for measuring task load [18].

4.5 Data Collection

Two cameras were used to record video: one focused on the test subject's eyes for eye glance recording and the other was focused directly down on the subject's interaction with the phone. These two images were mixed together with the front-view screen (the side-view screens were not recorded) to produce the material which was recorded on DVD and later analysed, see Figure 4.

The simulator was programmed to create log files for each condition. These log files included values such as distance to the centre of the lane, distance to the followed car, velocity, user reaction time, crashes and task completion times.

4.6 Data Analysis

Two aspects were analysed manually from the video recordings: near-crash incidents and eye glances.

Crashes and near-crashes are often used as a measure in the literature on driver performance. The simulator software was programmed to log crashes, while near-crashes had to be analysed manually.

Fig. 4. A subject making an eye glance while typing on a smartphone

For the purpose of this analysis, we used the related literature to define a near-crash as an incident where the driver was very close to crashing or only narrowly avoided a collision with the lead car or another car or object. For example, if the driver came into the opposite lane and only in the last fraction of a second avoided a head-on collision; or if the driver had to emergency brake to avoid the car he was following.

For eye glances, there was an error in the video recording for one test subject which implied that the eye glances could not be detected. Therefore, the data on eye glances are only from 27 test subjects. The one missing is from the city condition. The 27 recordings were analysed frame by frame for eye glances. The frame where the subject's eyes started moving away from the front window was registered as the first frame in the eye glance, and the frame before the first frame where the eyes were back on the front window was registered as the end of the eye glance. The length of each eye glance was calculated from the number of frames (it was recorded as 24 fps). This was done separately by three of the authors. After the analysis, they compared their results. The disagreements were mostly a couple of frames in the beginning or end of an eye glance. We calculated the Fleiss Kappa for the three raters across all eye glances. The result was Kappa=0.844, which is categorized as almost perfect agreement.

The NASA TLX answers and weights were grouped as prescribed by the manual [18]. The data from the driving scenarios was cleaned manually and organized using software developed specifically for this task and then imported for statistical analysis in the R statistical software environment.

We performed repeated measures ANOVA with condition as the repeated factor and road type as a between-subject factor.

5 Results

This section contains the results from the experiment, grouped into driving performance, task performance and eye glances.

Table 2. Means and standard deviations of driving performance

	Condition		
Variable	Baseline	T9	Touch
City			
Reaction time (ms)	1206 (454)	2200 (1104)	2363 (720)
Crash or near-crash	0.36 (0.50)	1.36 (1.39)	1.71 (1.38)
Following distance (m)	23.52 (6.52)	29.30 (6.62)	29.21 (69.44)
Following distance variability	10.66 (3.42)	14.09 (3.85)	14.43 (4.14)
Lane crossings per kilometer	8.98 (1.46)	9.58 (2.92)	10.08 (2.27)
Time in wrong lane per kilometer (s)	48.60 (20.14)	82.23 (46.70)	97.40 (37.05)
Lane variability	1.79 (0.12)	1.61 (0.36)	1.71 (0.25)
Freeway			
Reaction time (ms)	1281 (397)	2723 (1181)	2568 (908)
Crash or near-crash	0.64 (1.00)	1.14 (0.86)	2.43 (2.24)
Following distance (m)	32.50 (8.39)	43.73 (14.14)	42.85 (12.18)
Following distance variability	12.55 (4.08)	23.59 (9.41)	23.42 (8.60)
Lane crossings per kilometer	0.00 (0.00)	0.22 (0.56)	0.81 (1.03)
Time in wrong lane per kilometer (s)	0.00 (0.00)	4.027 (10.45)	10.73 (20.21
Lane variability	0.27 (0.07)	0.45 (0.17)	0.57 (0.30)

5.1 Driving Performance

The results for driving performance are shown in Table 2.

Reaction Time. The reaction time was approximately doubled, an increase of a little more than one second, when writing text messages, and this difference is significant both on the freeway (F=21.4, p<0.001) and in the city (F=17.6, p<0.001).

Assuming the subject is driving at 50 km/h, as the speed limit was in the city scenario, one second of increased reaction time would translate to an extra 13 meters of breaking distance.

There was no significant difference in reaction time between the Touch and T9 conditions. The analysis showed no significant differences in reaction time when driving in the city and on the freeway.

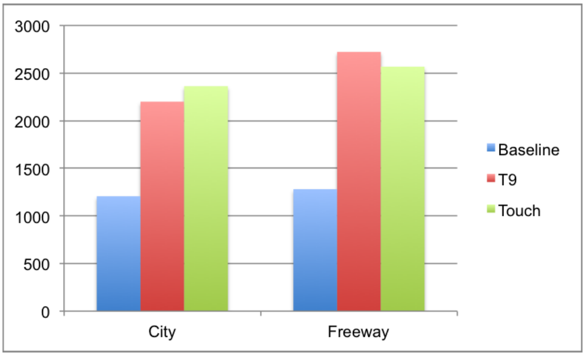


Fig. 5. Reaction time

Crash or Near-Crash. The number of crashes and near-crashes increased significantly between the baseline condition and the writing conditions both in the city (F=6.40, p<0.01) and on the freeway (F=5.13, p<0.05). There were significantly more crashes or near-crashes when writing with Touch than with T9 on the freeway (F = 5.92, p < 0.05). There is a similar tendency in the city, but this difference is not significant.

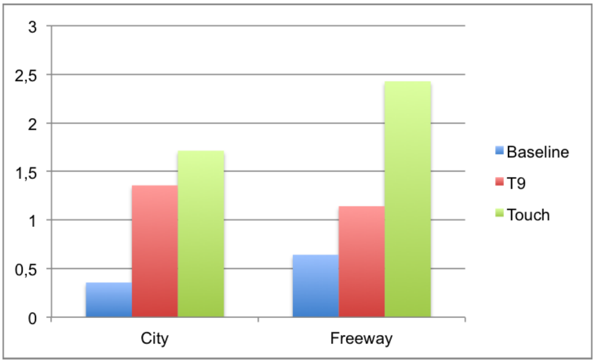


Fig. 6. Crash or near-crash

Following Distance. The average following distance shown in Table 2 reveals that the participants increased their distance to the car in front of them when they were writing text messages. The analysis showed significant difference between the baseline condition and the writing conditions both on the freeway (F=11.64, p<0.001) and in the city (F=11.59, p<0.001).

The variability in following distance increased significantly when the participants were writing text messages compared to the baseline condition both on the freeway (F=20.13, p<0.001) and in the city (F=11.04, p<0.001).

There were no significant differences in following distance or following distance variability between Touch and T9 conditions.

Lane Maintenance. The average lane crossings per kilometer increased in the writing conditions, but the increase was only significant on the freeway (F=6.28, p<0.01). The Touch condition had more lane crossings per kilometer than the T9 condition, and this difference was only significant on the freeway (F=5.66, p<0.05).

The difference in the time spent in the wrong lane between the baseline condition and the writing conditions is significant in the city (F=16.25, p<0.001), but not on the freeway. This means that although the number of lane crossings in the city is not significantly higher when writing, the time to correct and get the car back in the correct lane is longer.

The time spent in the wrong lane is also higher in the Touch condition than in the T9 condition, but this difference is not significant.

The variability within the lane increased when writing on the freeway, and this increase was significant (F=8.66, p<0.01). No such increase was found in the city, and there was no significant difference in lane variability between Touch and T9 conditions.

The lane variability indicates a severe change in driving behaviour, and swerving into another lane increases the risk of the driver being involved in a crash or near-crash situation. This risk increases the more time the driver spends in the wrong lane.

Table 3. Means and standard deviations of task performance

	Condition		
Variable	Baseline	T9	Touch
City			
Task load	26.42 (12.55)	52.67 (21.04)	58.19 (20.89)
Task completions	N/A	3.86 (1.56)	3.79 (1.58)
Task completion time (s)	N/A	38.17 (31.85)	44.87 (25.93)
Characters per minute	N/A	61.91 (24.92)	49.84 (26.59)
Freeway			
Task load	23.79 (15.29)	45.54 (26.42)	53.77 (26.89)
Task completions	N/A	4.64 (0.74)	4.00 (1.66)
Task completion time (s)	N/A	30.23 (8.85)	44.61 (29.34)
Characters per minute	N/A	53.55 (16.85)	47.14 (22.43)

5.2 Task Performance

The results for task performance are shown in Table 3.

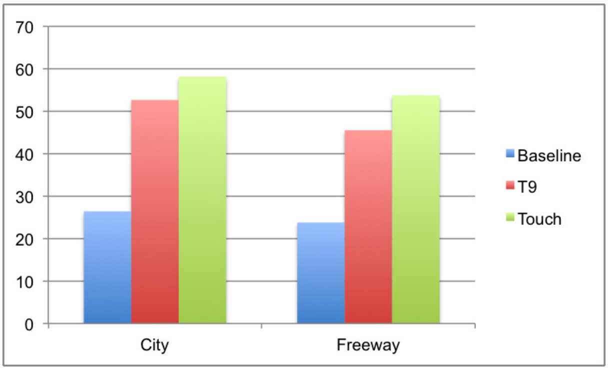


Fig. 7. Task load

Task Load. The subjects' average perceived task load doubled when writing text messages. This was significant both on the freeway (F=11.36, p<0.001) and in the city (F=17.73, p<0.001). The task load was also slightly higher when writing with Touch than it was when writing with T9, but this difference was found not to be significant. The average perceived task load was slightly higher when driving in the city than it was when driving on the freeway, but this difference was not significant.

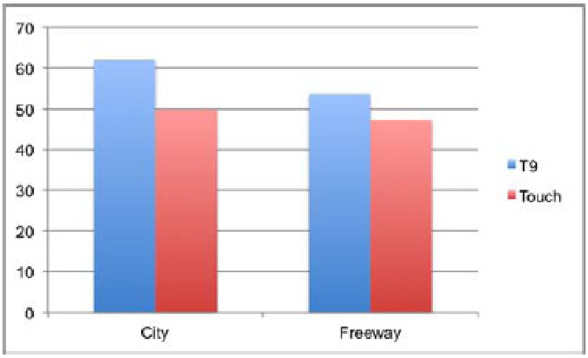


Fig. 8. Average characters per minute

Task Completion. The average number of successfully completed task per subject was higher for T9 than it was for Touch, but this difference was not found to be significant.

Writing Speed. The average number of characters per minute was higher with T9 than it was with Touch, but no significant difference was found.

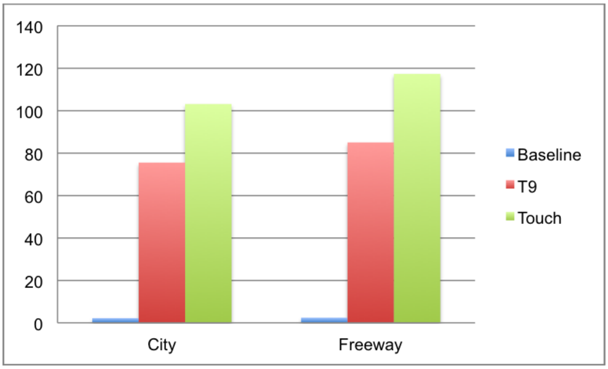


Fig. 9. Average time spent looking away

5.3 Eye Glances

The results on eye glances are shown in table 4. The effect of writing on time spent looking away was significant both on the freeway (F=48.81, p<0.001) and in the city (F=59.69, p<0.001). There was a significant increase in time spent looking away when writing with Touch compared to T9 both in the city (F=5.66, p<0.05) and on the freeway (F=4.87, p<0.05). This was caused by a significant increase in category 3 eye glances both in the city (F=9.92, p<0.01) and on the freeway (F=10.00, p<0.01). There was no significant difference in category 1 and category 2 eye glances between Touch and T9.

Table 4. Means and standard deviations of eye glances

	Condition		
Variable	Baseline	T9	Touch
City			
Time spent looking away (s)	2.18 (1.56)	75.52 (27.55)	103.14 (32.75)
Category 1 (< 0.5 seconds)	0.54 (0.66)	1.39 (2.06)	2.15 (1.95)
Category 2 (0.5-2.0 seconds)	2.54 (2.54)	45.31 (13.96)	42.31 (25.36)
Category 3 (> 2.0 seconds)	0.00 (0.00)	9.00 (8.85)	18.69 (8.92)
Freeway			
Time spent looking away (s)	2.40 (2.23)	85.02 (33.57)	117.35 (44.59)
Category 1 (<0.5 seconds)	0.14 (0.36)	0.50 (0.85)	1.07 (1.32)
Category 2 (0.5-2.0 seconds)	2.79 (2.69)	43.57 (11.69)	39.00 (23.90)
Category 3 (>2.0 seconds)	0.00 (0.00)	12.07 (9.30)	24.07 (9.93)

The results showed no significant difference on time spent looking away between driving in the city and the freeway.

5.4 Expert Comparison

A recent study assessed whether frequent texters are less dangerous in the traffic than less frequent texters. Based on their data, they concluded that this is not the case [34].

Our experiment was not designed to study this aspect. However, for all the results presented in the three tables above, we made between-subject ANOVA tests between T9 users in the T9 condition and Touch users in the Touch condition. The analysis showed no major differences compared to the overall results. This suggests that increased experience with a mobile phone does not change the effect on road safety when texting while driving.

6 Discussion

The results on driving performance show a clear change in driving behaviour when writing text messages compared to driving without interacting with a mobile phone. Between the baseline and the two writing conditions, we have significant results showing that the reaction time doubled, risk of crash/near-crash was 4-5 times higher, and the following distance, following distance variability and lane crossings increased in both the city and on the freeway, and so did lane variability on the freeway.

In the overview of related work, we mentioned several studies of mobile use for calling. Our results are consistent with these studies which demonstrate that text messaging is just as dangerous as other uses of a mobile phone.

The combination of reaction time and crash/near-crash risk shows that writing text messages while driving increases the risk of being in a dangerous situation by around 4 times in the city, and around 5 times on the freeway.

For several of our measures, there was a tendency that the smartphone decreased road safety compared to the traditional tactile phone. However, only some of these differences were significant. The main difference for performance was the number of crash or near-crash incidents and the lane crossings which were significant on the freeway. These results show that overall, text messaging with touch-based smartphone is more dangerous than with T9.

This increase in crash and near-crash risk from city to freeway may be caused by the straight and wider road of the freeway, compared to the many curved roads and intersections in the city. Virginia Research [7] found that texting while driving increased the odds of crashes or near crashes by 23 times. We only found an increase of four but this is possibly attributed to the difference in testing environments, as their testing was done in vehicles in real traffic and not in a simulator. Furthermore, our simulator did not have elements such as pedestrians or cyclists, which, if present, could have increased the likelihood of crash and near-crash situations. On the other hand, our users were in the younger end of the scale, and they may be better able to deal with the distraction from the mobile phone.

Our subjects clearly increased their following distance when texting, suggesting that they subconsciously attempted to reduce the likelihood of being involved in a crash. This is consistent with other results [11].

We analysed our videos looking for explanations of this difference between city and freeway. In the city, our drivers seemed more stressed because they had to follow a car that was stopping, starting and turning frequently. They also had to watch traffic lights and cars driving in the opposite direction. This seems to increase their attention on the traffic. On the freeway, there is less activity and they are more relaxed, which makes them pay less attention to the driving. This is also consistent with the results on lane variability. Some of the related studies have results indicating that there are fewer crashes or near-crashes in the city.

The results on task performance point in the same direction. The task load doubled from the baseline to the two texting conditions. Thus the subject perceived it to be much more demanding to write text messages while driving. There was a tendency that T9 was faster on task completion and writing speed compared to the Touch phone, but this difference was not significant.

The results on eye glances show that writing messages causes the driver to look away from the traffic more often and for longer periods of time. The drivers spent 30-50 times more time looking away when they were writing text messages. The number of eye glances, where the driver looks away from the traffic, increased significantly. In addition, the increase was on the long (between a half and two seconds) and very long (over 2 seconds) eye glances. Research on road safety has shown that these longer eye glances are much more dangerous than the shorter ones.

When we compare the two mobile telephones, the users with the Touch phone spent significantly more time looking away, and they had the double number of very long eye glances compared to the users of the T9 phone.

The test subjects lane variability, crash/near-crash ratio and time spent in the wrong lane as well as time spent looking away was increased when using a smartphone. This suggests that the lack of tactile feedback on the phones caused the subject to look away for longer periods of time, which is why only the number of category 3 eye glances (above 2 seconds) was significantly higher when using a smartphone. These results confirm our original expectation that the touch-based smartphone requires more visual attention.

Overall, the results confirm that text messaging during simulated driving has significant impact on road safety. On some parameters, there was a tendency that texting on the smartphone reduced road safety even more than the traditional phone. On the more severe parameters, there were significant results showing that the touch phone was more dangerous than the traditional tactile mobile phone.

7 Conclusion

In this paper we presented results from an empirical study of the effect of text messaging on road safety where we specifically compared the use of a traditional mobile phone and a smartphone for writing text messages during simulated driving. The results clearly show that writing text messages using any of the two mobile phones significantly decrease road safety. This is caused by significant changes in driving behaviour related to reaction time and lane keeping. Writing text messages while driving increased the risk of being involved in a crash or near-crash situation by a factor of 4 to 5.

The results suggest that while the touch-based smartphone is a step forward in many respects, it decreases road safety when used for writing text messages while driving. The subjects spent significantly more time looking away when they used the smartphone, and this increased the risk of crash or near-crash incidents. Texting while driving is increasing rapidly, and our results show that when a smartphone is used for this, it increases driver distraction as this type of phone provide no tactile feedback, leaving only visual feedback when writing text messages. Our results suggest that drivers should never text while driving, especially not with a smartphone.

There are some limitations in our study. The experiment was conducted in a simulator and not in real traffic. However, our results point in the same direction as studies of real traffic, and comparisons of car simulators and real traffic confirm the relevance of the results from simulators. Our test subjects were relatively young, and, thereby, more experienced in texting. This could imply that they are less dangerous; on the other hand, young people are more likely to be involved in accidents.

For a future study, it could be interesting to identify the factors of mobile phone interaction while driving that affect distraction. Related to this, it would be interesting to further identify the exact points where touch-based smartphones differ from traditional tactile mobile phones. This could include studies with different keyboards and types of feedback on the smartphone.

Acknowledgements. The research behind this paper was partly financed by the Danish Research Councils (grant number 09-065143). We are grateful to the test subjects who participated and the anonymous reviewers.

References

1. Countries that ban mobile phones while driving, `http://www.cellularnews.com/car_bans/`
2. Car crash statistics based on age and location, `http://www.autos.com/drivingandsafety/carcrashstatisticsbasedonageandlocation`
3. Alm, H., Nilsson, L.: Changes in driver behaviour as a function of handsfree mobile phones – A simulator study. Accident Analysis and Prevention 26, 441–451 (1994)
4. Alm, H., Nilsson, L.: The effects of a mobile telephone task on driver behaviour in a car following situation. Accident Analysis and Prevention 27, 707–715 (1995)
5. Arif, A.S., Lopez, M.H., Stuerzlinger, W.: Two new mobile touchscreen text entry techniques. In: 36th Graphics Interface Conference (2010)
6. Battestini, A., Setlur, V., Sohn, T.: A large scale study of text messaging use. In: MobileHCI (2010)
7. Box, S.: New data from VTTI provides insight into mobile phone use and driving distraction. Driving Transportation with Technology (2009)
8. Brookhuis, K.A., De Vries, G., De Waard, D.: The effects of mobile telephoning on driving performance. Accident Analysis and Prevention 23, 309–316 (1991)

9. Charlton, S.G.: Driving while conversing: Mobile phones that distract and passengers who react. Accident Analysis and Prevention 41, 160–173 (2009)
10. Company, T.N.: How teens use media. A Nielsen report on the myths and realities of teen media trends (2009)
11. Curran, K., Woods, D.: Investigating text input methods for mobile phones. Telemat. Inf. 23(1), 1–21 (2006)
12. Drews, F.A., Yazdani, H., Godfrey, C.N., Cooper, J.M., Strayer, D.L.: Text messaging during simulated driving. Human Factors (2009)
13. Entner, R.: Smartphones to overtake feature phones in u.s. by 2011. The Nielsen Company (2010)
14. Favalora, L.: Texting and driving kills. Dr. Frank Ryan's death while texting and driving is a warning. Global Press Release Distribution (2010)
15. Foundation, A.: Text messaging and mobile phone use while driving. Safety Culture (2009)
16. Gartner. Mobile Phone Sales Q3 (2010), `http://www.gartner.com/it/page.jsp?id=1466313`
17. Gilbert, J.E., Martin, A.M., Eugene, W., Alnizami, H., Moses, W., Morrison, D.: Driving transportation policy through technological innovation. Brave New World (2010)
18. Hart, S.G.: Nasa-task load index (nasa-tlx); 20 years later. NASA-Ames Research (2006)
19. Hosking, S.G., Young, K.L.: The effects of text messaging on young drivers. Human Factors (2009)
20. James, C.L., Reischel, K.M.: Text input for mobile devices: comparing model prediction to actual perfor-mance. In: Proceedings of CHI 2001, pp. 365–371. ACM (2001)
21. Jensen, B.S., Skov, M.B., Thiruravichandran, N.: Studying driver attention and behaviour for three configurations of GPS navigation in real traffic driving. In: Proceedings of CHI 2010, pp. 1271–1280. ACM (2010)
22. Joergensen, M.: Mobiltelefoner og bilkoersel. Trafik og Veje (2010)
23. Kircher, A., Vogel, K., Tornros, J., Bolling, A., Patten, N.L.C., Malmstrom, T., Ceci, R.: Mobile telephone simulator study. Swedish National Road and Transport Research Institute. Swedish National Road and Transport Research Institute (2004)
24. Klauer, S., Dingus, T.A., Neale, V.L., Sudweeks, J., Ramsey, D.: The impact of driver inattention on near-crash/crash risk. Technical report, U.S. Department of Transportation (2006)
25. Kristiansen, A.H., Larsen, F.: QdQ: A new keyboard for writing text on small soft keyboards. Master Thesis, Aalborg University (2009)
26. Lee, J.D., Strayer, D.L.: Preface to the special section of driver distraction. Human Factors (2004)
27. McEvoy, S., Stevenson, M.R., McCartt, A.T., Woodward, M., Haworth, C., Palamara, P., Cercarelli, R.: Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crosover study. BMJ 331 (2005)
28. McKnight, A.J., McKnight, A.S.: The effect of cellular phone use upon driver attention. Accident Analysis and Prevention 25, 259–265 (1993)
29. McPhee, L.C., Scialfa, C.T., Dennis, W.M., Ho, G., Caird, J.K.: Age differences in visual search for traffic signs during a simulated conversation. Human Factors (2004)
30. Olswang. Convergence Survey (2011), `http://www.olswang.com/convergence2011/`
31. Redelmeier, D.A., Tibshirani, R.J.: Association between cellular-telephone calls and motor vehicle collisions. New England Journal of Medicine 336, 453–458 (1997)

32. Reed, M.P., Green, P.A.: Comparison of driving performance on-road and in a lowcost simulator using a concurrent telephone dialing task. Ergonomics 42, 1015–1037 (1999)
33. Salvucci, D.D., Markley, D., Zuber, M., Brumby, D.P.: iPod Distraction: Effects of Portable Music-Player Use on Driver Performance. In: Proceedings of CHI 2007, pp. 243–250. ACM (2007)
34. Samuel, S., Pollatsek, A., Fisher, D.: Texting While Driving: Evaluation of Glance Distributions for Frequent/Infrequent Texters and Keypad/Touchpad Texters. In: Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (2008)
35. Strayer, D.L., Drews, F.A., Crouch, D.J.: Fatal distraction? A comparison of the cell-phone driver and the drunk driver. University of Utah Department of Psychology (2003)
36. Wood, C., Hurwitz, J.: Driver workload management during mobile phone conversations. In: Driving Assessment 2005: Proceedings of the 3rd International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design (2005)

Ageing, Technology Anxiety and Intuitive Use of Complex Interfaces

Raghavendra Reddy Gudur[1], Alethea Blackler[1], Vesna Popovic[1], and Doug Mahar[2]

[1] School of Design, Queensland University of Technology, Australia
[2] School of Psychology and Counselling, Queensland University of Technology, Australia
{r.gudur,a.blackler,v.popovic,d.mahar}@qut.edu.au

Abstract. This paper presents the outcome of a study that investigated the relationships between technology prior experience, self-efficacy, technology anxiety, complexity of interface (nested versus flat) and intuitive use in older people. The findings show that, as expected, older people took less time to complete the task on the interface that used a flat structure when compared to the interface that used a complex nested structure. All age groups also used the flat interface more intuitively. However, contrary to what was hypothesised, older age groups did better under anxious conditions. Interestingly, older participants did not make significantly more errors compared with younger age groups on either interface structures.

Keywords: Prior-experience, Technology anxiety, self-efficacy, Intuitive interaction, Ageing, Complex Interfaces.

1 Introduction

Many older people have difficulties in using contemporary consumer products that have complex interfaces and extensive functionalities. Past decades have seen a substantial increase in the use of technology in all aspects of daily living. Inability to use contemporary technologies such as computers, mobile devices, the Internet, and ever increasing self-care medical devices puts the older population at a disadvantage in terms of their ability to live and function independently [1]. This research hypothesised that designing technological products that are more intuitive for older people to use can solve this problem. An intuitive interface requires minimal learning as it mostly relies on technology prior experience/familiarity of the users for effective interaction [2, 3].

However, gradual shift of consumer products from hardware-based to microprocessor controlled software-based technologies has brought a higher level of abstraction into interaction with products [3, 4]. Older generations, who grew up with relatively older technological paradigms, have been left behind. Although the use of technologies such as computers and the Internet is increasing among older people, an age-based digital divide still exists [1]. Research also suggests that older people encounter more difficulties when using interfaces that use complex (nested) interface

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 564–581, 2013.

structures compared with simple (flat) interface structures [5, 6]. Despite this a gradual move in product interface interaction towards touch-based input systems with small screens has necessitated extensive use of multi-layered interface structures.

Not being pace with contemporary technology or lack of technology prior knowledge could lead to low perceived technology self-efficacy, which in turn has the potential for causing technology anxiety. Both low domain-specific prior knowledge and technology anxiety can impede successful use of complex contemporary technological products[7]. In addition, some research [8] also suggests that stress, anxiety and oppressive environments are not conducive to intuitive thinking. However, what is not well understood is the impact of these conditions on intuitive use of a product. This paper discusses a study that investigated the impact of anxiety, complexity in interface structure (nested versus flat) and age on intuitive interaction.

1.1 Intuitive Interaction

Interaction design professionals often feel that they understand what ‘intuitive interaction’ is; however, very few really define it clearly [9, 10]. In the context of interaction design, Blackler [9] and Hurtienne [11], based on their literature reviews of the nature of intuition, suggest that - intuitive use of product interfaces involves non-conscious use of user’s prior-knowledge related to the product in use. In other words, the user is familiar (based on their earlier encounter with similar products or other relevant experiences) with different features and functions of the product [11, 12]. Intuitive use of an interface can be recognised by the following characteristics: It is fast and effortless, it is generally non-conscious and does not involve conscious reasoning [9].

1.2 Complexity in Contemporary Product Interfaces

There is a gradual shift in most consumer products towards a touch-based interface paradigm. If this trend continues, it will not be long before most of the technology older people encounter in their day-to-day activities will be based on this interface paradigm. On a brighter note, recent research suggests that touch-based products are much easier to learn and older users could successfully use them regardless of their age-related cognitive or physical deficiencies [13-15]. Interestingly, Umemuro’s [16] research on older people’s aptitudes to computer found that the anxiety factor declined significantly in the touch screen condition. However, further research needs to be done in this regard to clearly establish which types of input devices are optimal for older people [17].

On the other hand, this shift to touch screen has also resulted in smaller screen sizes with little or no tactile feedback [18]. In addition, with increased functionality and small screen sizes, it has become necessary either to decrease the size of interface elements so as to fit them on the small screen, or to resort to some amount of nesting of the interface to fit all the functions of the device on the screen. Decreasing the size of text and icons will result in visibility and readability issues, especially for older people. Then again, a nested design, if its structure is too deep or complex, could increase the cognitive load on the users.

1.3 Breadth versus Depth in Interface Design Structure

Fundamental characteristics of menu structures have been well researched over the past couple of decades. In particular, tradeoffs between breadth (the number of options in a level) and depth (the number of levels) in menu structures have been extensively empirically investigated and analysed; for example: [6, 19-23]. Most of these studies agree that broad menu structures result in shorter times and better accuracy.

A flat/broad interface is one that has only one level of menu with all the options arranged in a grid fashion, as shown on the left in Figure 1. A nested/multi-layered interface has more than one level of menu. In a nested or multi-layered interface, menu options for the second level onwards are only displayed when one of the menu options in the first level is activated, as shown on right of Figure 1.

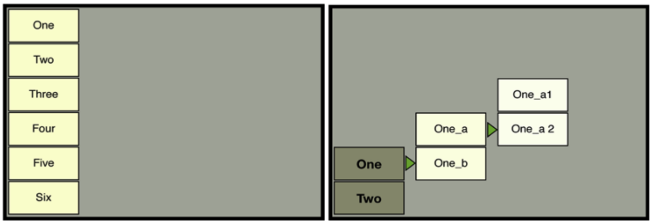


Fig. 1. Broad/flat interface versus nested/multi-layered 2-choices and 3-levels menu system

Menu structure on mobile devices presents a different set of problems. Their screen sizes are often smaller and they usually use touch or pen-based interfaces. A study of preferences for menu structure of mobile devices found that participants preferred a layered (nested) rather than broad menu structure [21]. More specifically, they preferred a structured hierarchy with four to eight items per level.

Research suggests that older people encounter more difficulties when using interfaces that use nested/multi-layered structures compared with flat/broad interface structure [5, 6, 24]. They are also much slower using nested interfaces and tend to prefer shallower interfaces that offer better spatial orientation [19]. A nested interface, by its very nature, offers very few cues to the spatial location of different nested functions. Therefore, these types of interfaces impose a greater demand on the user's working memory. In general, working memory function deteriorates with age [1]. Research suggests that broad or flat structured interfaces rely much less on cognitive abilities and, hence, may be much more beneficial for older users [4, 6].

1.4 Anxiety, Stress and Intuitive Interaction

Although anxiety and stress are known to interfere with intuitive thinking [8], there is no research available that clearly points to the impact of anxiety on intuitive interaction. Blackler's [2, 9] research investigating the hypothesis that intuitive use of products is based on prior experience with similar products did attempt to use the presence of anxiety in the participants as an indication of non-intuitive use. However, it was noted in the study that the psychophysiological tools used to measure heart rate and electrodermal activity (to indicate when participants were anxious) did not provide useful data due to latency issues. On the other hand, it was noted that there were occasions when, towards the end of the session, it was observed that the participants were making mistakes with features they had used correctly earlier. While causes of these moments were not ascertained, it was noted that they could be a result of anxiety or fatigue interfering with intuition [2, 9]. Overall, the outcomes of these experiments were inconclusive in terms of establishing a clear relationship between anxiety and intuitive use.

2 Method

2.1 Experiment Design

The objective of this experiment was to investigate the relationship between age, anxiety, and intuitive use of complex interface structure (nested versus flat). The following hypotheses were framed to investigation this relationship.

1. Nested interface structure will have an adverse effect on time to complete the task, intuitive uses and errors on older when compared with younger participants.
2. Participants who score low on the Technology Prior Experience Questionnaire will also score low on the Self-efficacy Questionnaire and report high anxiety on the State-Trait Anxiety Inventory (STAI) questionnaire.
3. Anxiety, induced by stressful conditions, will have an adverse impact on the time to complete the task, intuitive uses and errors in both younger and older participants.

Table 1. Experiment design and Independent Variables

Independent variables	Levels of Independent Variables	
Interface type	Repeated measures (interface structure type)	
	Flat interface	Nested interface
Age	Young (17 to 34 years)	
	Older young (35 to 49 years)	
	Younger old (50 to 64 years)	
	Old (65 to 72 years)	
	Older old (73+ years)	
Anxiety	Low stress condition	
	High stress condition	

This experiment used a mixed-factorial design (Table 1). The Dependent Variables (DVs) for this experiment were *time on task*, *percentage of errors* and *percentage of intuitive uses*. Independent Variables (IVs) were, *interface type*, *age* and *stress* with their levels listed in Table 1. *Interface type* -flat or nested – was the repeated factor and *low* or *high* stress conditions were the between subjects factors.

2.2 Participants

Overall 54 participants (29 male, 25 female), in five age groups (17 to 34, n = 12; 35 to 49, n = 10; 50 to 64, n = 12; 65 to 72, n = 10; 73+ Years n = 10), between the ages 18 to 84 years (*M* = 54, *SD* 18) participated in this experiment. They were recruited from different sources to maintain a good sample of the general population. Individuals from organisations (for example, sports clubs, educational institutes, recreational facilities and retirement resorts) were asked if they could volunteer to take part in the study. All the participants were screened for visual acuity (corrected or uncorrected) using a Snellen chart. All participants were given a small gift as a token of appreciation (a $3AU worth scratch lottery ticket or coffee voucher) and were entered into a $200AU shopping voucher draw.

2.3 Apparatus and Measures

Technology Prior Experience Questionnaire
This was a two-part questionnaire to capture participants' exposure to, and knowledge of, technologies related to the mediator product used for this experiment.

Self-efficacy Questionnaire
This questionnaire was based on a well-tested instrument to measure perceived general self-efficacy (GSE) that was suggested by Schwarzer, and Jerusalem [25]. For the specific self-efficacy (SSE), a short questionnaire, as suggested by Cassidy and Eachus [26] was used, SSE was specifically designed to measure perceived self-efficacy on using domain-specific technology.

STAI Questionnaire
The State-Trait Anxiety Inventory (STAI) is a self-evaluation questionnaire administered to measure level of state and trait anxiety [27]. A short form six-question questionnaire was used to measure current state of anxiety [28].

Mediator Product
A touch-based device (iPad) was used as the mediator platform for the experiment. This decision was based on observing the increasing use of touch-based interfaces for consumer products. The mediator interface for this experiment was designed based on research [29, 30] that suggested that there was a significant negative correlation between visuospatial sketchpad capacity and time to complete the task and the number of errors. The number of controls in the mediator product for this experiment was kept at 5, to keep it within the visuospatial sketchpad capacity of an average older participant [31, 32] . Interface structure was kept at an optimal two levels [23]. At the same time, the product interface was designed so that, to avoid floor or ceiling effects, it was not too easy for the younger and not too difficult for the older participants. The design of the interface and the task difficulty was established based on the outcome of two pilot studies. Overall 12 participants participated in the pilot studies in three age groups young (20-39 years), middle-aged (40-60 years) and old (61+).

The tasks for this experiment were designed to emulate a real-life, meaningful situation, as Hawthorn [33] suggests that involving the participants in the task is very important to gain meaningful data for the experiment. The task scenarios were based on the real life situation of pet sitting. Moreover, they provided enough interest for both younger and older participants during the pilot study. The tasks that participants were asked to perform were scripted to make sure all participants went through the same number of steps to complete them successfully.

Figure 2 illustrates the layout of the screen with different elements. Interface for the task was displayed only in black, white and greyscale. This was done to control any colour perception issues that the participants might have.

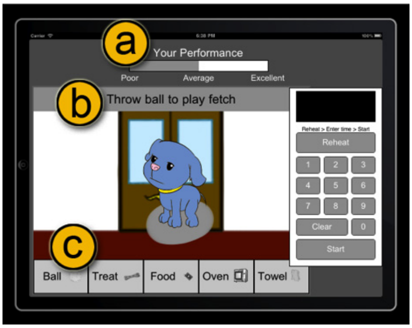


Fig. 2. Screen layout: a) performance feedback bar b) message window and c) interface

Figure 3 illustrates the two repeated measures, flat and nested conditions. Each condition has 5 menu options. In the flat condition, all the five menu items were presented in a row at the bottom; in the nested condition, they were nested under two pop-up menus, one located on the right and the other one on the left bottom corner.

Fig. 3. Flat and nested interface

Figure 4 illustrates the two between subject factors, *low* and *high stress conditions*. The *low stress condition* was induced through positive feedback provided through a bar located on the top centre of the screen. Similarly, *high stress condition* was induced through negative feedback provide through the feedback bar. Feedback was not tied to the actual performance of the participant. Depending on the condition (*low or high stress*) the performance bar changed its status (towards excellent or poor) at certain pre-determined milestones of the task. In the *low stress condition*, the feedback given was always positive, no matter what the actual performance of the participant was. Similarly, on the *high stress condition*, the feedback was negative no matter what the actual performance was.

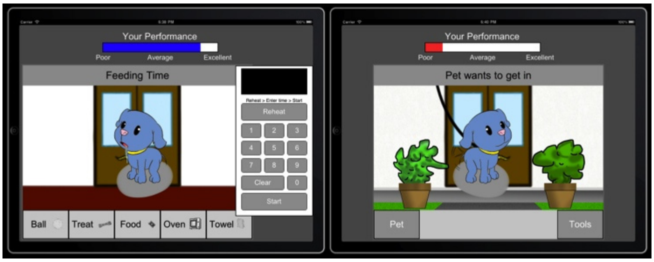


Fig. 4. Positive feedback and negative feedback display bar on the top of the task window

Task Sheet
Two task sheets were prepared that described the two task scenarios, Play and Walk, for the participants. They were asked to read the task scenario before they were asked to start on the tasks.

3 Procedure

The Experiment was conducted in a controlled laboratory setting. The whole experiment was scripted to make sure all the participants followed similar patterns of events, timing, sequence of tasks and instructions from the experimenter. Participants within each age group were randomly assigned to one of the stress conditions, and each completed two task scenarios (Play and Walk), one each on a flat and nested interfaces. To avoid sequencing effects, the tasks were counter-balanced by alternating the sequence of interfaces (flat and nested) and the two task scenarios (Play and Walk) on the device. Two cameras were used to record the experiment for later analysis using the Observer software. One camera was positioned to record participants' facial expressions and body language; the second camera was positioned to record tasks performed by the participant on the screen.

Once a participant was allocated to a group, they were introduced to the lab environment and experimental setup, and were screened for visual acuity using the Snellen chart. They were provided with the information package and consent form. Participants were also informed that they could stop the experiment at any time and could request the deletion of all the record of their participation.

Participants were then provided with the Technology Prior Experience Questionnaire followed by the combined General and Specific Self-efficacy Questionnaire. As not all participants were familiar with the mediator platform (iPad), they were asked to use a simple calculator application on the platform to get a feel for the device. This was also done to ensure that the seating and the position of the device was comfortable for the participants, as it was observed during the pilot study that some of the older participants were bothered by the glare that could occur when the device was placed at certain angles and positions.

Once the participants were comfortable with the setup, they were provided with the STAI Questionnaire and were asked to complete it based on their current state. Then after this they were given the Task 1 scenario sheet, and after making sure they understood what was expected of them, they were asked to start on the task.

After completing Task 1, they were provided with the STAI Questionnaire for the second time to record their current state of anxiety. Once they completed the STAI, they were provided with the Task 2 scenario sheet. After making sure they understood what was expected of them they were asked to start on the Task 2. Soon after completing the Task 2 they were asked to complete the STAI Questionnaire for the third and final time.

After the participants completed the third STAI Questionnaire, they were debriefed. This was especially needed for the participants in the high stress condition, as some tended to get a bit upset with their performance. It was explained to them that the performance bar was rigged and was not a reflection of their actual abilities or performance.

3.1 Data Analysis

The dependent variables *time on task, percentage of intuitive uses, and errors* were central to this experiment. *Time on task* is an important indicator of intuitive use. As 'Intuitive interaction' is defined as fast and generally non-conscious [9]. *Time on task* and *errors* were relatively easy variables to measure. On the other hand, collecting data on actions that are non-conscious was much more challenging. Data on intuitive use was acquired based on observations using Noldus Observer software. Section 3.2 describes heuristics for coding different IVs. The data from the Observer application, the questionnaires and the CogLab application were later exported into statistical analysis software (SPSS) for analysis.

3.2 Coding Heuristics

The coding heuristics were based on earlier studies of intuitive interaction [12]. There are altogether seven codes used for coding the tasks in this experiment. Data from four of these codes are presented in this paper.

Time on Task: This code records a start and finish time and calculates time on task.

Intuitive Use: As intuitive process does not involve conscious reasoning, it is one of the most difficult codes to operationalise [9, 34]. Blackler [12], based on her work on the nature of intuitive interaction, suggests that an intuitive use can be recognised by five indicators: 1) Lack of evidence of conscious reasoning (An intuitive use is often based on very fast decision making with no evidence of reasoning. There is often a lack of verbalisation and, at times, verbalisation follows an action.); 2) Expectation (Participants with very specific prior knowledge about the event are certain about the outcome of their actions.); 3) Subjective certainty of correctness (Confidence of participants executing an event.); 4) Response latency (When a participant executes an event quickly without hesitation.); and, 5) Prior knowledge (When participants indicate their earlier encounter with a similar event). All of this information was extracted from audio video recording of the tasks.

An event was coded as *intuitive use* only when a participant showed two or more of the above indicators [9]. The most certain way of recognising an intuitive use is when an event is executed quickly without hesitation, and when verbalisation follows the action. However, in a study such as this where participants with very diverse sensorimotor and cognitive abilities are involved, it becomes difficult to establish a baseline for both of these indicators. It is well established that ageing slows down motor responses and this slow-down is not linear [35]. In other words, two people might share the same chronological age but may have very different reaction times. These issues were resolved by establishing a baseline response time for each participant. Each participant's observational data was coded multiple times until the differences between *correct non-intuitive use* and *intuitive use* was clearly recognisable. To further minimise coding errors, two independent experienced raters coded the audio-visual recording. An inter-rater reliability analysis using the Kappa statistic was performed to determine consistency of coding *for intuitive use* between the two raters. The inter-rater reliability for coding intuitive uses by the two raters was found to be Kappa = .77 ($p < 001$). Kappa values between 0.60 to 0.79 are considered to indicate substantial agreement between the raters [36].

Correct Non-intuitive Use: Participant completed an event successfully with the use of reasoning or when they had learnt from earlier error. Use of reasoning is indicated by: hesitation in action, latency between the exposure to the event and response, and verbalisation preceding the response.

Error: This code was used when a participant was unable to complete an activity. Indicated by their using a wrong function or overlooking it.

4 Results

The results are organised into two sections: one will look at the effects of different measures used in this experiment; and other presents the relationship between the *type of interface*, *anxiety* and DVs.

4.1 Effects of Different Conditions and Measures Used in the Experiment

Technology Prior Experience

Older people were much more diverse in their capabilities and exposure to the technology. Younger people tended to score much higher on technology prior experience score and were also much more homogenous in their capabilities, 17 to 34 years (M= 46 , SD = 4 , N = 12), 35 to 49 years (M= 43 , SD = 6.3 , N = 10), 50 to 64 years (M= 36.4 , SD = 8 , N = 12), 65 to 72 years (M = 36.6 , SD = 7.2 , N = 10), 73+ years (M = 34.4 , SD = 14.8 , N = 10).

Effect of Stress Condition on Reported State Anxiety

Before proceeding to analyse the rest of the data it was important to check if the method used for inducing stress had the planned effect on the participants. A two-way ANOVA with 5 age groups and two stress conditions as it factors revealed that there was a significant effect of *age* [$F(4,44) = 3.73$, $p = .011$, $\eta_p^2 = .25$] and *stress condition* [$F(1,44) = 6.45$, $p = .015$, $\eta_p^2 = .13$] on the *anxiety* reported on STAI Questionnaire. There was no significant interaction between the *age* and *stress conditions* [$F(4,44) = 0.75$, $p = .561$, $\eta_p^2 = .064$]. A TukeyHSD post-hoc test revealed a significant difference between age groups 35 to 49 and 73+ years ($p = .016$). Except for the 35 to 49 years age group, all other groups reported more *anxiety* on *high stress condition* compared with *low stress condition*. Overall, the method used for inducing stress appears to have worked to a large extent.

Perceived Self-efficacy, Technology Prior Experience and Anxiety

As anticipated, participants with low technology prior experience (TP) had reported low perceived specific self-efficacy (SSE) and high anxiety on the STAI. Although there was a significant correlation between perceived general self-efficacy (GSE) and technology prior experience (TP), [r (52) = .453, $p = .001$] between GSE and SSE [r (52) = .526, $p < .001$] and between SSE and TP [r (52) = . 650, $p < .001$] there was no significant correlation between the perceived GSE and *anxiety* [r (52) = -.172, $p = .212$] or between TP and *anxiety* [r (52) = -.198, $p = .151$]. On the other hand, SSE had a significant correlation with reported anxiety [r (52) = -.269, $p = .049$].

Overall, TP has a significant positive correlation with both the SSE and GSE and, SSE has a significant negative correlation with the reported anxiety. In other words, participants with more TP reported higher SSE and participants with higher SSE reported lower anxiety on the STAI.

Interface Structure

It is also important to stress that the nested interface requires a higher number of responses ($M = 22$, $SD = 7.3$, N= 54) from participants, when compared to the flat interface ($M = 16$, $SD = 4.5$, N= 54). The nature of the nested interface is such that the controls are accessed through two pop-up menus and the additional action of using this two menu controls increases the number of responses needed to complete a task.

4.2 Age, Type of Interface, Anxiety and DVs

All ANOVA analyses reported in this section are three-way, *interface type* (flat, nested) x *age* (18 to 34 years, 35 to 49 years, 50 to 64 years, 65 to 72 years and 75+ years) x *stress condition* (Low, High), mixed factorial design, with *interface type* (flat and nested) as the repeated measure factor.

Time to Complete the Task

Due to the violation of homogeneity revealed by Levene's test, for flat interface [$F(4,49) = 12.72$, $p < .001$], and nested interface [$F(4,49) = 12.09$, $p < .001$], a strict alpha of .025 was used for this analysis [37]. A three-way mixed ANOVA showed that there was a significant effect of *interface type* on *time to complete the task* [$F(1,44) = 24.53$, $p < .001$, $\eta_p^2 = .36$] (Figure 5). Participants took significantly more time to complete the task on the nested interface (M =134 seconds SD = 71) when compared with the flat interface (M =102 seconds SD = 40).

Age also had a significant effect on *time to complete the tasks*: $F(4,44) = 26.69$, $p < .001$, $\eta_p^2 = .71$. TukeyHSD post-hoc test revealed that the 73+ years age group took significantly more time to complete the task when compared with the four younger age groups ($p < .001$). The age group 65 to 72 years also took significantly more time than the youngest (17 to 34 years) age group ($p = .003$). There were no significant differences between the three younger age groups.

There was a significant *interface type* by *age* interaction: $F(4,44) = 3.63$, $p = .012$, $\eta_p^2 = .25$. This shows that the *time to complete the task* in age groups differed between the flat and nested interfaces. To break down these interactions, contrasts were performed comparing different age groups between nested and flat interfaces. *Type of interface* had a significant effect on the 73+ ($P < .001$) and 65 to 72 year age group (p .013). Both of these age groups took more time to complete the task on the nested interface when compared to the flat: 65 to 72 years (flat M = 106 seconds, SE = 10, nested M = 149 seconds, SE = 15) and 73+ years (flat M = 154 seconds, SE = 10, nested M = 234 seconds, SE = 15).

There was also a significant *interface type* by *stress condition* interaction: $F(1,44) = 5.68$, $p = .021$, $\eta_p^2 = .12$. This indicates that the time to complete the task on both interfaces differed between stress conditions. Contrasts revealed that the time to complete the task on the nested interface differed significantly between High and Low stress conditions ($p = .012$). On the flat interface there was no significant time difference between the Low (M = 101 seconds, SD = 47) and High stress (M = 103 seconds, SD = 33) conditions. Interestingly, on nested interface, participants took significantly less time in High stress condition (M = 120 seconds, SD = 48) when compared with Low stress condition (M = 149 seconds, SE = 86).

There was also a non-significant (due to violation of homogeneity) interaction between *age* and *stress condition:* $F(1,41) = 2.78$, $p = .032$, $\eta_p^2 = .21$. Although the interaction was not significant, it is mentioned here as the effect size is very large. This interaction indicates that the effect of stress differed between the age groups. Figure 5 shows these differences clearly. On both stress conditions (Low Figure 5A, High Figure 5B) all age groups took most time to complete the task on the nested interface. As can be seen, the oldest age groups took a lot more time to complete the tasks on both interface types; also, the time differences increased between the nested and flat interfaces. Differences between the other three age groups were not significant.

Interestingly, in the high stress condition, the younger age group (17 to 34 years) took less time on the nested interface compared to the flat interface under High stress condition. Moreover, older age groups took a lot less time to complete the task on both interfaces under High stress condition compared with Low stress condition.

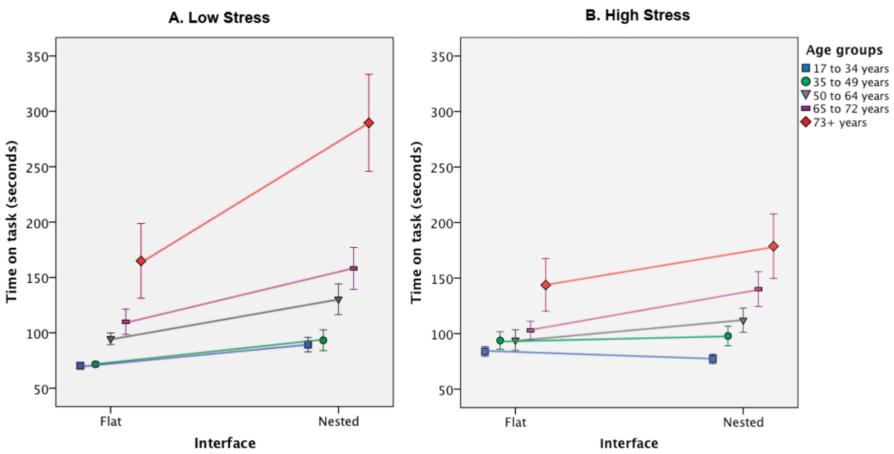


Fig. 5. Time *to complete the task* on flat and nested interfaces under A) Low stress and B) High stress condition

Percentage of Intuitive Uses

A three-way mixed ANOVA with *percentage of intuitive uses* as one of its factors revealed a significant effect of *type of interface* [$F(1,44) = 4.45$, $p = .041$, $\eta_p^2 = .09$] on *percentage of intuitive uses*. This indicated that the participants used flat interface ($M = 17$, $SD = 7$) more intuitively when compared with nested interface ($M = 15$, $SD = 8$).

Age had a significant effect on *percentage of intuitive uses*: $F(4,44) = 5.33$, $p = .001$, $\eta_p^2 = .33$ (Figure 6). TukeyHSD revealed that the age effect was significant between the age groups 17 to 34 and 65 to 72 years ($p = .008$), 35 to 49 and 65 to 72 years ($p = .004$), 35 to 49 and 73+ years ($p = .048$). There was no significant difference among the older three age groups.

There was also a significant three-way interaction between *interface type* by *age* by *stress condition*: $F(4,44) = 2.97$, $p = .029$, $\eta_p^2 = .21$. This indicates that the *interface type* by *age* interaction was different for Low and High stress conditions. Contrasts were performed to reveal the *age* by *interface type* interaction under Low and High stress conditions. These revealed that in Low stress condition (Figure 6A) age had a significant effect on both flat , $F(4,22) = 3.38$, $p = .027$, $\eta_p^2 = .38$, and nested, $F(4,22) = 6.36$, $p = .001$, $\eta_p^2 = .54$, interfaces.

A pairwise comparison using Bonferroni correction revealed a significant difference between the age group 35 to 49 and 65 to 72 years on flat interface, $p = .031$. There was a significant difference between the 35 to 49 and 50 to 64 ($p = .013$) age groups, and between the 65 to 72 ($p = .003$), and 73+ ($p = .003$), age groups on nested interface. Basically, the 35 to 49 age group's behaviour was a mirror image of the rest of the age groups (Figure 6A).

In High stress condition (Figure 6B) the effect of *age* was significant for nested interface [$F(4,22) = 3.85$, $p = .016$, $\eta_p^2 = .41$]. A pairwise comparison using Bonferroni correction revealed no difference between age groups on the flat interface

(Figure 6A). However, there were significant differences between the 17 to 34 and 65 to 72 years (p = .025) age groups, and between the 17 to 34 and 73+ (p = .049) years age groups on Nested interface.

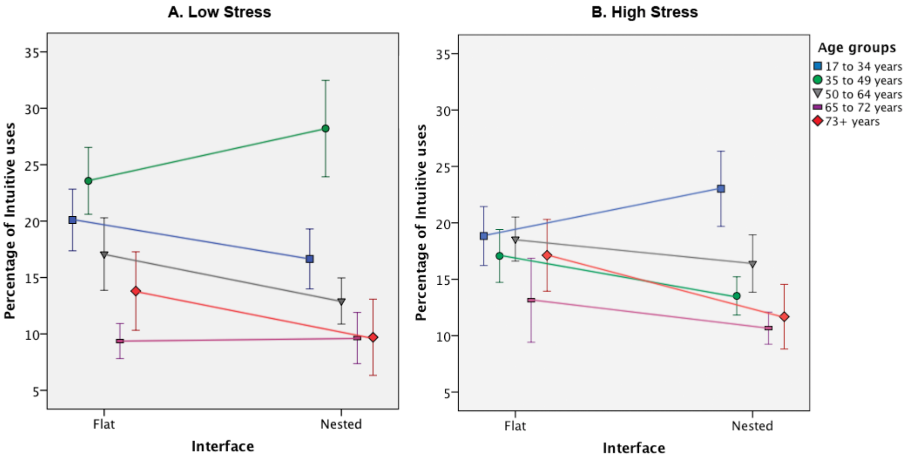


Fig. 6. Percentage *of intuitive uses*, *interface type* and *age* under A) Low stress B) High stress conditions

Interestingly, similar to the results for *time on task*, older age groups did better under stressful conditions. The 50+ age groups used both types of interfaces more intuitively under the High stress condition.

Percentage of Errors

Age had a significant effect on the *percentage of errors*: $F(4,43) = 3.11$, $p = .025$, $\eta_p^2 = .22$. Overall, older age groups made more errors on both types of interfaces when compared with the younger age groups. However, the *type of interface* had no effect on errors made (Figure 7).

There was a significant interaction between *age* and *stress condition*: $F(4,43) = 2.64$, $p = .047$, $\eta_p^2 = .20$. This indicates that the *percentage of errors* made between age groups differed between the Low and High stress conditions. Contrasts revealed that in Low stress condition (Figure 7A), age had a significant effect on use of the flat interface: $F(4,43) = 4.94$, $p = .002$, $\eta_p^2 = .31$. A pairwise comparison using Bonferroni correction revealed that significant differences were only observed between the 35 to 49 and 65 to 72 years age groups with the flat interface ($p = .001$).

Under Low stress condition on the flat interface, the 65 to 72 years (M = 17, SD = 7) age group made substantially more errors compared to the 35 to 49 year (M = 3, SD = 2) age group. Similar to other DVs, older age groups appear to have done better under High stress condition when compared with Low stress condition.

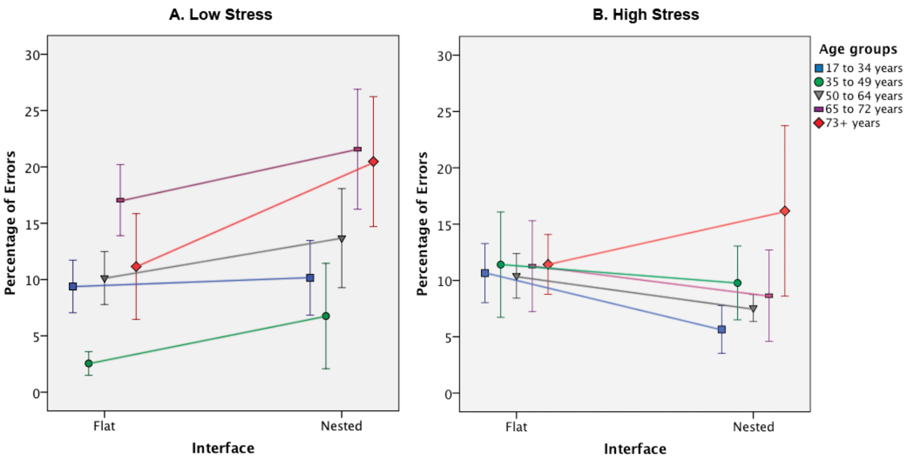


Fig. 7. Percentage *of errors*, *interface type* and *age* under A) Low stress and B) High stress conditions

5 Discussion

This experiment was designed to investigate the relationships between *age*, *anxiety*, and *intuitive use* of complex interface structure (nested versus flat).

As hypothesised, people who scored low on Technology prior experience (TP) and Specific Self-efficacy (SSE) reported higher anxiety. This supports the research that suggests a relationship between age, cognitive decline, technology adoptions, self-efficacy and technology anxiety [38, 39]. Similarly, all age groups took significantly more time to complete the tasks on the nested interface when compared with the flat interface, probably because the nested interface needed more actions to complete the task. However, these differences between nested and flat interface use increased with age. Age also had a significant effect on time to complete the tasks on both types of interface. All age groups used the flat interface more intuitively compared with the nested interface. This finding supports existing data that suggests that older people find nested interfaces more difficult to use [5, 6].

Contrary to what was hypothesised, although older age groups made more errors, there were no significant differences in error rates between use of the nested and flat interface types. Similarly, anxiety had different effects on the younger and older age groups. Older people (65+) reported more anxiety than both the younger age groups. This supports Eisma et al.'s [7] research which suggests that older people may experience more anxiety when they interact with new technologies. However, surprisingly, the performance of the two older age groups was better in the high stress condition. These groups completed the tasks faster and used the interfaces more intuitively under the high stress condition.

This supports Attentional control theory [40], which suggests that anxiety affects processing efficiency more than performance effectiveness. In other words, highly anxious individuals, under stressful conditions, trade time for accuracy in achieving their goal. They also use increased effort and more working memory resources. However, if the task does not overwhelm available resources, increased effort on the working memory resources actually enhances the performance. Since the experiment tasks were designed to be only moderately difficult, high stress condition did not induce high enough levels of anxiety to have a catastrophic effect on the older age groups' performance. This could be explained by the inverted U-hypothesis of anxiety-performance, which states that, for any given task optimal performance is achieved at some intermediate level (the peak of the inverted U) of arousal. Performance starts deteriorating as the level of arousal increases or decreases from its optimal level. This could be the most probable reason behind the performance increase in the older age groups under high stress condition.

The behaviour of the 35 to 49 years age group under stressful condition was different to that of the other age groups. Unlike other age groups, they reported low anxiety on high stress condition and high anxiety on low stress condition. However, while they performed better in a high anxiety state as older age groups did, their performance was opposite to that of the older age groups since they experience high anxiety under low stress condition. This behaviour did pose a few challenges in interpreting the results. For example, ANOVA showed that age had a significant effect on number of errors. However, contrasts showed that the difference was significant only between the 35 to 49 and 65 to 72 year age groups. The probable cause of this difference was that the 35 to 49 years age group made a lot fewer errors on low stress condition, whereas the other age groups did not differ that much. One of the possible reasons for the peculiar behaviour could be, as noted by Kosnik, Winslow [41] study of the perception of vision related problems through adulthood, middle-aged people are more concerned about age-related changes that start becoming noticeable at this age. This behaviour of the middle age group needs further investigation.

The findings from this experiment also support Processing-speed Theory [42], which suggests that older people tend to trade speed for accuracy. Although older people took more time to complete the tasks compared to younger people, overall they did not make significantly more errors than younger groups on both types of interfaces. However, as discussed earlier, middle-age groups differed significantly from the 65 to 72 year age group.

Overall, older people scored lower on technology prior experience, took more time to complete the tasks and used the interfaces less intuitively. However, the number of errors is one of the most crucial indicators for successful use of a product interface. This data suggests that when the interface is designed with consideration of the cognitive limitations of older people, the differences in its use among age groups can be minimal. Apart from the oldest age group (73+) the differences in terms of intuitive use of the interface were not significant. This supports research which suggests that working memory deficiencies in ageing are mediated by coping mechanisms adopted by older individuals, especially when the task is simple [43]. Surprisingly, although most of the older participants had never used an iPad (mediator product for this experiment) they were at ease in using it. This supports recent research on touch-based products that has made similar observation [13, 15].

6 Conclusion

This experiment was designed to investigate if complexity of interface, in terms of its structure, has any impact on older users under high and low Stress conditions. The tasks used in this experiment were designed to consider older people's cognitive limitations. The outcome showed that, as expected, older people took more time on the nested interface compared to the flat interface. However, the type of interface structure had no significant effect on errors made. In addition, the age differences in terms of errors made were also minimal, except for the oldest age group (73+), who were significantly slower and used the interfaces less intuitively. Furthermore, contrary to what was hypothesised, older age groups, although they reported higher anxiety levels, did better under high stress conditions.

This research is significant as the findings from this study demonstrated that, age differences are minimal when complex interfaces are designed to accommodate age related cognitive limitations (Section 2.3), Technology anxiety played positive role in use of interfaces in older people and, older people are at ease in using products with touch based interface. Based on these findings and related research, we are currently developing an adaptable interface framework for inclusive design that takes advantage of flexibility afforded by touch-based interfaces.

References

1. Czaja, S.J., Lee, C.: The impact of aging on access to technology. Universal Access in the Information Society 5(4), 341–349 (2007)
2. Blackler, A.: Intuitive interaction with complex artefacts. In: School of Design 2006. Queensland university of Technology, Brisbane (2006)
3. Hurtienne, J., Blessing, L.: Design for intuitive use - Testing image schema theory for user interface design. In: International Conference on Engineering Design, ICED 2007, Paris (2007)
4. Docampo Rama, M., Ridder, H.D., Bouma, H.: Technology generation and Age in using layered user interfaces. Gerontechnology 1(1), 25–40 (2001)
5. Docampo Rama, M.: Technology generations handling complex user interfaces. In: J. F. Schouten Instituut for User-System Interaction Research 2001, pp. 1–134. Technische Universiteit Eindhoven, Eindhoven (2001)
6. Detweiler, M.G., Hess, S.M., Ellis, R.D.: The effects of display layout on keeping track of visuo-spatial information. In: Rogers, W.A., Fisk, A.D., Walker, N. (eds.) Aging and Skilled Performance: Advances in Theory and Applications, pp. 157–184. Lawrence Erlbaum Associates, Mahwah (1996)
7. Eisma, R., et al.: Mutual inspiration in the development of new technology for older people. In: Proceedings of Include 2003 (2003)
8. Bastick, T.: Intuition: evaluating the construct and Its impact on creative thinking 2003, xxxvi, p. 494. Stoneman & Lang., Kingston (2003)
9. Blackler, A.: Intuitive Interaction with Complex Artefacts: Empirically-based research 2008, p. 324. VDM Verlag Dr. MÜller, SaarbrÜcken (2008)
10. Turner, P.: Towards an account of intuitiveness. Behaviour & Information Technology 27(6), 475–482 (2008)

11. Hurtienne, J., Weber, K., Blessing, L.: Prior Experience and Intuitive Use: Image Schemas in User Centred Design. In: Langdon, P., Clarkson, P.J., Robinson, P. (eds.) Designing Inclusive Futures, pp. 107–116. Springer, London (2008)
12. Blackler, A., Popovic, V., Mahar, D.: Investigating users' intuitive interaction with complex artefacts. Applied Ergonomics 41(1), 72–92 (2010)
13. Häikiö, J., et al.: Touch-based user interface for elderly users. In: Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 289–296. ACM, Singapore (2007)
14. Taveira, A.D., Choi, S.D.: Review Study of Computer Input Devices and Older Users. International Journal of Human-Computer Interaction 25(5), 455–474 (2009)
15. Isomursu, M., et al.: Experiences from a Touch-Based Interaction and Digitally Enhanced Meal-Delivery Service for the Elderly. In: Advances in Human-Computer Interaction 2008 (2008)
16. Umemuro, H.: Lowering elderly Japanese users' resistance towards computers by using touchscreen technology. Universal Access in the Information Society 3(3-4), 276–288 (2004)
17. Czaja, S.J., Lee, C.C.: Information Technology and Older Adults. In: Sears, A., Jacko, J.A. (eds.) Human Computer Interaction: Designing for Diverse Users and Domains. Taylor & Francis Group (2009)
18. Ziefle, M.: Information presentation in small screen devices: The trade-off between visual density and menu foresight. Applied Ergonomics 41(6), 719–730 (2010)
19. Zaphiris, P., Kurniawan, S.H., Ellis, R.D.: Age related differences and the depth vs. breadth tradeoff in hierarchical online information systems. In: Carbonell, N., Stephanidis, C. (eds.) UI4ALL 2002. LNCS, vol. 2615, pp. 23–42. Springer, Heidelberg (2003)
20. Kiger, J.I.: The depth/breadth trade-off in the design of menu-driven user interfaces. Int. J. Man-Mach. Stud. 20(2), 201–213 (1984)
21. Geven, A., Sefelin, R., Tscheligi, M.: Depth and breadth away from the desktop: the optimal information hierarchy for mobile use. In: Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services 2006, pp. 157–164. ACM, Helsinki (2006)
22. Landauer, T.K., Nachbar, D.W.: Selection from alphabetic and numeric menu trees using a touch screen: breadth, depth, and width. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1985, pp. 73–78. ACM, San Francisco (1985)
23. Miller, D.P.: Depth/breadth tradeoff in hierarchical computer menus. In: 25th Annual Meeting of the Human Factors Society (1981)
24. Lim, C.S.C.: Designing inclusive ICT products for older users: taking into account the technology generation effect. Journal of Engineering Design 21(2-3), 189–206 (2009)
25. Schwarzer, R., Jerusalem, M.: Generalized Self-Efficacy scale. In: Weinman, J., Wright, S., Johnston, M. (eds.) Measures in Health Psychology: A User's Portfolio, Causal and Control Beliefs, pp. 35–37. NFER-NELSON, Windsor (1995)
26. Cassidy, S., Eachus, P.: Developing the Computer User Self-Efficacy (CUSE) Scale: Investigating the Relationship Between Computer Self-Efficacy, Gender, and Experience with Computers. Journal of Educational Computing Research 26(2), 133–153 (2002)
27. Spielberger, C.D.: Understanding stress and anxiety. Life cycle series. Thomas Nelson Australia, West Melbourne (1979)
28. Marteau, T.M., Bekker, H.: The development of a six-item short-form of the state scale of the Spielberger State- Trait Anxiety Inventory (STAI). British Journal of Clinical Psychology 31(3), 301–306 (1992)

29. Reddy, G.R., et al.: Ageing and use of complex product interfaces. In: IASDR 2011 Diversity and Unity. TUDelft, Netherlands (2011)
30. Reddy, G.R., et al.: The effects of cognitive ageing on use of complex interfaces. In: OzCHI. ACM, Brisbane (2010)
31. Charness, N.: Visual Short-term Memory and Aging in Chess Players. Journal of Gerontology 36(5), 615–619 (1981)
32. Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 24(1), 87–114 (2001)
33. Hawthorn, D.: Interface design and engagement with older people. Behaviour & Information Technology 26(4), 333–341 (2007)
34. Hurtienne, J.: Image schemas and Design for intuitive use. In: Fakultät, V. (ed.) Fakultät V - Verkehrs- und Maschinensysteme, Technischen Universität, Berlin (2009)
35. Fisk, A.D., et al.: Designing for Older Adults: Principles and Creative Human Factors Approaches. CRC Press, Hoboken (2009)
36. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics 33, 159–174 (1977)
37. Keppel, G., Wickens, T.D.: Design and Analysis: a Researcher's Handbook, 4th edn. Pearson Education, Inc., New Jersey (2004)
38. Czaja, S.J., et al.: Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychology and Aging 21(2), 333–352 (2006)
39. Bandura, A., Freeman, W., Lightsey, R.: Self-Efficacy: The Exercise of Control. Journal of Cognitive Psychotherapy 13(2), 158–166 (1999)
40. Eysenck, M.W., et al.: Anxiety and cognitive performance: Attentional control theory. Emotion 7(2), 336–353 (2007)
41. Kosnik, W., et al.: Visual changes in daily life throughout adulthood. Journal of Gerontology 43(3), P63 (1988)
42. Salthouse, T.A.: Major Issues in Cognitive Aging 2010. Oxford University Press, Oxford (2010)
43. Brébion, G., Smith, M., Ehrlich, M.: Working memory and aging: Deficit or strategy differences? Aging, Neuropsychology, and Cognition 4(1), 58–73 (1997)

Emerging Technologies and the Contextual and Contingent Experiences of Ageing Well

Toni Robertson[1], Jeannette Durick[1], Margot Brereton[2], Kate Vaisutis[2], Frank Vetere[3], Bjorn Nansen[3], and Steve Howard[3]

[1] Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
[2] Science and Engineering Faculty, Queensland University of Technology, Australia
[3] Computing and Information Systems, The University of Melbourne, Australia
{Toni.Robertson,Jeannette.Durick}@uts.edu.au,
{M.Brereton,Kathleen.Vaisutis}@qut.edu.au,
{F.Vetere,hansenb,showard}@unimelb.edu.au

Abstract. Based on a series of interviews of Australians between the ages of 55 and 75 this paper explores the relations between our participants' attitudes towards and use of communication, social and tangible technologies and three relevant themes from our data: staying active, friends and families, and cultural selves. While common across our participants' experiences of ageing, these themes were notable for the diverse ways they were experienced and expressed within individual lives and for the different roles technology was used for within each. A brief discussion of how the diversity of our ageing population implicates the design of emerging technologies ends the paper.

Keywords: Ageing population, ageing well, social technologies, tangible technologies, diversity.

1 Introduction

We are engaged in a project that investigates sociophysical interaction conceptualised as the entanglement of tangible and social interaction. With commitments to phenomenological and participatory approaches, as well as the recognition that all human action is both embodied and situated, our project is grounded by the exploration of the roles that emerging social and tangible information and communication technologies can play in ageing well. This context provides the actual living bodies, objects and physical environments that such research depends on. It also responds to its pressing significance to many developed societies where governments and other agencies are seeking to evolve policies and strategies to provide appropriate, adequate and cost-effective services to their ageing citizens within limited and tightening budgets. Design and use of new and emerging technologies can contribute to the alignment of the desires of ageing people to remain active and in control of their lives with the goals of governments to encourage healthy, active and financially independent senior citizens and so minimise their reliance on expensive government services (eg Andrews 2002).

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 582–589, 2013.

We have begun to explore this space where actual ageing bodies meet with new social and tangible technologies, as well as with still-developing social policies, through an initial series of scoping interviews with ageing people. This paper is not intended as a comprehensive discussion on the findings of those interviews (see Robertson et al. 2012 for a preliminary discussion of findings), nor is it an attempt to systematically challenge the myths and stereotypes that exist about older people. Instead, our interest is to explore the major issue that 'shouted' from our data: the diversity of our participants, which existed despite the limited demographic and the assumed homogeneity of ageing people that we encountered both in prevailing social attitudes and much of the research on ageing and gerontechnology (see Lindley, Harper & Sellen 2008 for a literature review of this topic).

We begin by briefly describing the series of interviews before teasing out the areas of diversity and related richness that were both shared and defining of the lives of our participants. From there we explore the highly contextual and contingent ways some common aspects of ageing were experienced and expressed in the day-to-day lives of our individual participants and the role new and emerging technologies were playing in these experiences. We end the paper with a brief discussion of the implications such diversity has on the design of emerging technologies.

2 The Interviews

Scoping interviews are designed as a small, open corpus of loosely-structured interviews that are used to evaluate, ground and refine the initial understandings, assumptions and concepts of a research team in projects such as this one where the development of ethnographic understandings of the domain of intended use are an important part of the process (Robertson et al. 2012). Both their design and analysis were driven by the needs of the project as design research (Cross, 2007). This means that our concerns are less about exhaustive analysis of what currently exists and more about checking our assumptions, focusing the research team with shared understandings of what ageing well meant to those who were actually doing the ageing and so developing an informed orientation to the aims of the research project. The interviews were intended to provide a ground from which to investigate what might be technologies of value in the future and how we might initiate their design and uptake. They were designed to solicit both demographic and experiential data including fitness and activity levels, community and social engagement, use of and attitudes to various kinds of information and communication technologies and applications, future work and retirement plans, and current practices around keeping in touch with friends and families. Each participant was also asked to describe what ageing well meant to them and to add anything else about ageing that they felt relevant to our project.

Twenty interviews from four groups of ageing people have been completed, transcribed and analysed. We interviewed five members of each group. The groups were chosen to represent an aspect of ageing of particular relevance to the overall project: maintaining health and mobility, community engagement, and social and family relationships. We sought people aged between 55 and 75 who were still very active, including some still in full time employment, some working part time and some who were retired from paid employment but engaged in community activities of various kinds. Participant contexts included patients in a community rehabilitation hospital in Sydney who had joint replacement surgery 10-14 days prior to their interview; members of an ethnic community organisation in the Illawarra district of NSW; members of large national and voluntary community service organisations in Brisbane and a group who, for various reasons, were childless. Men and women participated equally in the interviews and the group included people from cities, regional and rural areas of Eastern Australia.

The interviews were completed and transcribed in batches with initial analysis of the first ten used to generate areas of attention that could be further addressed in the remainder. The analysis of the full series included combining and comparing different participants' contributions to each area of questioning as well as identifying both commonalities and areas of difference that might be more thoroughly explored by field study and iterative prototyping in the remainder of the project. In our analysis we were mindful of Star's recommendations to listen for the special language used by the participants and for anything that appeared 'strange, weird and anomalous' (Star 2010, p. 605). The first recommendation maps closely to the central HCI design principle expressed by variations of ‘Know the user population’, while the 'strange, weird and anomalous' can help us develop insights into the diversity of our participants, the richness of their experience and relationships, and how these might refine our focus on the roles social and tangible technologies might play in ageing well.

3 The Diverse Experiences of Ageing and Technology Use

Summarised very broadly, our participants’ views of what ageing well meant to them were remarkably similar. Good health, with a particular emphasis on mobility and fitness; social agency in their interactions with family and friends; independence—including both physical and financial independence in their daily lives; and some kind of ongoing work and/or community involvement were repeatedly identified as central to the maintenance of current and future well-being. Our participants all expressed their desire to *maintain* for as long as possible, their capacities to actively experience the areas of their lives that defined their shared sense of ageing well.

However, while our participants' views of ageing well were very consistent, their actual ageing bodies were not. Their bodies were the bodies of people who had led long lives. They were old enough for their various individual genetic, cultural and environmental influences to have expressed themselves, along with the inscriptions made on their bodies by their actual life experiences, such as the work they did, how well they had 'looked after themselves', their access to health care, good food and a safe environment, the communities in which they lived and the kinds of social activities, cultural mores and expectations these provided, as well as various random life events and accidents. So what it would mean and require for each participant to maintain good health, mobility and fitness varied enormously in practice. While the desire to remain active was also common, the activities our participants wanted to continue were similarly varied. And while all our participants wanted to maintain their social agency and ongoing relationships with families and friends, their social networks and how they were experienced and managed were very different.

Our participants were already using information and communication technologies and had learned to use them at work or when they had recognised a need or advantage to do so. The use of word processing, email and various web services, such as online banking was common. Given our focus on the roles that social and tangible technologies might play in ageing well, our participants' engagement with and attitudes towards new and emerging technologies were as contextually and contingently defined as other aspects of their ageing selves. In the remainder of this section we explore the relations between our participants' attitudes towards and use of communications, social and tangible technologies and three relevant themes from our data that, while common across our participants, were notable for the diversity of their experience and expression. These themes are: staying active, friends and families, and cultural selves.

3.1 Staying Active

Maintaining an active lifestyle was cited by all participants as being central to ageing well and framed as being part of a cycle of habits and behaviours, including social activities and physical exercise that all impacted one another. Maintaining good health was considered central to keeping mobile and maintaining the social agency and physical independence that each of our participants valued so highly:

- "*your body deteriorates a lot quicker if you don't keep moving*"
- "*poor health stops you living well. It stops your quality of life*".

All our participants recognised that their bodies had changed with age in ways that had affected their mobility and the activities they enjoyed. Several were dealing with recent cancer scares and most had some kind of age-related health condition such as diabetes, arthritis, high blood pressure and general aches and pains. But the health effects of ageing were very specific to each individual; impacted each of them in distinct and diverse ways and each responded to them differently. One participant told of the shift from soccer playing and a life of physically demanding jobs to games of billiards and extensive volunteer work at his local community club. Another longed to enjoy gentle travel and gardening after a working life involving a great deal of hectic, work-related travel and active sports that had led to early joint degeneration. Another still played tennis regularly (but no longer in competition), watched cricket instead of playing it and walked a great deal, including daily walking a dog that was jointly owned with several other similarly-aged friends, who also regularly walked him.

At least half our participants were still working either full time or in various stages of reducing their work hours and moving into full retirement. Those who were fully retired were active in various kinds of community and volunteer work even holding leadership positions in their various community groups. The activities our participants engaged in varied widely and included (among just twenty people): gardening, sport (eg golf, tennis, pistol shooting and bowls); explicit exercise (eg cycling, dancing, hydrotherapy, aerobics, gym programmes, swimming, walking–alone, with others or with pet/s); playing chess, Scrabble, billiards and single-user computer games; actively contributing to community organisations and community projects; travelling (domestically and internationally)—individually, together or in larger organised groups (eg. cruises, bicycle and walking tours); taking a caravan around Australia; studying, both formally and informally; and engaging in a rich and varied range of social interactions with family and friends. The common element among all participants and their chosen activities was their desire to maintain their capacity to continue them.

Only one of our participants reported using any kind of technology as part of staying physically active —in this case a mobile app was used to record speed, distance, energy consumed etc during daily walks. Most participants had some knowledge, even positive experiences, of console games gained via interactions with their children and/or grandchildren. However, none were interested in acquiring a game console of their own. Most participants explained they were "*too busy*" for games. Several added they were not interested "*I'm not a games person; not even of the board variety*" while others indicated that if relevant they might be interested. For example, one had played online Scrabble with his adult child as an informal way to stay connected while another reflected that he would like "*someone to play a game of chess with*". But tangible technologies involving active bodily movement as input remained largely unfamiliar and unused among our participants at the time of the interviews.

3.2 Friends and Families

Our participants all had families and networks of friends that varied in their conventionality, size, geographic spread, frequency of interaction and emotional connectedness. Most participants had children and grandchildren; some had been divorced, others widowed and some had remarried and/or settled with new partners. Others still had never married yet maintained close, large and supportive networks of friends: "*I've never been a family person. I've always been more into friends*".

Participants noted that it wasn't only their bodies that were affected by the changes ageing brings, but also their relationships with other people. Staying active was central to maintaining these relationships. One participant, a retired semi-professional golfer said: "*Physical activity is my number one priority […] I know that grandchildren will be on the way. I want to be able to physically be with them. I don't want them looking at grandma sitting in a chair and who can't do anything, because I want to enjoy them.*" Another commented: "*whilst we can still interact with the young ones we will. I like to get on the floor with them but I'm not as able as I was two years ago*." Others were concerned about needing to be cared for by friends and families or noted the changing needs of their ageing friends: "*old friends who really do decline or change or have to move or that sort of thing. Dealing with that is something that I was not prepared for*".

Phones, both landline and mobiles were the usual means of keeping in touch with others. Only one participant did not have his own mobile phone but he used his wife's mobile when necessary. Another used his mobile very infrequently, mainly when travelling, but used email intensely for both work and social contact. For others, mobile phones were enmeshed in their everyday lives. SMS messages were popular especially within close relationships: "*I will SMS forever and a day, particularly to the kids*". Email use was frequent and well-established. All participants had heard of Skype with around half of them using it to maintain contact with distant family and friends. One was using it very frequently to maintain contact with a close friend in another country who was very ill. Facebook and even LinkedIn were also used to maintain contact with distant relatives and friends. Another was following a blog written by an old school friend who was travelling around the country in a caravan.

Physical distance between our participants and their geographically dispersed friends and family was reflected in varying ways in the regularity of interaction and the types of technology used. Phones, SMS and email were the first choices for close and regular contact—both for organising social activities and/or for just keeping in touch that is, maintaining relationships— but these were complemented by the use of Skype and Facebook for maintaining contact with those further away and with larger social groups such as wide family relationships and old school friends: "*for staying in touch with people. Even if it's just to see what they're doing, as opposed to actually communicating with them.*" We did not find any examples where our participants had used Facebook for forming new friendships or engaging with apps or marketing.

3.3 Cultural Selves

Ageing people, by definition, have long embodied histories; they have lived though many relationships and many changes over time, not just with people but with places and objects including a great variety of technologies. Their social identities have been produced and shaped by the lives they have led. All our participants had their own very specific histories that shaped how they defined and lived their social roles, their values and the priorities of their current lives. Among our participants there were still differences in how communication technologies were used, particularly among our older participants who had migrated to Australia as young adults. Aspects of the cultural self that we discuss here pertain to cultural origins, life histories, gender roles and generation.

All participants noted the importance of family and friends. Some of the participants, who had migrated from Europe, had large, extended families with many members still living overseas. The group from the Illawarra-based community club were skewed towards the older end of our demographic and shared a common ethnic and cultural identity. Traditional roles of husband and wife seemed stronger within this group though even here there were still some anomalies. For example, in one household, the husband used a variety of communication technologies regularly to keep in touch with family and organise get-togethers. This included using Facebook and videoing family events, editing these and then copying the final product to DVDs that he posted to relatives abroad. On the other hand his wife, who claimed to be uncomfortable with technology, used her mobile phone extensively for immediate and frequent social contact and organisation with close relatives and friends.

> *Well, we've both got a mobile phone, and of course I've got the Internet too. I keep in contact with SMS and emails. Most of our kids work. The way to get to them is to send them a message–SMS 25 cents. So even if they are working it doesn't matter. My wife doesn't use SMS. She rings. She's not a real tech . . ."*

But in another household within the same group, the female participant possessed the technical skills (aided by her granddaughter's guidance) and her husband, whose health was poor, relied on her to organise and maintain frequent contact with their extended family.

Differences in technology use and attitudes to it also reflected the type of work our participants did or had done. For example, one participant, a farmer who also ran a small earth moving business, explained: *"<female partner> does most of my computer work for me. I sit there and dictate and she writes it out. It's a typing issue . . . I find it a bit frustrating to sit down at a computer*". He had always done his paperwork manually but had learned to use internet banking, to search for information online and to read email. He understood and could manage the computers that were part of his earthmoving equipment and had learned to maintain home's Internet connection despite living in an area with frequent lightning strikes. It was his partner who did most of the emailing and used both Skype and Facebook: "*<female partner> uses Facebook, but I would never go on Facebook; I reckon it's a waste of bloody time."* Among the younger participants and the city-dwellers there were no obvious gender differences in either familiarity with or use of the different technologies.

Our participants readily acknowledged generational differences in technology usage within their own families and circles of friends—particularly the use of mobile phones and Facebook. While this was at times accompanied by dismissive comments about Facebook from some of the male participants, for most it was a simple acknowledgement of technological change and generational differences. One participant, after commenting on his grandchildren's addiction to Facebook, added "*if they haven't got a telephone, it's like cutting one of their arms off or something. They're lost*". But later in the interview, commenting on his current use of a mobile phone and his own youth, remarked: "*I don't know how people used to get on then without phones and iPods or whatever they call them*".

4 Discussion

The demographic for our interviews was a relatively narrow one: white, middle-class, between the ages of 55 and 75 and living in the eastern states of Australia. All participants shared a strong sense of what ageing well meant to them, that is, one that privileged maintaining health, mobility and independence, and active engagement with their families, friends and local communities. Yet within this group of people—chosen because they fitted an 'ageing' demographic—some were still working, some were semi-retired and others fully retired; some were very computer literate while others were comfortable only with (mobile) phone use and basic email; all had very different physical and/or personal needs that were being managed in a wide range of ways and their reasons for engaging with technologies, the selection of which technologies to use and what they were actually used for varied widely.

Our participants made choices about the technologies they used based on specific needs, relevance and the availability of supporting infrastructure. They had already appropriated a number of communications technologies into their everyday lives and were in various stages of engaging with newer social technologies. There were different levels of familiarity with some of the newer tangible technologies but little interest, at this stage, in expanding their use of those currently available. Most of all, our interviews provided a snapshot of the ongoing appropriation by our participants of new technologies that were not specifically age-related but part of the general social, technological and physical infrastructure in which ageing people live. And they reminded us that our potential users' familiarity with ICTs is dynamic, diverse and unbounded. Most importantly our findings highlight the diversity of the ageing population and the need for technology compatible with both diverse use and diverse users.

Diversity within our chosen demographic was multi-dimensional and often unpredictable, confirming the need to continue to question common assumptions about both ageing and the use of new and emerging technologies by ageing people. It was clear that our participants lived their lives within the range of norms and expectations available within their culture, even as they insisted on negotiating their own pathways through ageing. In the process they were actively reinterpreting these same cultural norms and expectations. Importantly for our current context, the recognition of the specificity of ageing bodies and the activities that are available to them—including the appropriation of new and emerging technologies—foregrounds the further recognition that ageing people have always been and will always remain embedded within their ever-changing social and cultural worlds. Indeed it is that embeddedness that they are seeking to maintain as they continue to age.

Acknowledgements. This research is funded by the Australia Research Council Discovery Grants Program DP110101999. We are grateful to our participants for their contribution.

References

1. Andrews, K.: National Strategy for an Ageing Australia. Commonwealth of Australia, ACT (2002)
2. Cross, N.: Designerly Ways of Knowing. Birkhäuser Architecture (2007)
3. Hughes, J., King, V., Rodden, T., Andersen, H.: Moving out from the control room: ethnography in system design. In: Proc. CSCW 1994, pp. 429–439. ACM (1994)
4. Lindley, S.E., Harper, R., Sellen, A.: Designing for elders: exploring the complexity of relationships in later life. In: Proc. BCI HCI Conference, vol. 1, pp. 77–86 (2008)
5. Robertson, T., Durick, J., Brereton, M., Vetere, F., Howard, S., Nansen, B.: Knowing Our Users: Scoping Interviews in Design Research with Ageing Participants. In: Proc. OZCHI 2012, CHISIG/ACM, pp. 517–520 (2012)
6. Star, S.L.: This is Not a Boundary Object: Reflections on the origin of a Concept. Science. Technology and Human Values 35(5), 601–617 (2010)

Question-Answer Cards for an Inclusive Micro-tasking Framework for the Elderly

Masatomo Kobayashi, Tatsuya Ishihara, Akihiro Kosugi, Hironobu Takagi, and Chieko Asakawa

IBM Research – Tokyo, 5-6-52 Toyosu, Koto, Tokyo 135-8511, Japan
{mstm,tisihara,a1kosugi,takagih,chie}@jp.ibm.com

Abstract. Micro-tasking (e.g., crowdsourcing) has the potential to help "long-tail" senior workers utilize their knowledge and experience to contribute to their communities. However, their limited ICT skills and their concerns about new technologies can prevent them from participating in emerging work scenarios. We have devised a question-answer card interface to allow the elderly to participate in micro-tasks with minimal ICT skills and learning efforts. Our survey identified a need for skill-based task recommendations, so we also added a probabilistic skill assessment model based on the results of the micro-tasks. We also discuss some scenarios to exploit the question-answer card framework to create new work opportunities for senior citizens. Our experiments showed that untrained seniors performed the micro-tasks effectively with our interface in both controlled and realistic conditions, and the differences in their skills were reliably assessed.

Keywords: Micro-Tasks, Gamification, Skill Assessment, Ageing, Elderly, Senior Workforce.

1 Introduction

In many developed societies, increases in the elderly population and declines in the working-age population are serious social issues. For example, in Japan, which has the oldest population in the world, 24.6% of the population is over 65 as of March 2013 [1]. The retirement of baby-boomers is expected to reduce productivity while increasing the demand for healthcare and economic support. To address these trends, the senior workforce could be more utilized and their participation in work can help them maintain their independent lives. It is known that work benefits people in many ways, for example by reducing the risk of death while improving their well-being [2].

However, the age-related loss of physical, sensory, and cognitive abilities and declining health may prevent senior citizens from participating in full-time jobs. Only a few exceptional seniors are highly active, even though many senior citizens have a desire to work and contribute to their society [3]. As a result, a considerable fraction of the "long-tail" elderly workforce and their valuable knowledge and experience remain unutilized. They need more flexible work opportunities, so they can participate in part-time jobs whenever they want to and at their preferred locations [4].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 590–607, 2013.

The key to success is how to exploit their abilities that have not declined with age (e.g., linguistic knowledge [2]) while avoiding any risks related to declining abilities.

The information-communication technology (ICT) is a key to such flexible work. Online communication tools allow distant collaborations and outsourcing to remote freelancers. Internet-connected mobile devices remove restrictions on workplaces. Among such ICT capabilities, this study particularly focuses on *micro-tasking* as used in crowdsourcing. It is one of the most suitable ways to work with a long-tail workforce, allowing occasional participation in small tasks at home without requirements for continuous commitment. The examples of micro-tasking range from simple, labor-intensive tasks to complicated, intellectual tasks as sampled in Section 2. Statistics show that the number of crowdsourcing workers has grown by more than 100% each year, reaching 6.3 million as of 2011 [5].

However, most senior citizens are currently excluded from micro-tasking. For example, the participants in the Amazon Mechanical Turk (MTurk) are mainly younger workers [6]. A small study we conducted with local senior citizens in conjunction with [7] found that none of them had ever participated in micro-tasking, such as contributing to MTurk, Wikipedia, or Q&A services, even when they had Internet connections. The literature suggests two reasons: limited ICT skills and concerns about new technologies. Although Internet-connected personal computers (PCs) and mobile devices have become increasingly popular even among senior populations, the elderly are often regarded as passive users of ICT, whose uses of technologies tend to be limited and to favor existing functions [8][9][10].

Our work aims to create an encouraging framework to accelerate elderly participation in micro-tasking. The primary target users involve senior citizens who have ICT devices such as smartphones or PCs but who do not actively use such technologies, who appear to be the majority of the senior population, at least over the next decade. First, we conducted a questionnaire-based survey to identify the elderly's requirements for the framework and to investigate their current usage of ICT. Next, based on the survey results, we introduce two technical foci of the framework: an interface design that minimizes the effort in learning; and a skill assessment mechanism that allows skill-based task recommendation. The interface uses the metaphor of stacked question-answer cards to allow senior workers to participate in micro-tasks with guided, simple, and consistent interactions (Fig. 1). The skill assessment uses a probabilistic model to calculate the skill level based on the results of micro-tasks.

Here is the structure of this paper. In the next section, we discuss related work. We then describe the concept of our inclusive micro-tasking framework, followed by a summary of the questionnaires, descriptions of our interface design and skill assessment mechanism, and some practical scenarios for our framework. Next we describe the prototype implementation. In the first experiment that involved senior and young participants, we tested the usability of our interface and the feasibility of skill assessment with touchscreen tablets. The second experiment examined the framework with home PCs. We conclude by discussing the implications of the experiments.

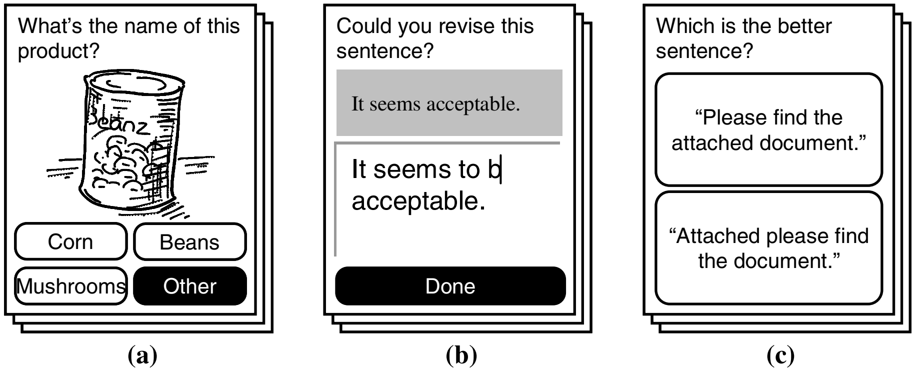


Fig. 1. Examples of micro-task cards, which ask (a) for a product's name (to help a person who is blind), (b) to revise a sentence, and (c) to choose the better sentence

2 Related Work

Micro-Tasking. Crowdsourcing, where workers perform small pieces of work called micro-tasks via the Internet, has been used in many ways such as image recognition to help people who are blind [11], character recognition [12], translation [13], and even user studies [14]. Crowdsourcing is typically regarded as a means to easily access inexpensive labor, but some approaches such as VizWiz [11] have been proposed to exploit social networks for more sophisticated participation. Also, crowdsourcing is useful for enriching the social network information itself. Guy et al. showed crowdsourcing can be used to gather tags for people and social networks [15]. They showed that micro-tasks that ask about social networks can help update obsolete information in social applications.

Crowdsourcing was originally used for solving simple tasks but has expanded into more complex tasks. Kulkarni et al. proposed a framework to decompose complex tasks into micro-tasks [16]. They showed that even the work of decomposing the complex tasks into simple tasks could be crowd-sourced. Soylent decomposes crowd-sourced text-editing into three steps: find, fix, and verify [17]. Noronha et al. showed that nutritional analyses can be crowdsourced by decomposing the tasks into several sub-tasks [18]. We assume that some of decomposed micro-tasks may require simpler ICT skills while still benefiting from the wisdom of age, and are thus especially suitable for senior workers.

Senior-Friendly Systems. What assists and hinders the use of ICT by the elderly? The user interface is a critical factor when designing a system for senior citizens. Leonardi et al. found that seniors preferred intuitive touchscreen interfaces over keyboards [19]. Kobayashi et al. reported that mobile touchscreen interactions are "enjoyable" for seniors and their skills with such interfaces improve rapidly [20]. Those studies motivated us to include touchscreen devices in the primary targets of our framework. The user interface needs of senior citizens have also been discussed in Web accessibility research [21]. There are no standard accessibility guidelines specific to the elderly. However, it is known that seniors have requirements that in many ways are similar to those of people with disabilities, since their physical, sensory, and cognitive abilities tend to decline as they age.

The Office for National Statistics in the U.K. reported that typical senior citizens use their mobile phones only for phone calls and text messaging and avoid using Web browsers and other applications [8]. Kurniawan concluded that senior citizens tend to be fearful of unfamiliar technologies [9]. She also found that seniors are passive users of mobile phones, which means they mostly receive phone calls and text messages rather than originating them. Leung et al. reported that the preference to use trial-and-error decreases with age and senior citizens prefer to have instructions [10]. Hiyama et al. used question-answer interactions through phone calls and text messages as an alternative means for the elderly to participate in social media [7]. Those studies led us to exploit passive interaction styles that can be used without trial-and-error.

Handling Worker Skills. A major problem of crowdsourcing is the difficulty of assuring the quality of the results. Since the tasks are distributed to many people, there are wide variations in workers' skills and risks of vandalism. There are several ways to remove low quality results. CrowdFlower [22], one of the major crowdsourcing services, provides a function to filter noisy workers by distributing some tasks whose ground truths are known. By checking the results of these tasks, the workers who do not have skills can be filtered from the results. Another approach to assure the quality is aggregating multiple workers' output while assessing the skill of each worker and the difficulty of each task at the same time [23]. This approach assumes a correct answer exists for each task and the more skilled workers will agree on the correct answer.

Heimerl et al. showed micro-tasks requiring expertise could be efficiently done by distributing tasks to skilled workers [24]. How to search for the experts is an active area in information retrieval research. For example, Macdonald et al. proposed a method that finds experts by using multiple sources (such as resumes and webpages) to build skill profiles [25]. Guy et al. showed that the content in social applications could help find people related to specific keywords [26]. These projects assumed that the candidates actively used social tools in their daily lives, an assumption that is often strained for senior citizens. As already noted, the elderly tend to be passive users of these technologies. Also, they tend to hesitate in sharing their personal information due to privacy concerns [27]. Thus we decided to devise a new approach to obtain the skill profiles, an approach based on analyzing the results of the micro-tasks.

3 Inclusive Micro-tasking Framework

This section describes our approach for elderly micro-tasking. As noted in Section 2, various types of tasks can be decomposed into micro-tasks and many of them are expected to be suitable for senior workers. For example, among the three steps (find, fix, and verify) in crowdsourced text-editing [17], "verify" seems to be the most suitable task for the elderly. It requires less ICT skills (i.e., simply vote to approve work), but benefits from the wisdom of age (i.e., linguistic knowledge).

In many workplaces it is known that older and younger workers have complementary knowledge, skills, and abilities [28], so younger workers can assist older workers with their ICT skills while the older workers may assist the younger workers with their vocational knowledge. If this type of collaboration is supported in micro-tasking, then elderly participation would be encouraged. Since the people with ICT skills can already use standard desktop interfaces, what we need is an easier gateway for senior citizens who are excluded from micro-tasking opportunities.

3.1 User Requirements Survey

A survey using a paper questionnaire was conducted to identify the elderly's requirements for micro-tasking framework. We gave the questionnaire to 179 senior citizens in a suburban city and 170 of them responded (for a response rate of 95%). The respondents consisted of 99 males and 71 females, ranging in age from 54 to 84 (*mean*=67.7, *SD*=5.3). This survey was conducted as a part of a larger survey regarding ICT and work.

For the usage of ICT, 78% were PC users, with 60% using a PC every day and 88% were mobile phone users, with 61% every-day users. Although most respondents were frequent users of PCs and mobile phones, their usage was limited. The tools they frequently used included word processers, spreadsheets, e-mail, and Web search engines. Meanwhile, they rarely used modern ICT tools such as video chat or a social networking service (SNS). These results indicate they are using specific applications similar to those they used before they retired from work, confirming previous findings that senior citizens tend to avoid active use of new technologies [9]. This suggests that, although the majority of senior citizens now use ICT, we cannot optimistically assume that they will start participating in micro-tasking without additional mechanisms to increase the elderly's easy access to micro-tasking opportunities.

As for requirements for micro-tasking, the majority of the respondents specified: detailed instructions prior to participation (72%) and task recommendations based on their skills and interests (63%). This also confirms previous findings regarding the preference for instructions and the passive usage of ICT [10]. Based on the results, we defined two technical requirements for an inclusive micro-tasking framework. First, the interface should be simple and versatile. The simplicity makes it easy to learn, while the versatility provides consistent interactions for various types of tasks. Second, the system should have skill profiles for task recommendations. As mentioned in Section 2, although skill profiles can be extracted from resumes, webpages, and social tools, these approaches tend not to work well for senior workers. We need a mechanism to assess workers' skills without explicit skill-related sources.

Finally, regarding their interests in micro-tasks, the respondents were particularly interested in the tasks that require linguistic abilities in their native language and tasks that help people with disabilities.

3.2 Question-Answer Cards

To make the interface simple and versatile, we use the metaphor of stacked cards. Each card shows a question that represents a micro-task. Workers perform the task by answering the question. As shown in Fig. 1, the question can be textual or graphical. The answer may be input as free-form text or from multiple choices. Once a task was completed, the card slides out of the screen and the next card appears. This allows the worker to perform multiple tasks in a passive manner without actively searching the task pool for questions to answer. Since a typical micro-task naturally consists of a simple request and response, this fits well with the question-answer interface.

For example, Fig. 1-a shows a micro-task that asks about the name of a product shown in a picture, which simulates a task to help people who are blind [11]. The worker is asked to choose the correct answer by selecting from buttons below the image, where the candidates were generated by a package recognition engine such as used in Yeh et al. [29].

The support of multiple choices is important to allow performing micro-tasks without text entry. It is known that senior people have trouble in using the keypad on mobile phones, particularly on a touchscreen [20]. The problem is more serious in East Asia, where frequent mode-switching is required to access thousands of characters. For tasks involving a question that is by its nature open, the candidate answers need to be generated in advance, so that senior workers can contribute to final decisions based on their deep knowledge in a particular domain. The candidates may be generated by programs (such as the recognition engine in the example above), or by other (possibly less reliable) workers who use a free-form text interface.

3.3 Skill Assessment

To obtain workers' skill profiles without explicit sources, we use a mechanism to estimate the skills from the micro-task results based on a probabilistic model introduced by Whitehill et al. [23]. This mechanism measures a worker's skill by aggregating the results of multiple-choice tasks. The probabilistic approach is needed because, in realistic micro-tasks, the correct answer is often unknown and it is impractical to provide all of the workers with the same sets of tasks to compare their skills. We extended the model in [23], which considers only the completed tasks, to use both completed and uncompleted tasks. This extension is needed because we aim to assess the skills for each worker, whereas the original work aimed to assess the ground truth for each task. Once the skill level is estimated for each task type, the system can recommend appropriate tasks for each worker.

The probability of worker i giving a correct answer for task j is represented as:

$$p(L_{ij}^c = k \mid z_j^c = k, \alpha_i^c, \beta_j^c) = \sigma(\alpha_i^c \cdot \beta_j^c),$$

where α_i^c is the skill of worker i in task type c, β_j^c is the difficulty of task j in c, L_{ij}^c is the answer by i for j in c,σ is a logistic function, and z_j^c represents the true label that cannot be known directly. The probability is always set to 0 when i has not yet completed j. The values of α_i^c and β_j^c are estimated by an Expectation-Maximization (EM) approach. A larger α_i^c means the worker has a higher skill while a larger β_j^c means the task is easier. If a worker has a higher probability of giving correct answers in a task type, then that worker is believed to have a higher skill for that type.

3.4 Elderly Micro-tasking Scenarios

In addition to micro-tasks such as those sampled in Section 2, there are other types of promising scenarios where our framework could be used for elderly micro-tasking. Since the knowledge of the elderly will be most effective in various types of advisory tasks, we believe that the question-answer metaphor is particularly suitable for the senior workforce.

- *Decision Making*: Based on their extensive experience, senior citizens could give advice on difficult topics such as careers and business options. Our card-style interface can be used both for collecting candidate answers and for voting to choose the best answers.
- *Authoring Support*: Senior citizens could contribute to the creation of educational material by using their expert knowledge and linguistic skills. The types of contributions may include translation, proofreading, and other supportive roles as well as writing new material.
- *Crowd Accessibility*: People with disabilities could benefit from working with remote senior workers in many ways. In addition to real-time assistance such as VizWiz [11], examples include captions and audio descriptions for videos, and making accessible books and websites.
- *Product Evaluation*: It is difficult for product designers to meet the usability and accessibility requirements of senior consumers. For example, text information such as menus and instructions must be easy to understand. Cards can show prototype interfaces for products and senior citizens can evaluate them.
- *Marketing Research*: Because of the rapid growth of the senior market, consumer products and service companies are reaching out to the senior population. In the process of developing products and services, enterprises need to research market acceptance and the card-style interface can be a mechanism for such surveys.

In many of these scenarios, the collaboration between the people who create candidate answers and those who make final decisions is a focus. Our framework provides interfaces for those who have weaker ICT skills or who use mobile devices, thus allowing senior citizens to participate in appropriate roles, whenever they want to and at their preferred locations.

4 Prototype Implementation

We implemented a prototype based on this framework as a Web application with support for various types of devices. The user interface written in JavaScript works on the client side while the skill assessment engine written in Java runs on the server side. The Web-based model allows using a standard server-side program for multiple

(a) (b) (c)

Fig. 2. (a) A card on a mobile touchscreen device, (b) A card seen as a widget in a desktop browser, and (c) Estimated skill visualization

types of micro-tasks, while sharing much of the client-side code for multiple devices. We currently have two variations of the implementation: one for mobile touchscreens and one for desktop browsers.

Fig. 2-a shows a card on a mobile touchscreen device. The card is asking for the name of the object in the picture, which represents a micro-task similar to [11]. The card can be completed by tapping an answer from several options. Once an answer is selected, the next card will appear on the screen with an animation. The worker can also skip cards or go back to previous cards by dragging across the card to the left or right, respectively. Fig. 2-b shows a card in a desktop Web browser. The card is embedded in an SNS interface as a widget, asking to revise a sentence which represents a micro-task similar to the “fix” task in crowdsourced text-editing [17]. The card can be completed by editing the text and clicking on the submit button. The basic design is to the same for the mobile and desktop, except that the desktop interface has “skip” and “back” buttons on top of the card.

The two targets, mobile touchscreens and desktop widgets, were chosen to increase the exposure of senior citizens to micro-tasks. The mobile touchscreen device is known to be easy to use and preferred by the elderly [20]. Although smartphone users are still a minority in the elderly population at this time, it is expected to be an increasingly common device in the next decade. This kind of devices is usually kept switched on and carried on all the day, so it will effectively increase their exposure. The desktop widget can be embedded in any existing interfaces that support HTML-based widgets. The exposure will increase by putting the widget in an interface the users frequently access, such as SNS websites. Although SNS websites are still unfamiliar to most senior citizens, they are becoming popular with them.

Fig. 2-c shows a view that visualizes the estimated skill levels, which we developed to motivate the workers. The skill values for each task type are normalized from 0 to 100 and presented in a radar chart. The average values of all of the workers are also displayed in the chart.

5 Experiment 1: Controlled Study

Our first experiment sought to test the usability of the card-style interface on a touchscreen device and the feasibility of skill assessment, tested in a laboratory setting with simple micro-tasks.

5.1 Settings

Participants. A total of 15 seniors in their 60s to 80s (*mean*=70.1, *SD*=6.8) participated in the experiment. We also involved 16 younger people in their 20s to 40s (*mean*=33.5, *SD*=3.6) as a baseline. A total of 8 of 15 seniors and 14 of 16 younger people had experience with mobile touchscreen devices before the experiment. The senior participants were recruited locally while the younger participants were employees in a global technology company. All of the participants were native Japanese speakers, and the experiment was conducted in Japanese.

Apparatus. The participants used an iPad. As senior citizens rarely have their own tablet devices at this time, the devices were issued by the experimenters and the experiment was conducted in a laboratory. We used the touchscreen interface presented in Fig. 2-a.

Target Micro-Tasks. We tested four types of micro-tasks asking about domain-specific knowledge: Fish, Flowers, Kanji, and English. Fish and Flowers asked for the name of a pictured fish or flower, which simulated crowdsourced image recognition [11]. Kanji asked for the readings of Chinese characters, simulating part of crowd-sourced proofreading of digital books [30]. English asked for the meanings of English words, simulating part of crowdsourced translation [13]. We hypothesized that the senior participants would be better at Fish, Flowers, and Kanji since they have deep knowledge about natural things and their native language from their long experience. We also hypothesized that the younger participants would be better at English since they were working at a global company.

The Fish and Flowers types used visual questions while the Kanji and English types used text questions. For all of the four types, we tested multiple-choice answers in the kind of interface shown in Fig. 1-a, where the candidate answers were provided in advance by software or other people. The free-form text input was omitted in this experiment because it is known that text entry on a mobile device is troublesome for the elderly as described in Section 3.2, and our skill assessment model is suitable for multiple-choice tasks as described in Section 3.3. For each of the questions, the experimenters knew the correct answer.

Procedure. Each participant saw 20 different cards for each task type. Thus they each processed (completed or skipped) a total of 80 cards. The presentation order of the cards was balanced. Prior to the tasks, the participants were given a few minutes of

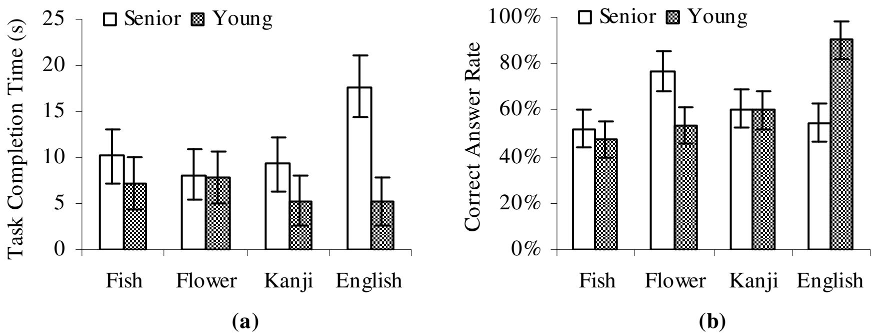


Fig. 3. (a) Task completion times, (b) Actual correct answer rates, by task type with 95% CIs

instructions on how to perform the tasks. After the participants completed all of the micro-tasks, they then answered several multiple-choice questions that asked for feedback on our framework. Each experimental session took less than 30 minutes for most of the participants.

5.2 Results

Task Completion Times. Fig. 3-a shows the average time to complete one task for each task type (skipped cards were excluded from these calculations). We used this measurement to examine whether untrained users could use the card-style interface without confusion. For the senior participants, the overall average values were 10.2, 8.1, 9.3, and 17.7 seconds for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 7.2, 7.7, 5.2, and 5.2 seconds. Analysis of variance showed significant main effects for the age ($F_{1,30.54}$=9.59, p<.005) and task type ($F_{3,2162}$=2.96, p<.05). There are also significant interaction effects between the age and task type ($F_{3,2162}$=5.96, p<.001). A post-hoc analysis found that the senior participants were significantly slower than the young participants for Kanji (p<.05) and English (p<.001). There were no significant differences for Fish and Flowers. This indicates that seniors tend to take longer to process text questions, but they are as fast as younger people at visual questions.

Correct Answer Rates. Fig. 3-b shows the percentages of questions answered correctly for each task type. We used this measurement to assess the actual skill levels, which were calculated using knowledge of the correct answers, as a baseline to test the accuracy of our skill estimation mechanism. For the senior participants, the values were 52%, 77%, 61%, and 55% for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 48%, 53%, 60%, and 90%. Analysis of variance showed a significant main effect for the task type ($F_{3,87}$=10.64, p<.001). It also showed a significant interaction effect between the age and the task type ($F_{3,87}$=17.77, p<.001). A post-hoc analysis found that the senior participants were significantly better at Flowers than the young participants (p<.001). For English, the young participants significantly outperformed the senior participants (p<.001). The age did not significantly affect the results for Fish and Kanji.

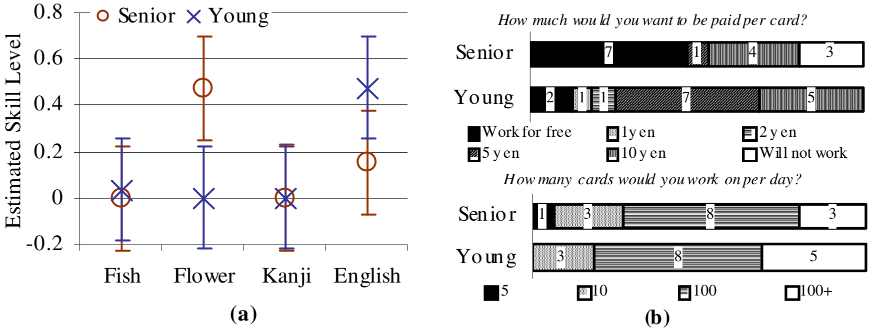


Fig. 4. (a) Estimated skills for each task type with 95% CIs, (b) Results of post-experiment questionnaires (100 yen ≈ 1 USD)

Estimated Skill Levels. To confirm that the skill estimation mechanism works, we compared the participants' estimated skills with their actual correct answer rates. Note that the actual rates are often unknown in realistic micro-tasking situations because the ground truths for the micro-tasks might also be unknown.

Fig. 4-a compares the average estimated skill levels for each task type, which were calculated without using knowledge of the correct answers, as described in Section 3.3. For the senior participants, the values were 2.0×10^{-4}, 4.7×10^{-1}, 1.2×10^{-3}, and 1.5×10^{-1} for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 3.4×10^{-2}, 1.5×10^{-3}, 3.8×10^{-4}, and 4.8×10^{-1}. Analysis of variance showed there was a significant main effect of the task type ($F_{3,87}$=3.94, p<.05). There was also a significant interaction effect between the age and task type ($F_{3,87}$=4.29, p<.01). A post-hoc analysis found that the senior participants had significantly higher skills in Flowers (p<.005) while the young participants had significantly higher skills in English (p<.05). For Fish and Kanji, no significant differences were found. This result matches up well with the actual correct answer rates. This confirms that the differences in skills, at least between the senior and young groups, were correctly detected from the results of the micro-tasks. This indicates that the same mechanism could effectively assess the individual workers. Note that, since the skill levels produced by the probabilistic model are relative values compared to other workers, the absolute values of the skill levels are not known.

The probabilistic model estimates the correct answer by aggregating the answers while considering the estimated skills of the workers. The correct answer rates of the aggregated answers are much higher than the average rates of the individual workers for all of the task types (English: 95.0%, Flowers: 85.0%, Kanji: 85.0%, Fish: 80.0%). This also indicates that the skill levels are estimated successfully.

As explained in the Section 3.3, we extended the probabilistic model proposed by Whitehill et al. [23] to support the cases in which workers did not complete, e.g., skipped, some cards. To confirm this extension works, we compared the estimated skills in English for four participants. They completed 20, 11, 9, and 0 English cards, respectively (the first participant skipped no English card while the last one skipped all of the English cards). They did not produce any incorrect answers for the cards they completed. Their estimated English skills were 1.8, 1.9×10^{-4}, 1.2×10^{-4}, -2.5×10^{-4}. The more cards skipped, the lower the estimated skill, as expected.

Subjective Feedback. First we asked "*How much would you want to be paid for each card if you were doing this work as a job?*" The answers were clearly different between the senior and young groups. More people in the senior group said they were willing to do this work even for free. At the same time, more of the seniors answered they would not work on cards regardless of compensation. This result suggests that seniors were less concerned about the rewards for the micro-tasks.

We then asked "*How many cards would you work on in one day?*" For this question, about half of the participants answered they could do about 100 cards in one day regardless of age. Finally we asked "*Was the system enjoyable?*" The majority of participants said that they enjoyed it. In informal post-experiment interviews, several participants commented that they felt the stacked card interface was enjoyable and game-like. The results for the first two questions are shown in Fig. 4-b.

6 Experiment 2: Live Study

The second experiment sought to test the card-style interface and the skill estimation mechanism in a realistic Web-based setting that involved more practical micro-tasks that would benefit more from the wisdom of age.

6.1 Settings

Participants. A total of 28 senior citizens participated in at least one task, with 9 answering the post-experiment questionnaires. For reference, 13 younger people performed at least one task, with 8 answering the questionnaires. The senior participants were members of an experimental local SNS for senior citizens, with no overlap with the participants in the first experiment, but there may have been some overlap with the respondents from the anonymous preliminary survey described in Section 3.1. The younger participants were employees in a global technology company, including four of the earlier participants. None of the participants who answered the questionnaires had prior experience with crowdsourcing, except for one member of the young group. The experiment was conducted in Japanese.

Apparatus. The participants used their own desktop or laptop PC at home or in their workplaces. We used the widget interface shown in Fig. 2-b. For the senior group, the widget was embedded in the portal of the local SNS. For the young group, the widget was embedded in an enterprise SNS for the employees.

Target Micro-Tasks. We tested two types of micro-tasks for proofreading, which is similar to the "fix" and "verify" tasks in Soylent [17]. We chose these tasks since linguistic knowledge is known to be an advantage of the elderly [2]. Also, our preliminary survey (Section 3.1) suggested that senior citizens are motivated to use their native language skills. The Fix task asked participants to revise a sentence to make it more polite and correct. It first asked whether or not the sentence required any revision, and if the participant answered "yes", then the system asked for a free-form text answer via a card similar to Fig. 1-b. The Verify task asked the participant to choose the more polite and correct sentence from two candidates. It asked for a multiple-choice answer via a card similar to Fig. 1-c. For both types of task, a brief context (e.g., the person who will use the sentence) was included on the card as a note. The scenario assumed that at first some participants would create candidate revisions through the Fix interface, and then other people would make the final decisions through the Verify interface. This experiment is intended to test realistic micro-tasks and the sentences to be proofread were prepared based on actual problems and the correct answers were unknown for each task.

Procedure. Each participant was presented up to 45 and 20 different cards for Fix and Verify tasks, respectively. Thus they completed up to 65 cards at their own pace. The presentation order of the cards was randomized. Prior to the tasks, the senior participants were given a brief explanation of the purpose of this study and instructions on how to perform the tasks. The young participants received an introductory announcement via email or a blog post. After the experiment, they were asked for feedback as in the first experiment. The second experiment ran for approximately one week.

6.2 Results

Participation Status. During the experimental period, a total of 42 senior users were shown one or more cards (counted based on the user ID), and 28 of them completed 906 cards in total (*mean*=32.4, *SD*=27.4). A total of 387 young users were shown cards (counted based on their IP addresses), and 13 of them completed 104 cards (*mean*=8.0, *SD*=11.0). Since 12 of the seniors completed all of the cards for at least in one task type, it seems likely they would have performed more tasks if they were available. None of the young group completed all of the tasks. This might suggest that the card-style interface motivated senior people more than young people, but the relevant experimental conditions were not the same for each group.

Task Completion Times. Fig. 5-a shows the average time to complete one task for each task type (skipped cards and neglected cards, which we defined as cards that took more than 5 minutes to respond to, were excluded from the calculations). For the senior participants, the overall average values were 82.4, 27.5, and 22.1 seconds for Fix-Yes, where they actually fixed the sentence with the free-form text entry card, Fix-No, where they stated that the sentence did not require any revision, and Verify, respectively. For the young participants, the values were 77.9, 45.3, and 29.3 seconds. Note that the value for each trial may be affected by the network delay. Analysis of variance showed significant main effect for the task type ($F_{2,949}$=57.93, p<.001). The age had no significant effect. A post-hoc analysis found that Fix-Yes tasks took significantly longer time than Fix-No and Verify tasks (p<.001).

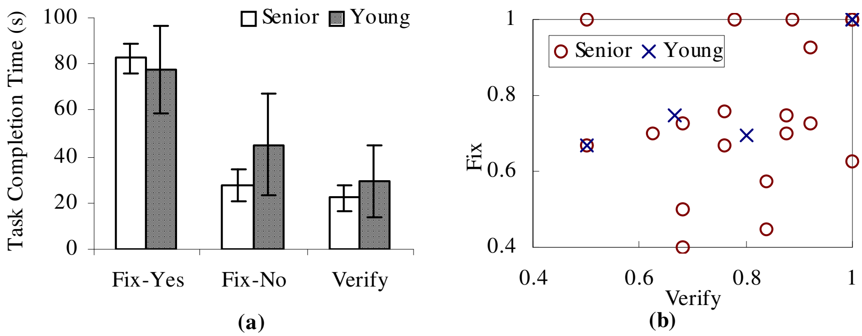


Fig. 5. (a) Task completion times for each task type with 95% CIs, (b) Distribution of estimated correct answer rates for each task type

Estimated Skill Levels. For the senior participants, the average values of the estimated skill levels were 0.11 and 0.23 for Fix and Verify, respectively. For the young participants, the values were 0.29 and 0.58. Note that the skill levels for Fix were calculated based on the results of the yes-no questions that asked the necessity for revisions. The actual content of revised sentences was ignored in this calculation due to a limitation of our skill assessment model, which cannot work well with open questions.

Analysis of variance showed significant main effect for the age ($F_{1,44.77}$=5.19, p<.05) and task type ($F_{1,31}$=5.18, p<.05). Despite the assumption that senior people would be better at linguistic skills, the young group showed higher skills in this experiment, perhaps because all of the young participants were highly educated and can be expected to have better-than-average linguistic skills. Meanwhile, the effect of the task type seems to be caused by the difference in the percentage of the "skipped" trials (17.1% for Fix, 9.3% for Verify). As confirmed in Section 5.2, skipped trials make the estimated skill lower in our skill assessment model.

Fig. 5-b shows the distribution of estimated correct answer rates, calculated by assuming that the aggregated answers are always correct. This indicates that the skill for Fix is not always positively correlated with the skill for Verify (R^2=.18) and the values are vary widely depending on the worker (SD=.28 for Fix, SD=.18 for Verify). This result suggests that we have to carefully choose skilled workers for each task to insure high-quality results.

Subjective Feedback. We asked similar questions as in the first experiment. Although only some of the participants answered the questions (since this experiment was conducted in a real-world setting), the results seemed to generally confirm the previous results. For the question asking about the compensation, the majority of the senior participants were less concerned about the rewards. The young participants tended to want 5 to 10 yen per card. To the question asking for a number, the participants generally responded that they could do around 100 cards per day. In contrast to the first experiment, many of the participants stated that they would not work on Fix and Verify cards regardless of compensation (50% for seniors, 38% for youths). This might indicate that realistic micro-tasks can involve some unenjoyable work, and thus some additional motivational mechanism would be needed.

7 Discussion

Based on the experimental results, we discuss the challenges and future work to invite senior citizens to do micro-tasks.

7.1 Interface Usability and Acceptability

Overall the results indicated that our approach was adequately usable for and accepted by the target users. The senior participants could understand how to perform all of the types of tasks and complete them with only brief instructions. We also confirmed that the proposed framework worked even in a live situation that involves realistic tasks. However the experimental results also indicated that the performance largely depends on the users and the task types. Since we did not control for the participants' ICT experiences or age-related losses of abilities, further research is needed to discuss the realistic usability and accessibility of our interface for specific users and tasks.

We also need to address the range of applications, since the card-style interface basically supports only simple question-answer interactions. Task-decomposition strategies such as discussed by Bernstein et al. [17] are promising ways to handle more complex tasks. They tried to assure the quality of their results by using a voting mechanism. As indicated in our second experiment, the mixed use of free-form text and multiple choice cards can support such an answer-vote strategy. Once the task is split into micro-tasks, the question-answer interactions will work well because each micro-task is typically represented as a pair with a simple request and its response. In such cases, multiple choice cards may be preferred and more effective than free-form text cards, i.e., with fewer skips and quicker responses. Some participants suggested that it would be helpful if voice input is available. As we used Web-based implementations, the voice input functions supported by modern Web standards are available.

7.2 Skill-Based Task Recommendations

Since the elderly are regarded as passive users of ICT, they will not actively explore a pool of tasks to find tasks they like, so our system should offer to the senior workers tasks that fit their skills. The experimental results confirmed that our skill assessment model could assess the differences in skills. This means that once a worker answered several cards for one type of task, the system could assess whether or not the task type is good for that worker. For example, if Alice is better at "fix" and Bob is better at "verify", then the system would provide Alice and Bob with more "fix" and "verify" tasks, respectively. The second experiment indicated that the skills of the elderly vary widely. Even if a worker is better at a type of tasks, it does not necessarily mean the worker is also better at a related, but different type of tasks (e.g., "fix" and "verify" tasks). As the probabilistic approach can estimate workers' skills with a small number of samples, it would allow quickly determining the skills for each specific domain.

A collaborative filtering approach [31] can help recommend task types even when a new worker has never tried those tasks. Our skill estimation mechanism produces a skill profile matrix that can be used for this purpose, where the value of each element represents the skill level of a worker for a task type. We can use a similar approach to Yuen et al.'s work [32], which used a task preference matrix to recommend tasks based on a collaborative filtering mechanism.

7.3 Motivation Management

Our experimental results show that seniors did not care much about the monetary rewards. This indicates that we need special care in the design of the incentive mechanism for senior workers. For example, some participants commented that the skill visualization (Fig. 2-c) allowed them to discover that they had greater knowledge in a particular domain, and that motivated them. The motivation may also be affected by the enjoyability. It is worth noting that many of the participants found that the card-style interface was enjoyable and like a game even though we did not intend to include any gamification mechanisms in our framework. However, we also have to note that, in the second experiment, the participants gave less positive feedback than in the first experiment. At this time we cannot conclude whether this resulted from the differences in the participants' communities, in the interfaces (touchscreen vs. desktop widget), or in the task conditions (laboratory vs. realistic), which means further investigation is needed.

Another key to success would be increasing the exposure of senior citizens to micro-tasking opportunities. We believe that the mobile interface and widget interface will be effective for this purpose, as described in Section 4. Also, the card-style interface could be applied to other situations. Some candidate situations include smart TVs, kiosk terminals, vehicle navigation displays, in-flight entertainment systems, and even when waiting in front of a microwave [33].

8 Conclusion

This paper described a framework that aims to increase the participation of senior citizens in various types of micro-tasks with a simple and versatile card-style interface. It uses a probabilistic model to assess their skill profiles, which is necessary for task recommendations and for quality assessments and improvements. The experimental results showed that even untrained seniors could handle micro-tasks by using the card-style interface in both controlled and real-world settings. We also confirmed that the differences in skills can be assessed by analyzing the results of the micro-tasks. We still have challenges to deploy the framework as a platform for senior citizens who want to work. The scope of tasks should be broadened, but more importantly, enjoyment and motivational factors should be evaluated and improved as well as the economic efficiency. We hope that the proposed framework will help senior citizens participate in the workforce and fully utilize their broad experience and knowledge in the society.

Acknowledgments. This research was partially supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program.

References

1. Statistics Bureau, Monthly Report, `http://www.stat.go.jp/english/data/jinsui/tsuki/` (retrieved April 15, 2013)
2. Ouchi, Y., Akiyama, H. (eds.): Gerontology – Overview and Perspectives, 3rd edn. Univ. of Tokyo Press (2010) (in Japanese)
3. United Nations University, Active Ageing, `http://wisdom.unu.edu/en/active-aging`
4. Leibold, M., Voelpel, S.: Managing the Aging Workforce. John Wiley & Sons (2006)
5. Loten, A.: Small firms, Start-ups Drive Crowdsourcing Growth. Wall Street Journal (February 28, 2012)
6. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the Crowdworkers? Shifting Demographics in Mechanical Turk. In: Proc. CHI EA 2010, pp. 2863–2872. ACM (2010)
7. Hiyama, A., Nagai, Y., Kobayashi, M., Takagi, H., Hirose, M.: Question First: Passive Interaction Model for Gathering Experience and Knowledge from the Elderly. In: Proc. PerCol 2013, pp. 151–156. IEEE (2013)
8. Office for National Statistics: Use of ICT at Home (2007)
9. Kurniawan, S.: Older People and Mobile Phones: A Multi-Method Investigation. Int. J. Hum.-Comput. Stud. 66(12), 889–901 (2008)
10. Leung, R., Tang, C., Haddad, S., McGrenere, J., Graf, P., Ingriany, V.: How Older Adults Learn to Use Mobile Devices: Survey and Field Investigations. ACM Trans. Access. Comput. 4(3), Article 11 (2012)
11. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: VizWiz: Nearly Real-Time Answers to Visual Questions. In: Proc. UIST 2010, pp. 333–342. ACM (2010)
12. von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M.: reCAPTCHA: Human-based Character Recognition via Web Security Measures. Science 321(5895), 1465–1468 (2008)
13. Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk. In: Proc. EMNLP 2009, pp. 286–295. ACL and AFNLP (2009)
14. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing User Studies with Mechanical Turk. In: Proc. CHI 2008, pp. 453–456. ACM (2008)
15. Guy, I., Perer, A., Daniel, T., Greenshpan, O., Turbahn, I.: Guess Who? Enriching the Social Graph through a Crowdsourcing Game. In: Proc. CHI 2011, pp. 1373–1382. ACM (2011)
16. Kulkarni, A., Can, M., Hartmann, B.: Collaboratively Crowdsourcing Workflows with Turkomatic. In: Proc. CSCW 2012, pp. 1003–1012. ACM (2012)
17. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: Soylent: A Word Processor with a Crowd Inside. In: Proc. UIST 2010, pp. 313–322. ACM (2010)

18. Noronha, J., Hysen, E., Zhang, H., Gajos, K.Z.: PlateMate: Crowdsourcing Nutritional Analysis from Food Photographs. In: Proc. UIST 2011, pp. 1–12. ACM (2011)
19. Leonardi, C., Albertini, A., Pianesi, F., Zancanaro, M.: An Exploratory Study of a Touch-based Gestural Interface for Elderly. In: Proc. NordiCHI 2010, pp. 845–850. ACM (2010)
20. Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., Ifukube, T.: Elderly User Evaluation of Mobile Touchscreen Interactions. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 83–99. Springer, Heidelberg (2011)
21. Web Accessibility and Older People: Meeting the Needs of Ageing Web Users, `http://www.w3.org/WAI/older-users/`
22. CrowdFlower, `http://crowdflower.com/`
23. Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., Movellan, J.: Whose Vote should Count More? Optimal Integration of Labels from Labelers of Unknown Expertise. In: Proc. NIPS 2009, pp. 2035–2043 (2009)
24. Heimerl, K., Gawalt, B., Chen, K., Parikh, T., Hartmann, B.: CommunitySourcing: Engaging Local Crowds to Perform Expert Work via Physical Kiosks. In: Proc. CHI 2012, pp. 1539–1548. ACM (2012)
25. Macdonald, C., Ounis, I.: Voting for Candidates: Adapting Data Fusion Techniques for an Expert Search Task. In: Proc. CIKM 2006, pp. 387–396. ACM (2006)
26. Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., Farrell, S.: Harvesting with SONAR: The Value of Aggregating Social Network Information. In: Proc. CHI 2008, pp. 1017–1026. ACM (2008)
27. Pfeil, U., Arjan, R., Zaphiris, P.: Age Differences in Online Social Networking – A Study of User Profiles and the Social Capital Divide among Teenagers and Older Users in MySpace. Comput. Hum. Behav. 25(3), 643–654 (2009)
28. Staffing Mature Worker Survey, `http://www.goldenworkers.org/images/publication/mature_workers_survey_2012_adecco.pdf`
29. Yeh, M.-C., Tai, J.: A Hierarchical Approach to Practical Beverage Package Recognition. In: Ho, Y.-S. (ed.) PSIVT 2011, Part I. LNCS, vol. 7087, pp. 348–357. Springer, Heidelberg (2011)
30. Kobayashi, M., Ishihara, T., Itoko, T., Takagi, H., Asakawa, C.: Age-based Task Specialization for Crowdsourced Proofreading. In: Stephanidis, C., Antona, M. (eds.) UAHCI/HCII 2013, Part II. LNCS, vol. 8010, pp. 104–112. Springer, Heidelberg (2013)
31. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems. ACM Trans. Inf. Syst. 22(1), 5–53 (2004)
32. Yuen, M.-C., King, I., Leung, K.-S.: TaskRec: Probabilistic Matrix Factorization in Task Recommendation in Crowdsourcing Systems. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012, Part II. LNCS, vol. 7664, pp. 516–525. Springer, Heidelberg (2012)
33. Watanabe, K., Matsuda, S., Yasumura, M., Inami, M., Igarashi, T.: CastOven: A Microwave Oven with Just-in-Time Video Clips. In: Proc. Ubicomp 2010 Adjunct, pp. 385–386. ACM (2010)

An Experimental Study of Chinese Shopping Related Sharing Behaviors

Min Li[1], Jun Zhang[1], Zhengjie Liu[1], and Graham I. Johnson[2]

[1] Sino-European Usability Center, Dalian Maritime University, Dalian, 116026 P.R. China
[2] Consumer Experience Research, NCR Corporation Dundee, UK
minleely@gmail.com, {William,liuzhj}@dlmu.edu.cn, graham.johnson@ncr.com

Abstract. Social media has become an increasingly important venue for social interaction and communication. Most previous research have shown the information sharing on social media platforms, but few focus on shopping related sharing (SRS) information and how these information influence consumer decision-making. Through a combination of interviews with observations in our lab to the users of social media and online shopping sites, we want to find answers to the question: what are the factors that influence Chinese consumers' SRS behaviors? It was discovered that the factors like why, what, whom, where, when etc. have significant influence to Chinese consumers' SRS behaviors. The study offers insights into the relationship between SRS and Chinese consumer decision-making, and design implications are discussed.

Keywords: Social media, shopping, sharing, social influence, consumer decision-making.

1 Introduction

Social media is now flourishing worldwide and plays a more and more important role in most people's lives as a means of information acquirement and sharing. It is notable that more and more shopping related contents appeared along with the combination of social media and E-commerce, for example, pictures with links from other online shopping websites and consumer generated contents about the whole shopping experience. Businesses also create public profiles or make advertisements on social media platforms to broadcast their products and attract consumers.

Consumer decision-making process is complicated and influenced by many factors, and now more and more affected by social media information. However, what is the relationship between social sharing behaviors and consumer decision-making? We propose a conceptual framework to indicate how social influence impact consumer decision-making process, and want to recognize the role social media plays on Chinese consumer decision-making from some point of view by studying SRS behaviors. By doing this, we can get the first hand knowledge of SRS and help us further study how SRS influence consumer decision-making.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 608–615, 2013.

The remainder of this paper is structured as follows. First we introduce the background and related work. Then we propose the conceptual framework to guide our study, and then describe research methods and processes. Next we describe the findings, followed by a short discussion of the design implications of our work for business and those in the interaction design community. Finally, we highlight the conclusions and next steps of our research work.

2 Background and Related Work

Recent years, the studies on commerce-oriented reviewing and sharing platforms (CRSP) primarily addressed the problems of summarizing and analyzing online reviews. For example, researchers conducted data mining on reviewing text generated by consumers rather than measuring the rating scores[1].Some researchers studied the problem of generating a "rated aspect summary" of short comments, which is a decomposed view of the overall ratings for the major aspects so that a user could gain different perspectives towards the target entity [2]. Some other studies addressed the problem of sentiment classification of the comments [3][4]. Although researchers have explored many aspects of social media platforms (e.g. blogs, Facebook, and Twitter etc.), most of which have primarily focused on the social network connections, user activities (like what, how, why, where and when) and privacy concerns [2] [5]. Some researchers paid more attention to the information contents. They classified the contents into some categories, analyzed the political opinions by data mining tools and conducted sentiment analysis of the comments [6].

In literature [7], the researchers showed that E-commerce had been largely affected by social media. Researchers have proposed that consumers share shopping information on social media platforms through status messages [8]. They presented types of questions asked by users, and found that five percent of the questions asked were shopping related. However, they didn't explore in-depth how shopping related information could influence purchasing decisions, which was our research focus. Jansen et al. examined Twitter usage for brand sentiment, and gave a report about information sharing among 20% of status updates by Twitter participants, and specifically indicated that people shared generally positive opinions about brands [9]. Park et al. focused on sharing of daily deal URL on Twitter, and they adopted quantitative analysis method to analyze sharing behaviors of daily deal information [10]. They provided evidence of the potential contribution of Twitter in improving the daily deal sales performance. There are also researchers with focuses on advertisements and business public profiles on social media [11]. However, they still didn't pay much attention on Chinese social media, so we are interested, from an HCI and user experience perspective, in the different ways in which social media is being used for commerce in order that subsequent designs and affordances of such services can be improved.

Previous studies paid less attention to the relationship between SRS and consumer decision-making process. We propose a conceptual framework based on the model [12] proposed by Thompson et al. to illustrate social influence on three sub-phases of consumer decision-making process (need recognition, information collection and after purchase evaluation). Basing on the conceptual framework, we try to investigate social sharing behaviors of Chinese consumers and the motivations underlying the behaviors. In the future work, we will explore how the sharing behaviors on different social media platforms influence decision-making.

3 Conceptual Framework

Many scholars have argued that a virtual community is capable of disseminating product knowledge and awareness and influencing community members' purchase behavior [10][13]. On the basis of previous research [12], in which Thompson et al. proposed the consumer decision process model to illustrated the core decision-making process. The process begins with need recognition, then information search, alternative evaluation, purchase and after purchase evaluation. We focus on information sharing related parts and propose a conceptual framework to illustrate factors influence the three steps of decision-making process, see details in Figure 1. Factors related to need recognition (NR), information collection intention (ICI) and after purchase evaluation (ACI) include perceived benefit (PB) and perceived risk (PR). Factors related to PR include trust (T), privacy concerns (PC) and social influence (SI). And SI is a factor influence both T and PC.

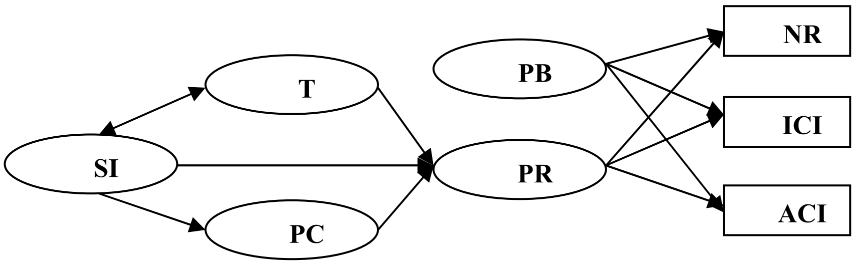


Fig. 1. Conceptual framework

PB refers to perceived benefit that is expected to result from the three steps. PR refers to perceived risk that is defined as the probability of any loss that can occur. T refers to trust that is defined as the belief that the result will be good. PC refers to privacy concerns that are defined as personal data or information privacy, users care about the unclear results from information leakage. SI refers to social influence that users may impacted by social media and/or other members in the social networks. We emphasize SI factor which is one of the most important factors in consumer decision-making process. And in this paper, we try to understand SRS behaviors and the social influence behind the behaviors.

4 Methodology and Process

We adopted lab-based sample user study. We recruited active users of social media sites and online shopping sites for interview and observation in our lab. They were asked to talk about their experience in SRS behavior and simultaneously show us on a PC how they were doing it. For example, we asked participants to answer whether or not share shopping related information and how, when, where and who to share. They talked about the latest SRS behaviors on social media platforms or other offline conditions. By using this method, we gained the knowledge about real online buyers' shopping and sharing experience quickly and directly.

We recruited 30 participants via an agency and got 29 qualified from them which were numbered from 1 to 29 respectively, see details in Table 1. We collected data of participants' SRS behaviors, demographic information, applied social media applications and online shopping behaviors by screenshot, audio recording and note taking. Each session lasted around 1.5 hours.

Table 1. Distributions of basic information of the participants

Category	Item	Value
Gender	Male/Female	14/15
Age group	18-24/25-30/31-36	9/10/10
Education degree	Junior college student/ Undergraduate/ Postgraduate	15/12/2
Online shopping websites used	Taobao.com	29
	Groupbuy websites	11
	Dangdang.com	5
	Paipai.com	4
	Amazon.cn	3
	360buy.com	2
	Others	4

We used grounded theory to analyze data and had some interesting findings during the study and summarized the factors that influence SRS behaviors, like why and what to share, where to share, whom to share with, and when to share. We present the findings of the study in details below.

5 Results

We will present the findings related to Chinese SRS behaviors both online and offline. We focus on the following aspects: 1) why and what to share, 2) where to share, 3) whom to share with, and 4) when to share.

Why and What to Share. Participants usually do not want to share shopping related information, because they have privacy and personal image concerns (20/29, means20 participants mentioned this point out of total 29 participants). But the situation is not unalterable, they will share when there are rewards for doing so, e.g., sharing groupbuy related information.

Participants are afraid that it may harm their privacy by sharing on social platforms. They do not want friends on their online social network to know what they bought, neither personal products (gifts for lovers, health medicine, etc.) nor daily goods. They regard shopping as a private activity:

"I don't want to share shopping information with others because it is personal thing." Participant 19, female.

Dholakia et al. proposed several social influence variables that could affect users' behaviors, such as social identity [8]. In our study, we found Chinese consumers cared a great deal about their personal images both online and offline, which was consistent with social identity. On one hand, some kind of information can improve image, leaving a good impression to others that the person is skillful, special, etc. On the other hand, some kind of information can damage image, like fake information, so participants do not want to take the risks introduced by sharing.

It is notable that personal image and privacy concerns may be weakened when there are rewards for doing so, e.g., sharing group-buy related information. We saw that the group-buy information was shared on Renren by participants 11. If any friend clicked the hyperlink to go to the group-buy website and at last bought the product, the participant could get a discount from the group-buy website in the form of a voucher.

Where to Share. Participants tend to share particular kinds of information on different platforms with different market positioning or segment. There are four major kinds of sharing platforms being mentioned frequently in our study: 1) built-in consumer reviewing platform and attached commerce-oriented social network services (SNS) platforms on online shopping websites, 2) relationship-emphasized or socially-oriented SNS platforms, 3) instant messenger (IM) platforms, and 4) dedicated consumer reviewing platforms.

Taobao is a typical online shopping website with a built-in consumer reviewing platform. Buyers can search the comments left by other buyers on it before they make purchase decisions. After the deal, they can also make a comment related to the products they bought. An example of the attached commerce-oriented SNS platforms is i.taobao.com, which is attached to Taobao and encourages users to participate in social networking activities as they do on other SNS platforms. Participants do have SRS behaviors on these platforms, but we find that the motivations of the SRS behaviors are rewards in the forms of reward points and discounts provided by Taobao but not social purposes to recommend better products to other users on the platform.

Renren, Sina Weibo and Tencent Weibo are typical examples of socially-oriented social media platforms. Even active users seldom share shopping related contents on these platforms. The main reason is concern on personal image and privacy. Another reason is that they regard shopping related contents so trivial that too much sharing may cause information overload to others.

Almost all the few SRS behaviors (5/29) we observed on these platforms were group-buy related, because strong-tie friends are regarded as the potential group-buy partners. This will potentially bring benefits to both the sharer and the friends:

"I will share the group-buy food to my friends on Renren, because I can get the reward of 5 Yuan from the website if my friend also buys this food through my link. That's a win-win. Why not?" Participant 8, female.

QQ, Fetion, MSN Messenger and Wangwang are typical examples of IM platforms. QQ is the most popular IM tool in China. Almost every Chinese web user has one account at least. Fetion (developed by China Mobile) and MSN Messenger also have large number of users in China. Wangwang is an IM tool attached to Taobao, and is most widely used between buyers and sellers when they discuss the deal related information. In our study, we found that most of the SRS behaviors happened among QQ-group members and users even have dedicated shopping group names like "shopping" and "group-buy". Participants regard this kind of platforms as appropriate places for SRS probably because QQ IM communication is so popular and timely that there is a great chance that there will be always a few friends already online and can get feedback immediately.

Dianping (www.dianping.com) and Mogujie (www.mogujie.com) are typical dedicated consumer reviewing social platforms in China, like Epinions in west countries. Users use these platforms to exchange shopping related information and comments. The majority of the information shared on these platforms is systematic and in-depth comments related to the products, even hyperlinks to online shopping websites, and users tend to seek for helpful information from these platforms in their shopping decision-making processes.

Whom to Share with. We will summarize people whom participants are more likely to share with. We find that there are mainly three kinds, 1) friends nearby in the physical world, 2) people that can potentially bring benefits, and 3) people that have similar preferences or experiences with the participant.

We find that participants are more likely to share with friends geographically nearby in the physical world (28/29). Apart from using IM tools like QQ, in more cases participants share by face to face or phones, because they live near to each other or have the chance to meet frequently in daily lives and know each other well:

"The products will be delivered to my company on working days, and my colleagues can see what I have bought, and discuss about them" Participant 14, female.

Participants are likely to share with those who can potentially bring benefits (8/29), e.g., share with potential group-buy partners. Participants told us that good products or services should be shared or recommended to others to bring benefits to them. But we think they also had concerns about the rewards for doing so based on our observations.

Participants are likely to share with those who have similar preferences or experiences with them (5/29), for information from those people tends to be more valuable to them. The phenomena can be explained by social influence theory that users want to keep consistence with peers [14], which can verify the conceptual framework we proposed.

When to Share. We find that there are two main situations when participants share, one is when the quality and attached services of the products are very good or very bad, and the other is when participants just want to show themselves, especially when they buy special things.

Participants are more likely to share when the quality and attached services of the products are very good or very bad (16/29). If the product is very good and may be suitable for others, then the participant will suggest potential consumers to buy it.

Sometimes participants share comments on bad products or services to others to express their negative feelings and also to remind other consumers not to buy it, and at this time the comments will be longer.

Sometimes participants just want to show themselves, especially when they have bought some special products. Participant 3 showed us a picture of a bought white mouse on his Qzone. He explained that the white mouse is bought online as the food for his pet python. He thought that the sharing was so unique and made him different from others.

6 Discussion and Design Implications

We have achieved a basic understanding of Chinese SRS behaviors based on the analysis of the qualitative data collected in the lab-based sample user study. We identified the factors that could influence SRS behaviors. It is notable that Chinese consumers are very much concerned about personal image online, but when there are economical rewards, the concerns may be weakened. We believe that market planners can make use of the insight to encourage consumers to share SRS information by giving them more rewards, such as coupons or discounts.

Chinese IM tools also should be mentioned, especially QQ. Because of its convenience for using and so large number of users, shopping related topics are diffusing frequently among QQ users and QQ group members. So, it is convenient to use IM tool to gather consumer needs to keep current consumers and attract potential consumers at the same time.

Some participants mentioned that often they can't find the right channel to access shopping related information except search engines like Google and Baidu. There are chances for companies to make better use of social media platforms to broadcast contents about their brands or products to more potential consumers. On another hand, recommendation system should give the most relevant and valuable SRS contents to the right people who really need them.

Another important aspect is peer's pressure. Consumers care much about the comments from nearby friends and people with similar preferences or experiences. Consumers want to keep consistent with peers, so they regard peers' reviews as most valuable. So, some mechanism should be considered to support interaction about SRS between geographically nearby strong-tie relationships.

7 Conclusions and Next Steps

We have achieved a basic understanding of SRS based on the analysis of the qualitative data collected in the lab-based sample user study. Four kinds of factors that influence SRS behaviors were identified. In the near future, we will discuss the relationship between SRS behaviors and consumer decision-making, and verify the conceptual framework we proposed. And next we will start quantitative and qualitative analysis of questionnaire data we collected online to enrich and verify the findings we've mentioned. And then we will also begin to address other social media platform, such as location-based services (LBS) applications, to study SRS behaviors and enrich our knowledge about the relationship between SRS and consumer decision-making.

References

1. Zhang, Z., Li, X., Chen, Y.B.: Deciphering Word-of-Mouth in Social Media: Text-Based Metrics of Consumer Reviews. ACM Transactions on Management Information Systems 3(1) (2012)
2. Nylander, S., Rudstrom, A.: Questions, inspiration, feedback, and contributions: How entrepreneurs network online. In: Proceedings of C&T 2011, pp. 128–137. ACM Press (2011)
3. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of WWW 2007, pp. 171–180. ACM Press (2007)
4. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 339–346 (2005)
5. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: Proceedings of CHI 2009, pp. 211–220. ACM Press (2009)
6. Song, J.X., Finin, T., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (webKDD/SNA-KDD 2007), San Jose, California, USA, pp. 56–65 (2007)
7. Turban, E., Bolloju, N., Liang, T.P.: Social Commerce: An E-Commerce Perspective. In: Proceedings of ICEC 2010, pp. 33–42. ACM Press (2010)
8. Morris, M.R., Teevan, J., Panovich, K.: What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In: Proceedings of CHI 2010, pp. 1739–1748. ACM Press (2010)
9. Jansen, B.J., Zhang, M., Sobel, K., Chowdhury, A.: Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology 60 (2009)
10. Park, J.Y., Chung, C.W.: When Daily Deal Services Meet Twitter: Understanding Twitter as a Daily Deal Marketing Platform. In: Proceedings of WebSci 2012, pp. 233–242. ACM Press (2012)
11. Cvijikj, I.P., Michahelles, F.: Understanding Social Media Marketing: A Case Study on Topics, Categories and Sentiment on a Facebook Brand Page. In: Proceedings of MindTrek 2011, pp. 175–182. ACM Press (2011)
12. Thompson, S.H., Yeong, Y.D.: Assessing the consumer decision process in the digital marketplace. The International Journal of Management Science 31, 349–363 (2003)
13. Dholakia, U.M., Bagozzi, R.P., Pearo, L.K.: A social influence model of consumer participation in network- and small-group-based virtual communities. International Journal of Research in Marketing 21, 241–263 (2004)
14. Moscovici, S.: Social influence and conformity. In: The Handbook of Social Psychology, vol. 2, pp. 347–412. Random House, NY (1985)

Perceptions of Facebook Privacy and Career Impression Management

Danielle Pilcer and Andrew Thatcher

Industrial Psychology, University of Witwatersrand, Johannesburg, South Africa
dpilcer@gmail.com, andrew.thatcher@wits.ac.za

Abstract. Facebook was initially designed as a social tool representing a person's social life, yet today it functions as a platform synthesizing all facets of life, including the work context. Within the frame of the social capital theory, a survey was conducted measuring Facebook users' experience, perceptions of privacy and career impression management. Results found that higher Facebook experience was related to increased levels of trust and career impression management in terms of self- monitoring and work relations. Higher work privacy was related to career impression management in terms of lower work relations and higher perceived workplace outcomes; and career impression management in terms of higher perceived self-monitoring and workplace outcomes were related to lower levels of privacy awareness. It was concluded that how one uses Facebook and one's perceived value in presenting an image of an employable person is influenced by how one views their Facebook privacy.

Keywords: Facebook, privacy, impression management, employability, social capital theory.

1 Introduction and Background

Facebook is a platform that has the potential to synthesise all facets of a person's life in one space [1]. It allows users to present an overall picture of who they are rather than present a compartmentalised image based on the audience. The content and amount of information users share are within their discretion, however Facebook's influence, reach and domination of web-based experiences has implications for user's that are unaware of the exposure they receive on Facebook, especially within the context of their work life. The onus thus lies with the user to be cognisant of their exposure online, the level of accessibility of the content shared, and the potential effects of this shared information on their work image. Facebook has already reared its tentacles in the world of work as a recruitment tool [2]. Yet in the instance of already employed individuals Facebook can be used by employers, either consciously or unconsciously, to monitor their employee's activities outside of the office, or screen content that has the potential to place the company or the employee in a negative light [3].

Thus, this paper's line of enquiry is as follows: Is the image one presents of oneself on Facebook fitting for the work context and does one perceive importance and value in presenting self-worth as an employable person. The research questions were as

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 616–623, 2013.
IFIP International Federation for Information Processing 2013

follows: Is there a relationship between how one uses Facebook and how one manages one's work impressions on Facebook? Is the way in which one uses Facebook and one's view of Facebook privacy related? Does one's view of Facebook privacy relate to one's work impression management on Facebook?

2 Related Work and Theory

2.1 The Social Capital Theory

The Social Capital Theory reveals that Facebook's appeal is inherent in its social capital value, that is, the value that is derived from ones' social network [1]. This social capital is related to indices of psychological wellbeing including self-esteem [1]. Self-esteem, as a person's overall self-evaluation of their worth, assumes that all individuals have a vital need to maintain or raise their self-esteem and strive for positive self-presentations [4]. Translated into Facebook terms, it is expected that the value acquired from Facebook influences a person's self-esteem positively, which is then emulated in the way the person presents themselves online. That is, the value derived from using Facebook results in a specific image presentation aligned with ones' appraisal of self-worth. Thus, the social value drawn from Facebook will influence the image presented, but both the value and the image will be duly influenced by what Facebook is used for and who makes up one's audience: who are one's friends, how many different areas of one's life make up this group of friends, and how many watching eyes are within this network of friends [5]. Essentially the audience is vital as it supplies the platform from which social value is drawn. In this sense, Facebook architecture is an extension on Foucault (1977) imagery of the Panopticon (within a society the effect of constant observation is constant behaviour monitoring by the observed [6]). Facebook constitutes a society where one is constantly being observed yet in turn constantly observing others too [7]. Thus, there are different categories of users: those that use Facebook to share content, and those who use it to observe. Either grouping of users is subject to the same reality, which sees Facebook allowing large and diverse friend networks requiring monitoring of one's self-presentation on the basis of the convergence of social contexts (social convergence) [1;8]. However, what constitutes a positive image will not be the same across all social groups.

2.2 Facebook Career Impression Management and Privacy

A 'socially converged' audience challenges the ability to consciously control selected activities for a desired impression on a particular audience [9]. That is, self-representing in the Facebook context is not the same as traditional impression management contexts [10]. Thus, by drawing from existing theory this paper explores the degree to which the particular audience of the represented self is work-centred. Behaviour monitoring at work is directed at increasing the likelihood of fulfilling financial and social goals, and avoiding possible negative consequences [11]. Thus creating and sustaining a desirable image on Facebook suitable for work, assumes that users will monitor their image and activities on Facebook to maintain favourable relations at work, to present the image of an employable person or suitable candidate that is promotion worthy with career advancement prospects, and to avoid possible negative workplace consequences (such as dismissal or disciplinary proceedings).

Behaviour monitoring for a specific audience brings to focus a user's friend strategy [9]. A friend strategy is the approach employed when selecting who to add, accept, or reject as a friend. It is concerned with whom a person selects to be a part of their network and the selection criteria thereof [12]. Strategies range from accepting everybody requesting, to accepting only people known personally or face-to-face. The chosen strategy influences what and how much a user presents [13]. Furthermore, 'privacy control' as a component of one's friend strategy sees users determining which parts of their network can access specific information. Users can select different settings, 'everybody, friends of friends, friends only', in terms of every wall post, photo tag and status update [14]. Having the ability to select friends and what they can see, facilitates a sense of trust in Facebook and in one's friend network, thus encouraging and stimulating information disclosure [15].

Facebook's inherent privacy controversies centre on information exposure to the wrong audience. One's view of Facebook privacy directs the classification of the wrong audience and the amount of information at this audiences' disposal. How one views Facebook privacy will influence the degree to which privacy is managed, that will in turn affect the quantity and nature of the content shared, the friend strategy employed, and what Facebook is used for.

3 Instrumentation, Methodology and Results

Within the frame of the Social Capital Theory, a survey was conducted measuring Facebook users' experience, perceptions of privacy and career impression management. This section presents the sample, methods, and measuring instruments.

3.1 Participants

The snowball sample comprised of 217 participants (mean age 35.93; age range 20-67; standard deviation =12.11; male 86; female 123); of which 181 participants were recruited on various social networking sites and 36 recruited as part of a company sample. Within the overall sample of 217, 188 participants were classified as Facebook users, defined as those participants that had a Facebook account (mean=34.93; range 20-67; standard deviation=11.26), and 28 were classified as non-users, i.e. not having a Facebook account. The following industry categories were identified within the sample: 4.37% law; 7.28% human resources; 6.31% consulting; 5.83% fulfilling support staff positions (receptionist, secretary); 8.74% finance; 2.43% service industry (make-up artist, chef); 3.88 % information technology; 17.48% in academic roles; and 1.94% in advertising.

3.2 Procedure and Instruments

An electronic survey link was posted on the researchers' Facebook wall (with a network of 800+ friends) consistently over the 4-week collection period and the link was sent through private inbox messaging. It was posted on different group forums on LinkedIn as well as on the researcher's profile, and it was tweeted on the researchers' Twitter profile. The link was emailed to all of the researchers' contacts. Lastly, the link was distributed by the Director of the Human Resource Department across the corporate body of a South African based IT company. Ethics Clearance was obtained from the University's Ethics Committee. The first point of the snowball sample was the researchers' own networks which raises possible biases in sampling. This was addressed by drawing participants from various sources including Facebook, LinkedIn, Twitter, email as well as within a company. Although the data is largely South African it raises possible generalizability concerns.

The survey comprised of a Facebook experience scale, perceptions of Facebook privacy scales (privacy awareness, work privacy, trust) and perceptions of Facebook career impression management scales (self-monitoring, work relations, workplace outcomes). It allowed both users and non-users to participate in the research by framing items in the active and passive form. Non-users scored zero (0) for the Facebook experience scale. This was done to ensure that the whole range of use/non-use was included.

The Facebook experience scale was made up of six items (exemplar: "On average, how many times a day do you use Facebook?"). The perceptions of Facebook privacy scales comprised of three privacy awareness items ("My privacy settings on Facebook are important"); three work privacy items ("It is acceptable for my superior to be my friend on Facebook"); and three trust items ("I trust that my privacy is secure on Facebook"). Facebook career impression management scales were made up of six self-monitoring items ("I monitor what I post on my Facebook profile"); three work relations items, ("I use my Facebook profile as a work tool, post my current employment, work achievements etc."); and six workplace outcomes items ("I manage my impression on Facebook because I am concerned that it may have negative consequences on my employability"). All the items were scored on a 5-point Likert type scale.

3.3 Analysis and Results

An initial pilot study (N=34; age 19-41) validated the developed scales. Cronbach Alphas for the scales were as follows: Facebook experience scale $\alpha= 0.91$; privacy awareness $\alpha=0.72$, work privacy $\alpha=0.71$, trust $\alpha=0.78$; self-monitoring $\alpha= 0.79$, work relations $\alpha=0.67$ and workplace outcomes $\alpha=0.83$ [16]. One-way frequencies, Ward's Cluster Analysis and Pearson's Correlation Coefficients were used to explore the nature of the relationships.

- ***Trends in Usage:*** Within the Facebook user sample, 67.37% reported that they had the Facebook application installed on their cell-phone; 65.75% of users reported that they instantly check their notifications; and the period between 12:00-18:00 was reportedly the time of the day Facebook was most used. This period corresponds to traditional work hours indicating that Facebook has penetrated and is prevalent in people's work life by virtue of its reach.

- ***Facebook Friend Strategy:*** The sample majority reported having less than 400 friends (65.89%), with the most commonly employed strategy (79.37%) being 'knowing someone personally' to accept them as a Facebook friend. Furthermore, 84.13% of the sample did not accept friends on the basis of being friends with their friends; 96.3% reported that they will not accept everybody who requests; and 17.46% of participants would accept friends they have heard of through others. Furthermore in terms of privacy controls, 80.63% of participants reported that they were familiar with privacy settings on Facebook; 82.98% reported that they have adjusted their privacy settings with 64.47% adjusting them to friends-of-friends; 55.15% of participants adjusted their settings on the basis of being generally private people, yet 61.05% reported that they did not adjust their settings on initial profile set-up; 14.98% reported that they customized their privacy settings; and only 13.33% of respondents reported that they had adjusted their settings for work related reasons. It is interesting to note that a mere 6.38% had their settings adjusted to 'friends only' even though the majority reported that they accepted someone on the basis of knowing someone personally. Furthermore, participant's network size was correlated negatively with their level of perceived Facebook trust. Results ($r = -0.18$; $p=0.0084$) indicated that as network size increased, the level of trust decreased. Although a weak effect size, this finding suggests that a smaller network size has the ability to facilitate trust.
- ***Perceived Social Gains:*** 67.03% of users reported that Facebook facilitated social interaction.
- ***User Clusters:*** The cluster analyses conducted identified five groupings of users based on Facebook activities and level of usage. Table 1 presents the clusters and depicts advanced, high, intermediate, low and non-users based on their overall usage. This table illustrates that advanced users were most engaged across the spectrum of Facebook activities. Advanced and high users would be categorized as active users with intermediate and low users being described as passive users.
- ***Correlations Analyses:*** Pearson's Correlation Coefficients were conducted to explore the relationships amongst the constructs. All of the relevant conditions for these parametric analyses were met. The results are reported in Table 2.
- ***Users vs. Non-Users:*** T-tests were conducted to assess the difference between users and non-users in terms of their level of trust. Results ($t=4.36$; $p<0.0001$) indicated that non-users ($m=10.76$) perceive higher Facebook trust than users ($m=8.81$). This corresponds to the finding that increased Facebook experience is related to decreased Facebook trust.

Table 1. User Cluster Proportions

Activities	Users (%) Advanced n=50	High n=43	Intermediate n=63	Low n=28	Non n=28
Connect	88	67	74	39	
Keep in touch	90	100	91	4	
Make new friends	24	19	7	4	
Job prospects	18	14	4	4	
Upload pictures	100	100	12	18	
Tag pictures	74	9	1	4	
Update status	80	21	16	4	
View others profiles	92	53	47	46	
Wall posts/ inbox messages	96	40	56	11	
Post links	52	14	9	4	
Alternative to email	74	16	22	0	
Use/develop applications	80	2	1	4	
Advertise	14	2	4	14	
Join groups	16	9	1	0	
Chat	50	12	28	4	
Events	72	30	13	11	
Sum Usage	9. 48	5.09	3.88	1.69	0

Table 2. Pearson's R Correlation Matrix (* Significant at ρ<0.05)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) F. Experience	1.00						
(2) Privacy Awareness	-0.01	1.00					
(3) Work Privacy	-0.02	-0.15*	1.00				
(4) Trust	-0.16*	0.13	0.11	1.00			
(5) Self-Monitor.	0.26*	-0.23*	0.09	0.03	1.00		
(6) Work Relations	0.23*	-0.01	-0.30*	-0.03	0.28	1.00	
(7) Work Outcomes	-0.02	-0.22	0.27*	0.99	0.53	0.06	1.00

The trends in usage point to Facebook's position in users' lives and its prevalence in the work context. In terms of friend strategy, the sample majority reported that they had adjusted their privacy settings and had a friend network of less than 400. Correlations indicated that there is increased trust in smaller networks and as network size increased, trust decreased. Non-users have greater trust in Facebook than users, corresponding to the finding that as Facebook experience increases trust decreases. The perceived social gains indicated that participants recognized the social value of Facebook, aligning with the premise of the Social Capital Theory. Users were clustered into five groups indicating that there are varied levels of usage. The correlation analyses, although weak, indicated that there are associations between how one uses Facebook, views privacy, and manages their work impression.

4 Findings and Discussion

Table 2 shows that although significant relationships found were weak, how one uses Facebook and ones' perceived value in presenting an image of an employable person is influenced by how one views their Facebook privacy. With greater Facebook experience being related to higher perceptions of self-monitoring and work relation, it is deduced that the more active one is on Facebook the more likely one is to perceive value in monitoring the image of an employable person and using Facebook to manage impressions within existing work relationships. A user will only present an image of themselves that they consider favourable as it facilitates and stimulates the feelings of self-worth [5]. This favourable image aligns with that of a desirable employee.

Furthermore greater Facebook experience was related to higher perceptions of trust such that the more active the user the less trusting. Trust can be viewed as the currency of social capital, which is acquired through mutual exchanges and interactions between people [5]. With Facebook stimulating interaction by way of mutual exchanges the way in which Facebook is used and the level of usage influences one's level of trust. The fostering of trust may also be influenced by one's friend strategy. 'Knowing someone personally' as a pre-requisite for addition to a network allows for a select and limited group of friends, stimulates greater trust within this network. Perceiving greater trust could also be influenced by one's network size, such that smaller networks facilitate a more trusting environment as exposure is more limited (majority of respondents friend networks were less than 400). Even though the majority of users did not limit their settings to friends only, having the ability to select who can access shared content builds trust.

One's perceptions of self-monitoring and workplace outcomes were related to low levels of privacy awareness (Table 2). That is, monitoring activities to present the image of an employable person and believing that Facebook can have real workplace outcomes are related to decreased levels of privacy protection and information disclosure. In this way, impression management is related to decreased disclosure of information as shared content is selected and managed on the basis of presenting the specified image. This shows that how one views ones privacy online ultimately affects how much information is shared and how impressions are managed.

High levels of work privacy were related to low levels of perceived work relations. Believing that work-associated people should not form part of one's friend network as employers monitor activities is related to low levels of using Facebook as a work tool. Thus having a network that is free from work-associated people means that profiles do not have to be managed in line with work. Furthermore, high levels of work privacy were associated with high levels of perceptions of workplace outcomes. The belief that employers utilise Facebook to monitor employees is aligned with perceiving that Facebook can affect ones potential for promotion, career developments and possible dismissal or disciplinary action. Viewing privacy in this way will affect what information is shared and how it is managed in order to emulate the image of an employable person. In conclusion, the focus and findings of this paper explored the nature of relationships that have a direct effect on work practices and people's lives. Facebook users who recognise the link between their work and Facebook lives derive social value from Facebook and simultaneously appreciate the value in presenting a favourable work image online.

References

1. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook "friends": Social Capital and college students' use of online social network sites. Journal of Computer- Mediated Communication 12, 1143–1168 (2007)
2. Smith, W.P., Kidder, D.L.: You've been tagged (Then again, maybe not): Employers and Facebook. Business Horizons 53, 491–499 (2010)
3. Rosenfeld, P., Booth- Kewley, S., Edwards, J.E.: Responses on computer surveys: Impression management, social desirability, Big Brother Syndrome. Computers in Human Behaviour 12, 263–274 (1996)
4. Buys, L., Bow, V.: The impact of privacy on social capital. In: Proceedings on the Social Change in the 21st Century Conference, Queensland University of Technology, Brisbane, Australia (2002)
5. Mehdizadeh, S.: Self- presentation 2.0: Narcissism and self- esteem on Facebook. Cyberpsychology, Behaviour and Social Networking 13, 357–364 (2010)
6. Foucault, M.: Discipline and Punish. Pantheon, New York (1977)
7. Dunn, H.B.: Facebook and other social networking sites: Candy-coating surveillance. Comm-entary: The University of New Hampshire Student Journal of Communication, 86–94 (2009), `http://www.unh.edu/communication/media/pdf/commentary/spring09.pdf#page=89` (retrieved April 25, 2011)
8. Boyd, D.: Facebook's Privacy Trainwreck: Exposure, invasion, and social convergence. Convergence: The International Journal of Research into New Media Technologies 14, 13–20 (2008)
9. Rosenberg, J., Egbert, N.: Online impression management: Personality traits and concerns for secondary goals as predictors of self- presentation tactics on Facebook. Journal of Computer- Mediated Communication 17, 1–18 (2011)
10. Barash, V., Ducheneaut, N., Isaacs, E., Bellotti, V.: Faceplant: Impression (Mis)management in Facebook Status Updates. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (2010)
11. Goffman, E.: The Presentation of Self in Everyday Life. Doubleday, New York (1959/1990)
12. Debatin, B., Lovejoy, J.P., Horn, A.K., Hughes, B.N.: Facebook and Online Privacy: Attitudes, behaviours, and unintended consequences. Journal of Computer- Mediated Communication 15, 83–108 (2009)
13. Boyd, D.M., Ellison, N.B.: Social network sites: Definition, history and scholarship. Journal of Computer- Mediated Communication 13, 210–230 (2007)
14. Liu, Y., Gummadi, K.P., Krishnamurthy, B., Mislove, A.: Analysing Facebook privacy settings: User expectations vs. reality (2011), `http://conferences.sigcomm.org/imc/2011/docs/p61.pdf` (retrieved October 12, 2011)
15. Dwyer, C., Hiltz, S.R., Passerini, K.: Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings at the Thirteenth Americas Conference on Information Systems. Colorado, Keystone (2007), `http://csis.pace.edu/~dwyer/research/DwyerAMCIS2007.pdf` (retrieved April 25, 2011)
16. Cronbach, L.J.: Essentials of psychological testing. Harper and Row Publishers, New York (1984)

Simplifying Remote Collaboration through Spatial Mirroring

Fabian Hennecke[1], Simon Voelker[2], Maximilian Schenk[1], Hauke Schaper[2], Jan Borchers[2], and Andreas Butz[1]

[1] University of Munich (LMU), HCI Group
80333 Munich, Germany
[2] RWTH Aachen University
52056 Aachen, Germany
{fabian.hennecke,andreas.butz}@ifi.lmu.de,
{voelker,borchers}@cs.rwth-aachen.de

Abstract. Even though remote collaboration through telepresence is supported by a variety of devices and display environments, it still has some inherent problems. One of these problems is the definition of a unified spatial reference system for the shared workspace in combination with an immersive representation of the collaborator. To mitigate this problem we propose a technique we call spatial mirroring. It is based on a virtual collaboration environment using two curved displays and aims to eliminate possible communication errors due to left/right misunderstandings. We explain the working principle and ideas behind spatial mirroring, and present two consecutive user studies in which we were able to verify its benefits.

Keywords: Remote collaboration, Telepresence, Orientation problem, Spatial reference, Mirroring.

1 Introduction

Over the last decade remote collaboration and telepresence have been widely researched and became available to many users (e.g. Skype[1]). A first vision of remote collaboration [8] was sketched in the Starfire video prototype [14]. Since then many projects integrated the remote collaboration system into the actual working environment of the user and created the illusion of being there [7, 11]. Recently the MirageTable [1] combined two differently-oriented display areas with a stereoscopic projection of a remote collaborator allowing three-dimensional collaboration. Beside this, additional horizontal displays have also been used to visualize a remote collaborator's interactions based on video overlays on an entirely shared workspace [6]. Even a video representation of remote collaborators on additional vertical displays all around a horizontal display was proposed [4] and its effect on collaboration was evaluated [14]. However, these approaches suffer from a problem, which can even arise in

[1] `http://www.skype.com`

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 624–631, 2013.

co-located collaboration settings: it is a priori unclear which of the different spatial reference systems of each collaborator someone refers to [16]. Collaborators can share the same point of view, which minimizes reference problems, but this also destroys the feeling of 'being there' in a remote collaboration scenario. Contrary to this the collaborators can be positioned on different sides of the shared workspace [10], which offers a good immersion but suffers from different spatial reference systems. While this problem can be mitigated with gazes and gestures in a co-located setting it is hard to properly transfer these additional collaboration information into a remote system. As speakers tend to use their own spatial reference system more often during collaborative tasks [12] this can become a problem during remote collaboration.

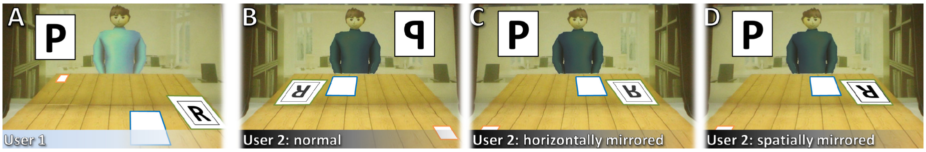


Fig. 1. Different views of the same table: (A) User 1: normal view, (B) User 2: normal view, (C) User 2: horizontally mirrored, (D) User 2: horizontally mirrored and content on the table additionally horizontally mirrored (spatially mirrored). Using a spatially mirrored visualization reduces misunderstandings about left and right and reduces selection times and error rates.

We propose a technique called spatial mirroring to overcome such spatial reference problems in remote collaboration environments. Spatial mirroring is basically a seamless combination of the two metaphors 'over a table' and 'through a glass window' described by Ishii et al. [5] and is based on the collaboration concept of the PerspectiveTable [3]. Based on two curved displays [17, 18] we utilize a double mirroring effect to align the spatial reference systems for both users in terms of left and right. Instead of 'my left' and 'your left' spatial mirroring creates 'our left' across the entire working area. In two consecutive user studies we were able to verify the benefits of spatial mirroring compared to a non-mirrored visualization.

2 Spatial Mirroring

A seamless virtual remote collaboration environment including differently oriented working areas allows for modifying the spatial reference system of a user without making this explicit or obvious. This way both users can refer to the same 'left' and 'right', while objects on the table still remain upside down for the opposite collaborator. (see Fig. 1.a and d). This avoids confusion and explanations about whose 'left' or 'right' they reference to.

Our setup for spatial mirroring is therefore based on a remote collaboration environment similar to the PerspectiveTable [3]. It is also based on two curved displays located in different cities. Using this setup spatial mirroring is realized with a mirroring along the central vertical axis of the entire video image using a simple pixel shader on one of the displays (see Fig. 1.b to c). While this is sufficient for content located vertically between both collaborators, an additional mirroring step is needed for content in the shared horizontal workspace to eliminate the problem of mirror-inverted appearance caused by the first mirroring (see Fig. 1.c). We are aware that this isn't the only possible combination of mirroring effects to achieve a spatial mirroring. But as we already used shaders to correct the image distortion caused by the curved displays we could implement the first mirroring step without much extra cost.

We opted for a virtual avatar like the PerspectiveTable [3] instead of a real video image as this allows us to make spatial mirroring go unnoticed. A mirrored video image of a known person can be easily identified. In case of a virtual avatar only the inverted handedness of the remote collaborator might reveal the mirroring.

3 Evaluation of Spatial Mirroring

To evaluate spatial mirroring we conducted two consecutive studies. The first study explored the basic effects on users' performance during simple selection tasks while the second study investigated the effects on actual collaboration. Participants of this second study solved a puzzle with a remote collaborator. The results of the first study show that users performed significantly better in terms of speed and error rate using spatial mirroring. The video analysis of the second study revealed that users also adapt to the mirrored reference system without problems. Both studies were conducted with curved displays of similar specifications to BendDesk and Curve [17, 18].

3.1 Study 1

In this study a virtual avatar at the opposite end of the remote collaboration table asked the participants to select one out of ten virtual playing cards. The cards' position was described relatively to the outer left or outer right cards using different instruction types (see Fig. 2.).

Participants and Design

19 participants took part in the study (9 females, 10 males, aged between 19 and 59, all right-handed). We used three different instructions (*audio*, *pointing* and *audio+pointing*) to describe the cards' position and two different visualizations (*normal* and *mirrored*) of the collaboration environment as within-group factors. The combination of these factors was counterbalanced with an 8×8 Latin square resulting in 64 actual selection tasks per participant. As both visualizations are the same for the instructions audio and pointing we interpreted the results of the normal visualization separately for both visualizations.

Task and Procedure

The participants were seated centered in front of the display. They had a training phase prior to the study where they were asked to select one out of two virtual playing cards using all instruction types of the study. There was no feedback if the selection was right or wrong during the entire study.

Instruction Type: We used three different instruction types: audio, pointing, and audio+pointing. For audio the participants were asked to select a particular card (“third card from the left”). In the pointing condition the avatar only pointed at that particular card while he combined both instructions in the audio+pointing condition (see Fig. 2.).

Visualization: Two different visualizations were used: normal and mirrored. In the normal condition the avatar referred to his left and right of the table, which could lead to a mismatch between audio and pointing in the combined condition (see Fig. 2.a). In the mirrored visualization the avatar referred to the mirrored spatial system and used left and right accordingly (see Fig. 2.b).

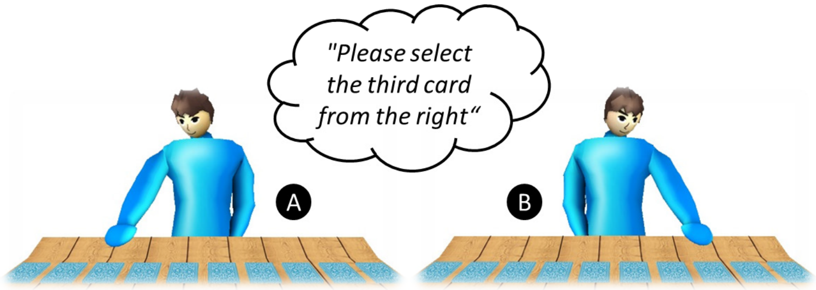


Fig. 2. Task design of our first study: A virtual avatar asked the participant to select a specific card from a row of ten cards using three types of instructions: audio, pointing, audio+pointing. The display was (A) not mirrord or (B) spatially mirrored across these three conditions.

Measures

We measured two types of objective data: selection time and errors. Selection time was defined as the time between the mention of the location information (end of the instruction) and the actual selection of a card by the participant via touch. It was also logged if the selection was correct. In the case of audio and pointing these measures were interpreted separately for normal and mirrored based on the same input data. After each session, participants answered an online questionnaire.

Statistical Tests and Analysis

Objective data was analyzed using a repeated measures analysis of variances (ANOVA). In cases in which the assumption of sphericity was violated, we used a Greenhouse-Geisser correction.

Results

As selection times for audio and pointing were identical due to the study design the selection times of the visualizations only differed for the audio+pointing instruction. Interestingly we still found a significant main effect of the visualization ($F_{1,18}$ = 12.628, $p < .05$) with mirrored (1953 ms) being faster than normal (2095 ms) across all instruction types. Besides instruction also had a significant effect on selection times ($F_{2,17} = 19.212$, $p < .05$). Post-hoc tests revealed significant differences between audio (2275 ms) and pointing (1642 ms) and also between audio+pointing (2156 ms) and pointing. Audio+pointing was even slightly slower than audio only in the normal visualization. We believe this is caused by the inconsistency between the audio and gestural instruction for audio+pointing in the normal visualization, which might have created a higher cognitive load.

There were also significant main effects of the visualization ($F_{1,18} = 222.493$, $p < .05$) with the spatial mirror (99%) being more correct across all instructions compared to normal (43%). This difference is caused by the large number of errors due to differing audio+pointing instructions in terms of left and right in the normal visualization – only 29% of these selections were correct while 100% of them were correct in the mirrored visualization. The instruction also had a significant effect ($F_{2,36} = 18.691$, $p < .05$) on the correctness of selection. Post-hoc tests revealed that pointing (100 %) performed best compared to audio+pointing (64 %) and audio (49 %) (all $p < .05$) across both visualizations. Besides we also found a significant interaction effect between visualization and instruction ($F_{1.074,19.340} = 82.899$, $p < .05$). Looking at each instruction separately shows that the mirrored visualization achieves nearly 100% correctness for all instructions while the normal visualization suffers from nearly 0% correctness in the audio only and 29% in the audio+pointing scenario.

3.2 Study 2

The second study consisted of a puzzle solving task with a remote collaborator similar to the study of Gergle et al. [2]. Two participants had to solve a 4×4 puzzle on the shared virtual table sitting remotely face-to-face while on screen was spatially mirrored to unify the collaborative working environment.

Participants and Design

16 participants (8 females, 8 males, aged between 21 and 31, 14 right-handed) were split into 8 teams (two participants per team). These teams were divided into two groups based on the visualizations already used in the first study: normal (no mirroring and no mentioning of the purpose of the study) and mirrored with an explicit explanation of the mirroring prior to the study. The remote collaborator was visualized with an animated virtual avatar based on Kinect tracking data for both groups.

Task and Procedure

The participants received a short introduction and training by solving a simple 3×3 puzzle on their own using direct touch. Afterwards the remote connection and an audio channel were established and the participants started to solve a 4×4 puzzle which showed an abstract tribal face in black and white (see Fig. 3). To enforce collaboration during the study one third at each end of the virtual table was only available to the corresponding user at that side. This way the users had to ask their collaborators for a specific tile if they could not access it.

Measures

We recorded each session at both displays in order to analyze the participants' behavior and interaction with the display. We counted the spatial descriptions of different puzzle tile positions during a session (left/right, front/back) and its according spatial reference system. We also counted pointing, description of the tile content and tile movement for each participant. Each description was also categorized into one of three categories: successful (e.g., participant received a tile after asking for it and describing its position), not-successful (e.g., collaborator moved the wrong tile), and descriptive (e.g., "I will drag this tile over here."). The participants answered an online questionnaire based on 5 point Likert scales after solving the puzzle.

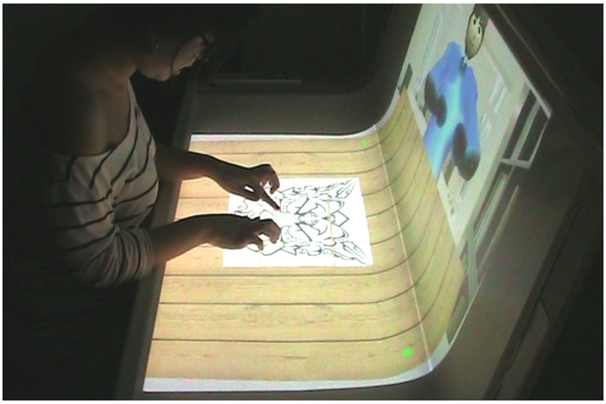


Fig. 3. A participant during the second study with a nearly complete puzzle and her remote collaborator in front of her

Statistical Tests and Analysis

Due to the limited number of participants and the subjective video analysis we didn't run statistical tests with the data and use frequencies instead to report our findings. They balance numeric differences in terms of left/right references, which might occur due to the explanation prior the mirrored condition compared to the normal condition. We combined the subjective ratings of the participants into 3 bins: 'Agree' ('1' & '2'), 'Neutral' ('3'), and 'Disagree' ('4' & '5').

Results

We recorded a total of 117 pointing and movement references and descriptions during the study. Some Participants also pointed at tiles or moved them to describe other tiles next to them. This happened to similar extents in both groups and therefore doesn't constitute a difference.

Normal: Surprisingly only 20% of the participants used their own reference system for left/right instructions, of which 88% were successful. This is possibly due to the rather unfamiliar remote collaboration setting [12]. In most cases they added the respective reference system (e.g., "on your left side") to the instruction and if they didn't their collaborator asked. Interestingly only 50% stated afterwards that they mirrored left and right in order to describe a tile's position. In terms of front and back, 65% used their reference system, which worked quite well (67% success, only 25% reported communication problems regarding “in front” and “in the back”).

Mirrored: All participants of this group used the unified spatial reference system regarding left and right: 62% were requests (88% successful) and 38% were descriptive. We found no usage of the pre-mirrored collaborator's reference system in this group. The self-assessment of the participants backs these results: 88% stated that they didn't have problems to determine what was ‘left’ and ‘right’. Nevertheless, 25% reported communication problems regarding ‘left’ and ‘right’ after the study. 80% of the front/back references were made according to the users' own point of view; they were mainly used for descriptive purposes (71%). However participants reported communication problems (25%) and problems to determine what is in the front and back (50%) after the study.

Further Findings: We noticed that subjects used an object's relative position between other objects to describe its position during the entire study. Some participants even used the display shape as a reference (e.g., “in the curve”). Contrary to our expectations participants only used a visual description of specific tiles after having a rough idea of the entire image and didn't describe specific visual properties of a tile. Finally, 88% agreed that a uniform point-of-view can help to avoid problems.

4 Summary and Conclusion

We presented spatial mirroring as a technique to avoid misunderstandings during remote collaboration about different interpretations of left and right. We implemented a spatial mirror based on a three-dimensional virtual collaboration environment similar to the Perspective Table [3]. We conducted two consecutive user studies to gather first insights on spatial mirroring. The first study showed that a spatially mirrored visualization not only allows faster selections but also minimizes selection errors. We assume that the faster selection times are due to a lower cognitive load during audio+pointing instructions as the instructions do not conflict during the mirrored visualization. The same applies for the selection error: The difference between the instructions for audio+pointing in the normal visualization led to wrong selections, while the mirrored visualization was completed nearly without errors. The second study involving a more realistic collaboration showed that a spatial mirror can help to ease spatial descriptions between collaborators during remote collaboration.

Based on these results we conclude that spatial mirroring can be used to ease remote collaboration by providing a unified spatial reference system. It reduces selection times due to fewer spatial reference conflicts and helps to avoid communication problems caused by misunderstandings about left and right. Nevertheless, one has to keep in mind that the position of the content within a document on the horizontal working area is also horizontally inverted to avoid mirror-inverted content. It will be interesting to investigate if and how this affects remote collaboration with text-based content [8]. Beside this allowing to fluently change the seating position around the virtual table could help to enhance the feeling of ‘being-there’. Finally the generalization to more than two users will be very interesting as it can be achieved by modifying the avatar movements to fit the virtual seating arrangement in each case but also opens up a lot of new questions about unified spatial references for all collaborators.

References

1. Benko, H., Jota, R., Wilson, A.: MirageTable: freehand interaction on a projected augmented reality tabletop. In: Proc. of CHI 2012, pp. 199–208 (2012)
2. Gergle, D., Kraut, R., Fussell, S.R.: Language efficiency and visual technology: Minimizing collaborative effort with visual information. Journal of Language and Social Psychology 23, 491–517 (2004)
3. Hennecke, F., Voelker, S., Schenk, M., Schaper, H., Borchers, J., Butz, A.: PerspectiveTable: Blending Physical and Virtual Collaborative Workspaces. In: Workshop 'Blended Interaction: Envisioning Future Collaborative Interactive Spaces' in Conjunction with CHI 2013 (2013), http://hci.uni-konstanz.de/blendedinteraction2013/
4. Hirata, K., Harada, Y., Takada, T., et al.: The t-Room: Toward the Future Phone. NTT Technical Review 4(12), 26–33 (2006)
5. Ishii, H., Kobayashi, M.: ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In: Proc. of CHI 1992, pp. 525–532 (1992)
6. Junuzovic, S., Inkpen, K., Blank, T., Gupta, A.: IllumiShare: sharing any surface. In: Proc. of CHI 2012, pp. 1919–(1928)
7. Kauff, P., Schreer, O.: An immersive 3D video-conferencing system using shared virtual team user environments. In: CVE 2002, pp. 105–112 (2002)
8. Matsushita, M., Iida, M., Ohguro, T.: Lumisight Table: A Face-to-face Collaboration Support System That Optimizes Direction of Projected Information to Each Stakeholder. In: Proc. of CSCW 2004 (2004)
9. Minsky, M.: Telepresence. OMNI
10. Olson, M.H., Bly, S.A.: The Portland experience: a report on a distributed research group. Computer-supported Cooperative Work and Groupware 34(2), 81–98 (1991)
11. Sadagic, A., Towles, H., Holden, L., Daniilidis, K., Zeleznik, B.: Tele-immersion Portal: Towards an Ultimate Synthesis of Computer Graphics and Computer Vision Systems. In: 4th Annual International Workshop on Presence (2001)
12. Schober, M.F.: Spatial perspective-taking in conversation. Cognition 47, 1–24 (1993)
13. Schwarz, T., Hennecke, F., Lauber, F., Reiterer, H.: Perspective + Detail – a visualization technique for vertically curved displays. In: Proc. of AVI 2012, pp. 485–488 (2012)
14. Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J., Buxton, B.: Three's company: understanding communication channels in three-way distributed collaboration. In: Proc. CSCW 2010, pp. 271–280 (2010)
15. Tognazzini, B.: The "Starfire" Video Prototype Project: A Case History. In: Proc. of CHI 1994, pp. 99–105 (1994)
16. Tversky, B., Hard, B.M.: Embodied and disembodied cognition: Spatial perspective-taking. Cognition 110(1), 124–129 (2009)
17. Weiss, M., Voelker, S., Sutter, C., Borchers, J.: BendDesk: Dragging Across the Curve. In: Proc. of ITS 2010, pp. 1–10 (2010)
18. Wimmer, R., Hennecke, F., Schulz, F., Boring, S., Butz, A., Hußmann, H.: Curve: Revisiting the Digital Desk. In: Proc. of NordiCHI 2010, pp. 561–570 (2010)

Social Influence from Personalized Recommendations to Trusting Beliefs of Websites: Intermediate Role of Social Presence

Yanan Li[1,2] and Yong Wang[1,*]

[1] Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, 100101 Beijing, China
[2] University of Chinese Academy of Sciences, 19 Yuquan Road, 100049 Beijing, China

Abstract. With the increasing prevalence of online shopping, many companies have added "personalized recommendation" modules on their websites to collect customers' click-stream data in real time and conduct relevant analysis, which will both assist the decision-making of the web owners and enrich the interactive experience of the customers. A website with more personalized recommender system allows users to experience others as psychologically present being, conveys a feeling of human warmth and sociability which is believed critical for the formation of consumers' trust towards online stores. In this study, a laboratory experiment was conducted to empirically examine the effects of the level of perceived personalization on consumers' trusting beliefs towards online store. The result also demonstrated that the perception of social presence can partially mediate the effect of perceived personalization on consumer's trusting beliefs.

Keywords: Personalized Recommendation, Personalization, Social Presence, Trusting beliefs.

1 Introduction

Web interface, the bridge between websites and consumers, not only undertakes the practical task, but also works as an emotional bond, bringing unique experience to form a social connection between websites and customers. Websites with personalized recommender system, with Amazon.com as the representative, learn consumers' preferences based on the click-stream data, so as to achieve a personalized presentation. Personalized recommending helps consumers to screen items they might be interested in from numerous commodities, and reduces information overload and increases choice efficiency. At the same time, it realizes the goal of persuading consumers by influencing their choice and preference [1, 2].

Apart from the practical functions mentioned above, the scenario of a one-on-one conversation is created between consumer and website during the recommending process, thereby generating a feeling that he/she is "being with other people". This is

[*] Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 632–639, 2013.

the foundation for the forming of the perception of social presence. Many investigations show that social presence is a key factor influencing the consumer trust towards the website [3, 4]. Previous studies, however, rarely discuss how it is affected by personalized recommendations. This study aims to investigate the effect of personalized recommendation on the consumers' trusting beliefs towards the website and the role played by the perception of social presence in this process.

2 Literature Review

2.1 Personalized Recommendation

Personalized recommendation is a representative of personalized service [5]. Most early recommendations are generated by asking about customers' preferred commodity attributes in advance. As the recommendation algorithm keeps improving and data mining technology is making advances, personalized recommender system as represented by Amazon.com can achieve real-time indirect conjecture on consumer demand based on click-stream data. Compared with earlier strategy, this context-based recommendation strategy moves closer towards personalization.

A personalized recommender system helps in alleviating information overload, improving choice efficiency, and enabling cross- and up- selling. Besides, it also influence consumers' cognitive processing and decision-making [6], their evaluation [7] and satisfaction, loyalty and trust to the website.

Researchers have been devoting to increase the accuracy of recommending algorithm. The bottleneck in this process is mainly associated with the constructive nature of consumers' preference. While evaluating personalization level, there is no baseline website to compare against, regardless of the recommendation method used. Consumers' perception of personalization varies depending on the tasks, context and time. So personalization is not a technical term, but a psychological variable. According to psychological theory, the information more relevant to the consumers' goal will result in more perceived personalization and deeper cognitive processing, which is the essence of web personalization [7].

2.2 Perception of Social Presence

The notion of social presence refers to that, in the process of communication, the media of communication gives the individual a social feeling that he/she is "being with other people". In its nature, the perception of social presence is closely related to two psychological concepts: Intimacy and Immediacy.

In earlier studies, the factors identified as triggering the perception of social presence include emotional texts, images, personalized greetings, as well as information technology(IT)-enabled support for the provision of live chats, forums, consumer reviews, live help, and virtual reality technology, etc. It has been revealed that the perception of social presence causes more positive assessment on the efficacy of website, the satisfaction, pleasure experience and intention to use [8]. However, few researchers investigate the effect of perceived personalization on the perception of social presence [8].

2.3 Trusting Beliefs towards Websites

Unlike traditional context, shopping online seems to be deficient in human touch. It is more difficult but more important to build consumer's trust towards online store. McKnight, Choudhury and Kacmar [9] believe that, in the context of online shopping, the trust towards a website consists of trusting beliefs and trusting intention. Trusting beliefs, reflecting consumer's concrete cognition and emotional belief towards the seller, can directly lead to trust intention and trust-related behavior.

In previous studies on user adoption of personalized recommendation, trust is a key psychological variable. Personalized recommendations can facilitate the formation of trust by improving the transparency, perceived control and internalization.

Some studies suggest that personalization can result in emotional sense or affective experience, like feelings of intimacy, perception of social presence [1], which can improve trusting beliefs towards online store. Therefore, it is a new trend of research to put emphasis on the emotional bond between the consumer and website, and on the formation of trust from the social aspect.

3 Hypothesis Development

Kumar and Benbasat [7] are among the first to have discussed the difference between personalized and non-personalized websites in creating users' perception of social presence. They filtered Amazon.com's contents in real time in laboratory experiment, and found that the personalized recommender service in Amazon.com improves its social presence. The study conducted by Choi et al [8] manipulated the recommendation pages with varying levels of personalization according to the connections between the recommendation source and the consumer, in order to explore the influence of products recommendations with different personalization levels on the perception of social presence and intention to use. The results indicated that, compared with non-personalized recommendation (for example, ranking lists), users experienced a higher level of social presence in personalized recommendation page.

We believe that the higher the personalization level perceived by consumers from a website, the more likely the consumers feel involved in a one-to-one dialogue between him/her and the website. The intimacy and sociability easily create a warm feeling, i.e. the perception of social presence. Hence we make the following hypotheses:

Hypothesis 1: Recommendation webpages with higher personalization level strengthens the consumer's perception of social presence.

The content, presentation, and timing affect and what the consumer feels about the personalization level of the recommendation, influence his/her judgments on the website. Previous studies mostly focus on the recommendation agencies which acquired information of user preference explicitly. It shows that consumers develop higher cognitive trust and emotional trust towards the website when their perception of personalization is higher [10]. Moreover, the stronger the perception of personalization, the more precise, in the consumer's point of view, the website's conjecture on his/her preferences would be, which enhances his/her trusting beliefs towards the website. Hence,

Hypothesis 2: Recommendation webpages with higher personalization level provoke stronger trusting beliefs from the consumers.

Through previous studies, it is not difficult to conclude that the perception of social presence is a key intervening variable in the process of users reading through the web interface elements to affect the user trust [3, 4]. In this study, further test is made to see whether user's perception of social presence will mediate the influence of perceived personalization on trusting beliefs towards the website. Hence,

Hypothesis 3: The perception of social presence mediates the influence of personalization on trusting beliefs.

4 Research Methodology

The task for the experiment consisted of asking participants to imagine that he/she is going to do some information searching for what he/she needs to buy, then browse web pages of a prototype shopping website named “Orangeebuy.com”. The study was designed as a one-factorial experiment manipulating two levels of perceived personalized in two independent groups. Participants were randomly assigned to the two groups, where each participant was exposed to only one level of stimulus or condition.

The prototype site was designed in terms of content and page layout to resemble typical e-commerce sites, such as Amazon.com. Product introduction page displayed basic product information of one item, followed by a set of recommendations—8 similar products—preceded by the words “products recommended”. For each condition, three product introduction pages involving three items, namely, headphone, TOEFL book and walnut, which were chosen in earlier studies [3, 8], were displayed. In order to confirm that the three items were within the scope of online shopping for college students in Beijing, an online questionnaire survey of about 200 undergraduates in Beijing was conducted. The results showed that books, digital devices and snack food were the top three product categories they most frequently purchased, and we also identified the commodity properties of the three items (price, brand, etc.) that the students were mostly concerned about when purchasing online.

All conditions, featured the same information of product introduction, but differed only in terms of recommendations. Following the strategy of “goal-relevance”, which is considered as the major feature of e-personalization [2], participants in $Group_{High}$, received more relevance recommendations to the product being introduced on current page based on the commodity property identified before than those in $Group_{Low}$.

Following the completion of the task in the main experiment, all items were statements and subject responses were offered on a 7-point Likert scale ranging from Strongly Agree(1) to strongly disagree(7) . All the measuring scales were previously validated, including scale for Trusting beliefs [4, 9], scale for perceived personalization [10], and scale for social presence [3] and measuring for demographical profiles. Each participant would get a reward of 10 RMB and some candies.

5 Data Analysis

A total of 114 undergraduates successfully completed the main experiment. Among them, 55 are arranged in $Group_{High}$ and 59 in $Group_{Low}$. All participants were recruited from universities in Beijing, China, including 39 males and 75 females with an average age of 23.5 years old. All participants reported that they had online shopping experience.

An independent-samples' t-test shows that the average score of perceived personalization between $Group_{High}$ and $Group_{Low}$ is significantly different ($Mean_{Low}$ =4.76, $Mean_{High}$ =5.65, t=5.172, $p<0.001$), which means that the experimental manipulation is successful.

We perform confirmatory factor analysis using AMOS 18.0 on the measurement model to assess internal consistency reliability, convergent validity, and discriminant validity. Trusting beliefs are represented as a second-order reflective construct, with benevolence, competence, and integrity beliefs represented as first-order reflective constructs [9]. Following the procedures recommended by Gorsuch, a first-order confirmatory factor analysis is used to confirm the structure of the three dimensions of trusting beliefs. The result shows, χ^2 / df = 1.782, NNFI = 0.906, CFI = 0.955, RMSEA = 0.083; the correlation coefficient of three sub-dimensions is 0.4~0.7 (0.65, 0.65, 0.48). In second-order analysis, the model fitting indexes are also acceptable. But the correlations between second- and first-order factors are stronger (0.66, 0.67, 0.96). Therefore, Trusting Beliefs as a second-order reflective construct can sufficiently express the concept connotation with a convenience of model simplification by averaging items from three sub-dimensions scales to represent the Trusting Beliefs.

A reliability test is conducted with an internal consistency test (Cronbach's alpha). The measurement items used in this study are shown to be reliable in that Cronbach's alpha values for all dimensions are at least 0.7. Then, the convergent and discriminant validity of the constructs are assessed and suggested to be acceptable as described in Table 1 and 2.

To test the hypotheses, we test the perception of social presence and trusting beliefs between the two groups using One-way ANOVA. It is found that significant difference exists (Social Presence: $Mean_{Low}$=3.97, SD_{Low}=0.96, $Mean_{High}$=4.68, SD_{High}=1.19, $F_{social\ presence}$=11.923, *sig.* = .001; Trusting Beliefs: $Mean_{Low}$=4.49, SD_{Low}=1.02, $Mean_{High}$=5.12, SD_{High}=0.71, $F_{trusting\ beliefs}$=13.992, *sig.* = .000). Thus, **Hypothesis 1 and Hypothesis 2 are confirmed**. That is, recommendation webpages with higher personalization level can create a stronger perception of social presence and trusting belief.

In all of the regression analyses that follow, we first enter a series of control variables before testing Hypothesis 3. These control variables include gender, age, education, hours spent on Internet per day, and shopping frequency in recent three months.

To test for mediation, we use the following four-step procedure outlined by Baron and Kenny (see Table 3): Model 1 shows that Perceived Personalization is significantly related to Social Presence (β=0.548, p<0.001). Model 2 suggests that Perceived Personalization is significantly related to Trusting beliefs (β=0.608, p<0.001). Model 3 shows that Social Presence is significantly related to Trusting beliefs (β=0.502, p<0.001). Support for the mediation hypothesis would be observed if the initially significant relationships we find between Perceived Personalization and Trusting beliefs disappear or decrease after we add Social Presence into the regression equation. Model 4 indicates that after Social Presence is added to the regression model, the initially significant relationship between Perceived Personalization and Trusting beliefs does not disappear (β=0.440, p<0.001). This result suggests that Social Presence only partially mediates the relationship between Perceived Personalization and Trusting beliefs. Thus, **Hypothesis 3 is confirmed**.

Table 1. Results of Reliability and Convergent Validity Analysis

Items	Standardized loadings	Cronbach's alpha	Items	Standardized loadings	Cronbach's alpha
PP1	0.755	0.778	TBb1	0.709	0.808
PP2	0.889		TBb2	0.803	
PP3	0.600		TBb3	0.790	
SP1	0.791	0.877	TBI1	0.875	0.915
SP2	0.737		TBI2	0.903	
SP3	0.689		TBI3	0.879	
SP4	0.867				
SP5	0.784				
TBc1	0.773	0.851	Trusting Belief Dimension		n/a
TBc2	0.826		Competence	0.884	
TBc3	0.843		Benevolence	0.753	
TBc4	0.661		Integrity	0.600	
χ^2 / df= 1.743 (df=129) CFI=0.923 NNFI=0.841 RMSEA=0.081					

Table 2. Average Variance Extracted (AVE) and Construct Correlations

	AVE	PP	SP	TB
Perceived Personalization(PP)	0.797	**0.892**		
Social Presence(SP)	0.882	0.565	**0.939**	
Trusting Beliefs(TB)	0.949	0.930	0.811	**0.974**

Note: Square root of AVE in bold on the diagonal. The AVE for each construct was > .5 as recommended. The square root of the AVE for each construct was also > the correlations with other constructs as recommended.

Table 3. Hierarchical regressions for mediation tests (N=114)

Variable	Model 1	Model 2	Model 3	Model 4
	Social Presence	Trusting Beliefs	Trusting Beliefs	Trusting Beliefs
Control				
Gender	-0.041	0.135	0.180	0.138
Age	-2.09	-0.213	-0.054	-0.153
Education	0.009	0.023	-0.049	0.022
Hours on the internet per day	-0.065	-0.13	0.009	0.006
Shopping Frequency	0.125	-0.05	-0.063	-0.084
R^2	0.232	0.037	0.193	0.193
Perceived personalization	0.548***	0.608***		0.440***
R^2	0.529***	0.452***		
Social Presence			0.502***	0.311***
R^2			0.624***	0.745***

* $p<0.05$ ** $p<0.01$ ***$p<.001$

6 Discussion

The experimental results show that, the recommending pages with two different levels of personalization arouse different perceived social presence and trusting beliefs towards online store. The higher the level of personalization that the users perceive, the stronger the perception of social presence, and the greater the trust towards the online store the consumers would have. The mediating effect of social presence is significant for the design of recommendation system.

Theoretical contributions of this study would be as follows. First, for users, personalization is a kind of psychological perception and evaluation of goal relevance. The greater the relevance to self needs, the higher the level of personalization he/she perceives. Second, personalized recommendations can directly affect trusting beliefs towards online store, and also affects indirectly by improving users' emotional experience. Third, perception of social presence is crucial for trust building process. Past researches mainly focus on improving the credibility of recommendation in order to enhance trust, while this study adds the social and affective influence of personalized recommending.

This study has several limitations. Our samples are composed only of undergraduate students. Future research should include participants within a more diverse range of age, online shopping experience and computer experience. And the personalized recommendation concerned in this paper is not a real one generated by a computer system. Therefore, the ecological validity of the experiment is affected.

Nonetheless, this study has already demonstrated the importance of personalized recommendation in terms of social impact. Managers who provide personalized recommendations on their online store should give concern to users' perception of social presence generated in the interaction with website. In addition, personalized recommendations might have different effects in combination with some social network elements and other environmental or psychological factors that might influence the social impact of personalized recommendations, which are all worthwhile for further investigation in the future.

Acknowledgements. This study was partially supported by the Knowledge Innovation Project of the Chinese Academy of Sciences (No. KSCX2-EW-J-8, 2011~2013), and the Special Fund for Beijing Key Discipline Construction (Applied Psychology, 2008~2012).

References

1. Liang, T.P., Li, Y.W., Turban, E.: Personalized Services as Empathic Responses: the Role of intimacy. In: PACIS vol. 73 (2009)
2. Tam, K.Y., Ho, S.Y.: Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes. MIS Quarterly 30, 865–890 (2006)
3. Gefen, D., Straub, D.W.: Consumer Trust in B2C E-Commerce and the Importance of Social Presence: Experiments in E-Products and E-Services. Omega 32, 407–424 (2004)
4. Hess, T., Fuller, M., Campbell, D.: Designing Interfaces with Social Presence: Using Vividness and Extraversion to Create Social Recommendation Agents. Journal of the Association for Information Systems 10, 889–919 (2009)
5. Tam, K.Y., Ho, S.Y.: Web Personalization As A Persuasion Strategy: An Elaboration Likelihood Model Perspective. Information Systems Research 16, 271–291 (2005)
6. Xiao, B., Benbasat, I.: E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. MIS Quarterly 31, 137–209 (2007)
7. Kumar, N., Benbasat, I.: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. Information Systems Research 17, 425–439 (2006)
8. Choi, J., Lee, H.J., Kim, Y.C.: The Influence of Social Presence on Evaluating Personalized Recommender Systems. In: PACIS (2009)
9. McKnight, D.H., Choudhury, V., Kacmar, C.: The Impact of Initial Consumer Trust on Intentions to Transact with A Web Site: A Trust Building Model. Journal of Strategic Information Systems 11, 297–323 (2002)
10. Komiak, S.X., Benbasat, I.: The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. MIS Quarterly 30, 941–960 (2006)

Social Network Users' Religiosity and the Design of Post Mortem Aspects

Cristiano Maciel[1,*] and Vinicius Carvalho Pereira[2]

[1] Instituto de Computação
[2] Instituto de Linguagens
Universidade Federal de Mato Grosso, Av. Fernando Correâ da Costa, nº 2367, Bairro Boa Esperança. Cuiabá – Mato Grosso - Brazil - 78060-900
cmaciel@ufmt.br, vinicius.carpe@yahoo.fr

Abstract. Social networks increase the challenges of designing real-world aspects whose computational abstraction is not simple. This includes death and digital legacy, strongly influenced by cultural phenomena, such as religion. Therefore, it is important to analyze youngsters' concepts of death in the web, as the Internet Generation outnumbers other groups of social network users. Besides, due to their age, many of them face other people´s death for the first time on the web. This paper analyzes to what extent these users' religion and the belief in afterlife may signal guidelines for a social network project that considers volition towards digital legacy. The data herein analyzed qualitatively and quantitatively come from a survey-based research with Brazilian high school students. The contributions for Human-Computer Interaction (HCI) studies comprise design solutions that may consider aspects of religion, death and digital legacy, also improving users' and designers' understanding on these issues in system design.

Keywords: social networks, Internet Generation, post mortem, religion.

1 Introduction

In the context of the diverse technological changes we face, especially the large use of social networks, many questions arise and make users and researchers think of their impacts on society. This challenge is bigger for software programs in which users connect, communicate and share content, because there are some real-world facts that migrate to the web and are difficult to undergo computational abstraction. Among these facts, users' post mortem aspects may be hard to design.

Being born, growing up and dying are natural events in human beings' life cycle, but things are not that simple on the web. The misalignment between life and death in the "real" and the digital worlds makes it harder to design post mortem aspects for social networks [13]. Software brings us to a new scenario, in which human beings can become immortal, even if involuntarily.

* Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 640–657, 2013.

Considering the possibility of immortality on the web, which involves bereavement [3] [17], posthumous interaction [15] and the preservation of users' memory [10], it is also important to discuss the user's volition about the destination of his digital legacy, so as to plan design solutions and provide legal basis for digital wills.

By means of a research with youngsters from the Internet Generation, [15] studied how these users' social representations of death influenced their concepts and practices of posthumous interaction. This research helped to design applications that considered post mortem aspects, by discussing, among other issues: a) which resources in dead people's social network (SN) profiles were most frequently accessed by these youngsters; b) what could be done to their SN accounts after their death, such as blocking some functionalities or password forwarding; and c) whether they would like their accounts to be removed from the SN after death. Other data therein collected are important to the present research, such as the fact that most of those youngsters had got in touch with someone's death for the first time in social networks.

However, other cultural representations, such as religion and funeral rites, are also related to human interactions facing death, which makes it important to study these aspects in a research on digital legacy. As every culture is a sign system, and human practices are systems of signification [9], one must investigate how these phenomena give new meaning to users' ideas about death and legacy. Therefore, we intend to understand in this research to what extent users' religious diversity, different belief systems and opinion about post mortem life can indicate guidelines for a SN project that consider users' volition, i.e. conscious choice, regarding their digital legacy.

In this research, we analyzed data from surveys answered by youngsters from the Internet (or Y) Generation, born between 1977 and 1997[22], who study in a Brazilian state school. The data we collected are interpreted considering previous researches and theoretical background on culture, religion and death. According to data from the Pew Internet research institute [18] on SN users in the United States, Facebook was used by 66% of Internet users (63% of the men and 70% of the women) in 2012, followed in terms of popularity by LinkedIn (20%) and Twitter (16%), although the latter two networks are most used by young adults. On the other hand, 83% of Facebook users are between 18 and 29 years old. In Brazil, in 2011, data from the Brazilian Internet Steering Committee (Comitê Gestor da Internet no Brasil - CGI) [6] also revealed that in 2011 83% of young people aged 16 - 24 use social networking sites, such as Orkut, Facebook and LinkedIn.

For the HCI field, we contribute in terms of design solutions that may consider religion, death and digital legacy, so as to deepen users' and designers' understanding on the role of death in system projects. From an epistemological point of view, bringing religion as a cultural aspect to HCI discussions is an attempt to study, from a computational perspective, another lens through which humans see reality and their interactive processes. In this research, religious diversity is regarded as an analysis variable, although science and religion are traditionally considered opposite areas. Thus, we have adopted Einstein's investigative paradigm: "science without religion is lame; religion without science is blind".

This paper is structured as follows: after this introduction, we present some theoretical background on the main issues discussed herein (section 3). Then, the research methodology is described. In section 4, the case study is presented and discussed. Finally, the conclusions and references are enlisted.

2 Theoretical Background

In order to analyze post mortem digital legacy, besides identifying the state of the art in this area, it is necessary to study culture, religion and death in our society.

2.1 Culture, Religion and Death

Religion is an inextricable part of life in society. [8], one of the first sociologists who studied religion scientifically, ascribes to religious practices the first human representation and symbolization systems, which determine abstract notions of time, space and causality. According to this sociologist, religion is a social fact, which makes it a common phenomenon to all human communities.

This idea of religion as a human practice emerging from social interaction is closely related to Geertz's view [9]: religion as a cultural system. In this context, the concept of system also applies to Anthropological Semiotics, which understands every culture as a sign system, and human practices as systems of signification.

The same definition of culture is employed by researchers on Semiotic Engineering, in the HCI field [20], to discuss projects and design solutions that consider cross-cultural aspects in computer-mediated communication, either between users or among users and designers. Therefore, among many definitions of religion (profane and sacred ones), this research spouses Geertz's view, as it is widely spread in HCI discussions and may bring some light to system projects facing digital legacy and cross-cultural religious diversity.

The connection between religion and death goes far beyond the comfort that sacred matter can bring to bereavers: religious funereal rites found culture [19], no matter if they are festive or mournful. This means that men abandon their state of nature and humanize in contact with someone's death, by waking, celebrating or mourning the corpse, i.e. ascribing meaning to a body that ran out of biological function. Then, the corpse has become a pure support for symbols, interface where men place values, feelings and material objects. Regarding these farewell rituals, we can see how human beings try to keep in connection to the dear ones that are no longer alive, driven by their imaginary and using symbols such as candles and flowers.

Actually, "death is a natural and non-classist fact. But it is also a social and cultural fact. As a social fact, it is stratified; as a cultural fact, it is intertwined with values and meanings" [21]. This set of cultural meanings, organized into religious systems, works as guidelines for people in their interactive practices with the dead: oblations, prayers, bereavement or bequeathment. People's beliefs determine who can interact with the dead, when, why and how to do it. [21] reinforces this idea by stating that "death entails the development of exclusion mechanisms, and their frontiers in different cultures are established by the limits between the sacred (the world of the dead) and the profane (the world of the alive). This duality lies underneath the collective imaginary and creates an invisible world parallel to the events of life.

In the Brazilian context, the three religions with greater amounts of believers are Catholicism (64,62% of the population), Protestantism (22,16%) and Spiritualism (2,01%), according to the data from the 2010 census [11], which quantified also the number of people with no religion (7.6%) and atheists (0,3%). In this discussion, it is not up to us to question God´s existence or post mortem life; instead, we focus on the fact that religion, as an element of culture, has its own meanings, possibly influencing the use and design of social software.

Based on these pieces of information and on the fact that most participants in this research are either Catholics or Protestants, it is necessary to understand how these religions deal with death. As they are both Christian religions, with a common historical background, Catholicism and Protestantism bare resemblance to each other when it comes to the understanding of death as a passage to an eternal life, denying reincarnation. However, whereas Protestants only believe in Heaven or Hell as afterlife destinies, Catholics consider also a third option: Purgatory. In terms of ritualism, the ceremonies of wake and burial in these religions are similar, although many Protestant groups do not employ candles as Catholics. Furthermore, whereas Catholics pray for the soul of the dead, Protestants tend to pray for those who are alive[7]. After the burial or cremation, it is common for Catholics to celebrate requiem masses seven days, one month and one year after the day of the death; on the other hand, such rituals do not exist in Protestantism[7]. This difference is related to the existence of a specific ritual for death among Catholics: extreme unction.

Some research has been done on religion and technology in the area of HCI. According to [25], although the design-oriented research examining religion is growing, only a handful of the world's major faiths have been studied. The uses of technologies in religious practices are also studied by [26], who presented results about the uses of technologies in three aspects of religious work: religious study and reflection, church services, and pastoral care. In another study, more focused on HCI, [24] analyzed how topics of interest to HCI researchers (e.g., material artifacts, routines, and Information and Communication Technologies use) are used for religious purposes. In their related works section, these authors presented an overview of researches on similar issues, grouped into three major categories: religion online, design explorations, and studies of appropriation. The aforementioned studies present interesting contributions about the relation between technology and religion, but do not specifically address this topic from the perspective of death and its taboos.

2.2 Related Works

In the HCI field, researches on this subject are recent and study the relation among death, mortality, bereavement and technologies. Particularly [16] have been doing research on HCI and the end of life in a lifespan-oriented approach. Other researchers, such as [3], investigate users´ behavior in social networks, based on their posts to dead people. In Brazil, researches have been done on post mortem data management.

In this sense, according to [13], social software must comprise a volition element, that is, software planning must consider the user's decision on the destination of his digital legacy. In his research, the author presented some possibilities that were considered in planning design solutions for posthumous digital legacy, such as:

1. attributing password power to third parties, while alive or in a will, attesting the wish for posthumous interaction. It is important to notice that password forwarding is a service that some specialized enterprises offer, studied by researchers like [4].
2. having a back-up copy of the digital legacy in some equipment or in a network, so that login and password are not necessary for access. Although this is an option to preserve digital legacy, this would be too demanding for users, because of the need for constant backup and software restrictions.
3. keeping a link with real-world institutions and documents to confirm one´s death, in the context of the social web. However, many changes would have to occur so that these processes were aligned, which would affect both governmental institutions and the laws of the jurisdiction where users live or the applications are hosted.
4. providing resources in the very social web applications to register users' volition. In this case, the user could choose via software the destination of his digital legacy. This would have a strong impact on the terms of use developed by software enterprises and agreed with by users.

Such solutions came out of a research with 83 software engineers [13]. That research also resulted in a list of requirements for social web applications that could provide users with volition functionalities for post mortem digital legacy (as in option d, from the aforementioned solutions).

In an extension of that research, [14] discussed how taboos and beliefs on death are an integral part of life in society and may affect, for instance, system development. To understand this phenomenon, software engineers' discourse was analyzed considering taboos and beliefs over death, which could condition the proposal of solutions to model digital legacy aspects for social networks. Seven categories of beliefs were generated and discussed, according to the software engineers' opinions: non profanable legacy, funeral rites, the immaterial beyond death, death as an end, death as adversity, death as interdiction, and the space required by death. In terms of results, that research found a predominance of the concepts of something immaterial beyond death and that death is an end in itself, two antagonist trends that might influence software development concerning death issues.

With regard to managing dead users´ digital legacy, [27] defined four possible destinations for such data in social networks: leaving (not doing anything to the account), deleting (removing the account), forwarding (giving the password or part of the assets to an heir) or archiving (keeping a backup copy of the data, maybe in material medium). Regarding the choice to leave passwords for an heir, [12] have not identified clauses about password transfer in case of death in the terms of use of Facebook, Google+ and LinkedIn, in three different countries (USA, UK and Canada). Finally, there are data that become unmanageable after the death of the account owner, as there are no means to access them.

The data of people who die, available on the web, permit posthumous interaction, that is, the system interaction with deceased people's data, or between users and deceased people via system [15]. These authors reinforce that it is worth stressing that the datum itself is not posthumous, as it was produced by the user when alive. Posthumous is the interaction occurring with data belonging to someone already dead, a process which is guaranteed by the asynchronous character permeating a large share of the communicative processes in the Internet. In that research, conducted with youngsters from the Internet Generation, it was discussed how the death representations of such generation influence their conceptions and practices on posthumous interaction.

There is evidence, however, that this posthumous interaction falls in frequency over time. [3] observed from empirical analyses on deceased users´ profiles on MySpace that comments on these pages reached a peak on the day of the death, decreased quickly within 10 days following this event and then reduced more slowly. The analysis of posthumous interaction is most often accomplished through empirical observations on websites than in interviews with the bereaved, given the emotional charge involving the death experience. The same applies to the study of real-world mourning practices, as it is more common to observe externally a funeral rite than to interview its participants [23].

Regarding the importance of semiotics in the design of systems that consider post mortem issues, it is necessary to discuss symbols or images related to this context. Proposing some useful guidelines for the development of systems that consider the possibility of digital mourning, [16] mentioned the importance of permitting the bereaved to control and choose their death-related symbols, determined in part by their belief systems, taboos and religious diversity. These authors also warn that some mourning symbols are extremely private, while others are meant to be publicly viewed.

However, are the users' wishes on the fate of their digital legacy influenced by religious beliefs? This issue is discussed by means of a case study in Section 3.

3 Methodology

This research was conducted with teenagers taking integrated secondary and technical education at a public school in Cuiabá (Mato Grosso – Brasil), called Instituto Federal de Mato Grosso. The school has 442 students in this segment, 78 of which participated in the research. Because of their age, these users are part of the Internet Generation, born between 1977 and 1997, according to Tapscott [22].

Initially, during the period from June 1st to 12th, in 2012, one of the researchers went to these students' classrooms, to provide them with a brief explanation about the research and give them a Term of Consent for answering the questionnaire, according to Brazilian Legal Resolution 196/96 [2], which establishes regulating directives and norms for researches involving human beings. As they were underage, the Terms of Consent had to be signed by their parents or person in charge, so that, at a set date, these young people responded to the questionnaire. However, this procedure was harmed because many students failed to bring the Term signed, saying they had either forgotten to bring it, or failed to show it to their parents. Hence, it was not possible to assess how many had not been granted permission to participate in the research and how many failed even to tell their parents about it. Maybe the very application of the questionnaire was affected by taboos concerning death or religious issues, hindering a more systematic reflection on the theme.

220 Terms of Consent were delivered and 78 students brought them signed. They participated in the research, filling in an individual anonymous questionnaire in the classroom. Considering all users who have expressed interest in participating in the survey, we have an intentional non-probability sample, composed of a known and finite population, which admits a maximum error of 4.7%, and desired confidence level of 95%.

The questionnaire had 33 questions, open and closed ones, divided into general data, knowledge on computing and on the internet, religion, social networks and posthumous data, and representations of death. To create the questionnaire, issues raised in previous researches and theoretical background in the area were taken into account, including a research instrumented with a 404-question questionnaire [1], the social representations of death categorized by [5], issues raised by software engineers as to the possibility of volition concerning the digital legacy [13], and these professionals' taboos and beliefs about death [14]. The planning of the data collection instrument was crucial to the research. It should be stressed that, before applying the questionnaires to the teenagers, a pilot research was conducted with the planned instrument. 19 students of a MSc program participated in it, who study education technologies and generational differences. The pilot questionnaire allowed significant adjustments to the instrument, especially in order to clarify some questions.

Initial data from this research were analyzed by [15]. In that first analysis, the authors investigated how the Internet Generation´s social representations of death influenced their conceptions and practices of posthumous interaction in social networks. 15 questions were analyzed then, and 5 correlations were made among the questions, via software. Now, in the present research, the questions about religion were analyzed and correlated with those on users´ volition about the destination of their post mortem digital legacy. To do so, the data were fed into an Excel spreadsheet and tabulated in SPSS (Statistical Package for the Social Sciences), for later quantitative-qualitative data analysis. The questionnaire questions were identified by the letter "P" for data analysis. From the total set of questions in the instrument, 18 were analyzed in this research, and 8 correlations were made among them, via software. That selection considered the focus of our proposal. In the qualitative analysis of the only open question, the teenagers were identified by the letter "J", followed by the questionnaire numbering. In the next section, these data are analyzed.

4 Data Analysis

In this section, the data tabulated are analyzed and discussed.

4.1 Demographic Data

Our research included 17.65% of the secondary education students of a federal school in Cuiabá (Mato Grosso – Brazil). All the teenagers belong to the "Internet Generation": 2.6% of them were born in 1993; 2.6% in 1994; 28.2 in 1995; 39.7% in 1996; 24.4% in 1997; and 2.6% in 1998.

Out of the teenagers participating in the research, 69.2% were female and 30.8% male. 93.6% answered they had a computer at home and 6.4% said they did not. About 73% of the teenagers declared to use the Internet for 1-4 hours a day. Among the most frequently accessed sites, 79% mentioned the SN Facebook, followed by Google (37.1%) and by Hotmail (25.6%), considering that each answer could contain up to three sites. Hotmail is observed to have the MSN communicator integrated to the system, which also constitutes a SN. 62 out of the 78 participants have Facebook accounts.

These sample data allow us to state that the great majority of the participant teenagers frequently use computers and the Internet frequently, making constant use of social networks. Thus, it can be considered that this research is really based on the opinion of the "Internet Generation".

4.2 Religion

The questionnaire had 5 specific questions that focused on religious issues and funeral rites. In question P8, 94.9% of the participants answered that they believed in God, whereas 3.8% do not share such belief and 1.3% did not answer the question. When asked about which religion they belong to, most students said that they are either Catholics or Protestants, among some other religions, as we can see in Table 1.

Table 1. Users´ religion

Religion	%
Catholics	33,3
Protestants	48,7
Spiritualists	5,1
Jeovah´s witnesses	1,3
No religion	10,3
No answer	1,3

Conversely to the Brazilian 2010 Census [11], this group of youngsters has more Protestants than Catholics, besides some Spiritualists and a single Jeovah´s Witness. Perhaps, this disparity is due to other demographic aspects, such as social classes, parental educational level, location of the school etc., which were not investigated in this research. Among the participants, there were no Jews, Buddhists or Muslims.

Question P10 asked if these youngsters engaged in religious rites. 38,5% answered "Yes, always"; 42,3%, "Yes, sometimes"; 7,7%, "No, never"; and 11,5% did not answer this question. Crossing the data from questions P10 and P9, the results suggest that most youngsters are not constant practitioners of their religions, as table 2 shows.

Table 2. Religion x Religious practices

Answer	Catholics (%)	Protestants (%)	Spiritualists (%)
Yes, always	34,6	48,6	50
Yes, sometimes	53,8	43,2	50
No, never	11,5	8,1	-

Such data suggest that those who are not constant practitioners of their religions might not be strongly influenced by their rites and dogmas concerning death and legacy. This is in close relation to the fact that religious systems are not homogenous, so that Catholic practices may differ from parish to parish, as well as Protestant attitudes vary from pastor to pastor.

In question P11, when they had to answer about post mortem life, 61,5% of the respondents stated that they believed on it, whereas 38,5% said they don´t. This question, different from the others, was not left blank by any participant, which suggests some kind of conviction on this topic. However, it is interesting to notice that post mortem life is a dogma that Catholics, Protestants and Spiritualists spouse, although their understanding on this issue is different. The fact that a significant amount of youngsters said that they do not believe in post mortem life confirms the aforementioned interpretation that their beliefs do not necessarily match the dogmas of their religions.

The 48 participants who answered that they believed in afterlife were asked another question (P12) on the meaning of life after death. They were given three answer options: "the migration of the soul to another physical body", chosen by 3 respondents; "life in parallel worlds", chosen by 29 respondents; and "other", chosen by 16 respondents.

Those who answered "other" were asked to explain, with their own words, the meaning of "*afterlife*". Analyzing the answers, it is very clear that the meaning of afterlife for these youngsters lies on where this life takes place, rather than what it consists of. Locative expressions, such as "*in Heaven*" or "*beside God*" are repeatedly present in their answers, as for instance: "*It means to go either to Heaven or Hell, according to how you lived on Earth*" (J12), and "*Living close to God after His return*" (J56). We also emphasize the predominance of typically Christian images in their explanations on afterlife, such as "Heaven", "Hell", "Jesus Christ" and the "Last Judgment", confirming the religious profile identified in P9. Here are some of the responses containing these images: "*Eternal life with God and His son Jesus Christ*" (J32), and "*When we die, we will wait for the final judgment*" (J71).

If we compare the results of P11 (belief in afterlife) and P9 (users´ religion), the results show that 76,9% of the Catholic respondents believe in post mortem life, 47,4% of the Protestants, 100% of the Spiritualists and 62,5% of those with no religion. The distinction between the answers of Catholics and Protestants confirms, once more, that these youngsters´ beliefs do not necessarily match the dogmas of their doctrines, as afterlife is a common element for Catholicism and Protestantism, even if slightly different in certain theological aspects.

4.3 Procedures after the User's Death

In question P30, the participants should choose up to three alternatives indicating what they wanted to be done to their SN accounts after death. It is important to recall here that, as explained in the methodology section, such alternatives consisted of procedures selected from previous researches on the subject [13] [14], conducted with software developers.

Table 3 shows the correlation between the answers to P30 and users´ religions (P9). As respondents could choose more than one option in P30, 204 answers were deemed valid. The acronym NR means the amount of choices for each option.

Table 3. Procedures after the user´s death vs users´ religion

Procedure after the user´s death (nr = 204)	**Catholics**	**Protestants**	**Spiritualists**	**No religion**	**%**
a1) A farewell message written by you should be shown on your profile, as soon as your death is confirmed (nr=42)	61,5	55,3	50	37,5	53,8
b1) A user previously selected by you should be given your login and password. However, he would have to obey the commands you left in a digital will (nr=37)	42,3	52,6	25	62,5	47,4
b2) An heir previously selected by you should be given your login and password, but free to do whatever he wanted to your account and data (nr=7)	15,4	2,6	0	12,5	9
c1) Your account should be immediately deleted, in case of official report of your death (nr = 17)	30,8	18,4	0	12,5	21,8
c2) Your account should be deleted by the end of a pre-selected period after the official report of your death. Within this period, your contacts would be normally able to access your account (nr=21)	15,4	34,2	50	12,5	26,9
d1) Your account should remain active on the SN, so that any user could see it (nr=14)	11,5	18,4	0	37,5	17,9
d2) Your account should remain active on the SN, but only your friends would be able to see it (nr=17)	26,9	18,4	25	12,5	21,8
d3) Your account should remain active on the SN, but only some previously selected friends would be able to see it (nr=5)	7,7	2,6	25	12,5	6,4
e1) Your data should be sent to a virtual cemetery or a memorial profile, containing only some information you chose, such as date of birth and date of death (nr=24)	26,9	34,2	25	25	30,8
e2) A backup copy of your data in the SN should be sent to an heir chosen by you (nr=20)	26,9	15,8	50	50	25,6

For a better analysis of these data, five categories were created, as described below.

a) Recording a posthumous message (item a1)

It is noticed that one of the most frequent options, chosen by 42 participants, expresses the desire to leave a posthumous message, displayed after the detection of the user's death. This option was the most frequently chosen, regardless of the user's religion, although it was more widespread among Catholics (61.5%). Then, communicating something can be seen as a last wish for most of these users, which confirms the use of social networks primarily for communication among people, even in the context of death. This seems to apply irrespective of religion, according to our sample.

Besides, confirming what [23] presented, the desire to leave a posthumous message shows that death on Social Networks is not a taboo issue that one should hide. This idea of death as an interdicted subject was also found by [14] as an infrequent concern in software developers´ solutions for digital legacy projects, which seems to match users´ answers in this research.

b) Naming an executor for your digital will (items b1 and b2)

Another frequent option, chosen by 37 participants, entails naming an executor for a digital will, by forwarding the user´s password and login to someone else (b1). This was the second most frequent option among representatives of all religions, except Spiritualists. Those who stated to have no religion were the most inclined to such procedure (62.5%). However, such solution requires a further discussion on legal issues related to digital assets.

On the other hand, only 7 participants chose the option b2, which implied further concerns on caution and privacy of the data. This option was seldom chosen, regardless of the user´s religion. Forwarding passwords is an alternative discussed by [4], but such action must not be disassociated from a will, as this could cause a series of problems analyzed by [13], just like misappropriation of identity, for example.

Making a will is also somehow related to the locative concern of afterlife respondents showed in question P12. Defining where (or whom) the data go to may provide some certainty in an area where doubts are predominant. Besides, the writing of a will is a very common practice in real world, so that the Internet Generation users expect this phenomenon to be modeled for computational purposes too.

c) Deleting the social network account (items c1 and c2)

38 participants chose options that involved account deletion from social networks, either immediately after death (c1) or after a previously set period (c2). The immediate exclusion was mostly chosen by Catholics (30.8%), whereas Protestants (34.2%) and Spiritualists (50%) opted for deletion by the end of a predefined time span. Such variation might be due to the differences between the funeral rites of these religions. The fact that Catholics keep celebrating requiem messes after death, besides other more intimate rites, might lead them to feel more comfortable with an immediate exclusion of the account, as they are more familiar to rites that do not involve the presence of the deceased person.

d) Keeping the account active (items d1, d2 and d3)

Conversely, 36 participants indicated that their accounts should be kept active on the SN. Those who claimed no religion were the ones who most frequently chose the option d1, in which the account would remain active and could be viewed by any user. This shows an indirect desire to remain interacting with the alive, even if posthumously, possibly because the lack of religion reduced incidental taboos that might prevent this kind of interaction. When presenting some design solutions that support similar users´ decisions, [16] affirm that "we can consider how systems might empower people who have died to maintain a digital identity that preserves their integrity and desires in this life; or, to deliver messages for loved ones into the future."

Among the three options in this category, d2 was the most frequently chosen, especially among Catholics (26.9% of these users) and Spiritualists (25%). Option d3, which demanded a more specific selection of users who would be allowed to see the dead user´s profile, was infrequent among the participants, regardless of religion, except for Spiritualists (25% of them). The unpopularity of this option might be due to the public nature of death in cyberspace, especially in social networks [23], so that users do not feel uncomfortable to share with their online friends the fact of death.

e) Preserving the digital legacy (items e1 and e2)

Another category of options indicated by the participants concerns data memory. 24 participants manifested the desire to have their data sent to a Virtual Cemetery or a profile memorial after death (e1), whereas 20 demanded that a backup copy of their data on the SN was sent to a previously selected heir (e2).

It was noticed that the preservation of the digital heritage is something desired by most respondents, regardless of religion: item e1 was the third most chosen option, selected by 30.8% of the participants. Item e2 was also chosen by 50% of the Spiritualists, and 50% of those who claimed no religion.

This data preservation is connected to the aforementioned locative concern about afterlife, as keeping the legacy entails the decision on where to keep it. Defining a place to store such data goes hand in hand with the religious concerns on where the soul is deemed to go after death, and with the legal concern about whom to give the legacy to. Therefore, most respondents preferred options in which the legacy was transferred to specific environments for this kind of data (as cemeteries and memorials), instead of keeping them like other data in the SN (as in items d1, d2 and d3).

The answers to question P30 were also correlated with those to P11, so as to investigate whether the belief in afterlife could influence users´ choices of procedures. However, this correlation did not present a significant difference between the choices of those who believed or not in post mortem life, especially when comparing Protestants and Catholics. Possibly, such similarities are due to the resembling interests and basilar doctrines of these Christian religions.

In question P22, users had to answer if they would like to set their SN accounts and determine their preferences about what should be done with their profile after death, as if making a will for digital assets. 67.9% of the participants answered "yes"; 29.5%, "no"; and 2.6% did not answer. Analyzing such data, it is clear that these young people intend to customize the destiny of their legacy, which confirms the customization characteristic [22] identifies with the Internet Generation. This is another option for social networks, which might prove useful in cases when the profile must be managed by an heir (as discussed in the items d1, d2 and d3, from P30).

Those who answered "yes" to question P22 were asked in question P23 to select from a list which resources they would like to block in their SN accounts after death. Notice that they could select as many options as they wanted. The data from P23 were correlated with those from P11, in order to analyze if the belief in afterlife could influence users´ choices on which resources to block. In general, we could see that those who believed in post mortem life opted to block more resources. Possibly, the idea of remaining somehow alive after death conflicted with the idea of not being able to manage their own account, so that these users opted for blocking most resources.

On the other hand, the two options of blocking most frequently chosen (blocking chats and blocking external applications) were slightly more popular among those who do not believe in afterlife. It is interesting that both functionalities would require some sort of response from the deceased user (even if in afterlife), so that those who do not believe in post mortem life tend not to consider relevant to keep these functionalities working. It is also noteworthy that the difference between the proportions of afterlife believers and non-believers was greater for the following options: blocking old messages visualization, blocking old notes liking, blocking content sharing and blocking access to the profile information. Table 4 presents these data in detail.

Table 4. Belief in afterlife vs. functionalities blocking

	Blocking resources in case of death	Belief in afterlife? (%)	
		Yes	No
1.	blocking chats	66,7	71,4
2.	blocking external applications	39,4	42,9
3.	blocking old messages visualization	39,4	19
4.	blocking history updating	33,3	23,8
5.	blocking note posting	30,3	23,8
6.	blocking comments about photographs	27,3	23,8
7.	blocking comments about old notes	27,3	23,8
8.	blocking old notes liking	27,3	9,5
9.	blocking content sharing	24,2	9,5
10.	blocking access to the profile information	21,2	0
11.	blocking access to photographs	12,1	9,5
12.	blocking no resource	12,1	4,8

However, new data arise when we correlate the choices on functionalities blocking (P23) with users´ religion (P9). Below, one can see Table 5, whose columns are numbered according to the caption of Table 4:

Table 5. Religion *vs*. functionalities blocking (%)

Religion	1	2	3	4	5	6	7	8	9	10	11	12
Catholics	76,5	41,2	23,5	35,3	29,4	35,3	29,4	23,5	35,3	11,8	11,8	5,9
Protestants	73,1	46,2	42,3	30,8	30,8	26,9	30,8	23,1	15,4	15,4	11,5	3,8
Spiritualists	50	50	--	--	50	25	25	25	--	25	25	25
No religion	40	20	40	40	--	--	--	--	--	--	--	20
Global average	68,5	40,7	31,5	29,6	27,8	25,9	25,9	20,4	18,5	13	11,1	9,3

By analyzing these data, we can see some common interests of Catholics and Protestants regarding which functionalities to block, possibly due to the similarities between these religions. There are only significant differences considering the options "blocking old messages visualization", mostly preferred by Protestants, and "blocking content sharing", mostly preferred by Catholics. Spiritualists also chose fewer resources to be blocked. Those who have no religion have chosen to block only 4 items, which suggests that the lack of religiosity might lead to fewer taboos regarding death, and therefore less fear from legacy profanation.

4.4 Symbols

When the choice on symbols to depict death (P32) is correlated with the users' religion (P9), there are some subtle differences between the signs chosen by each religious group, as shown in Table 6.

Table 6. Symbols vs religions

Symbol	**Catholics (%)**	**Protestants (%)**	**Spiritualists (%)**	**No religion (%)**
Cross	23,1	18,4	25	25
Tomb	19,2	10,5		
Gravestone	11,5	10,5		
Skull	3,8	7,9		25
Light	7,7	10,5	25	25
Coffin	15,4	23,7		25
Sky	15,4	10,5		
Others	3,9	8	50	

The analysis of such data shows that the cross is a common symbol for these users' religions, representing Christ's death and therefore common people's death as well. On the other hand, there is an important difference regarding the second most frequent choice on these symbols: whereas Catholics tended to prefer the tomb, Protestants felt more inclined to choose the coffin. This might be a result from the fact that Protestants tend not to visit cemeteries as frequently as Catholics [7].

Spiritualists opted for less morbid symbols, such as the cross and the light, which might reflect their view on death and life as cyclical phenomena, instead of sad mishaps. The sky was not among their choices, maybe because it is too closely associated with Heaven, in opposition to Hell, and this dichotomy is traditionally not part of Spiritualist beliefs. Also regarding morbidity, the skull was the most infrequent choice in this question, suggesting that users do not want to face physical decay or destruction as a symbol on their social networks.

Although the sample consisted of too few Spiritualists, we cannot neglect that half of them opted for the alternative “others”, a hint that the symbols proposed in the questionnaire were maybe too intertwined with Christianity and did not represent other religions. One of these users proposed the symbol of God moving toward the deceased, and the other chose the symbol of Ouroboros, a pagan myth of a serpent that devours its own tail.

Correlating question P32 with P11 (on users’ belief in afterlife), the results showed that those who believed in post mortem life tended to choose the light (18,8% of the believers) and the cross (16,7%). On the other hand, those who do not believe in afterlife tended to choose the coffin (26,7% of the non-believers) and the cross (20%). As it was said before, the cross is a dominant symbol in the answers due to its importance in Christianity, no matter if users believe or not in post mortem life. The choice on the coffin by those who do not believe in afterlife and consider death as an end in itself can be related to the material sense this symbol brings to death, instead of a more transcendental one possibly related to afterlife and the symbolism of light.

5 Conclusions

Religion, along with gender and ethnicity, is part of the diverse ways of understanding the world and individuals. Thus it is paramount to do research on elements from these fields that may contribute to the design of more universal applications which respect cultural diversity and meet the needs of digital life, until its very end.

However, many people are not comfortable with the idea of death, due either to fear of the unknown or to bonds of affection and friendship established by life in society and strongly influenced by culture. From the computational viewpoint, the treatment that is given to modeling users´ post mortem aspects requires further studies.

Regarding religion, there is an inextricable diversity that must be considered in cultural research. In the research presented herein, only the most popular religions in Brazil [11] are represented. However, there are different religious profiles in different countries, which have a strong influence on the design of computational systems, and sometimes on political and legal systems that may influence software use. Thus, there is a great challenge to the analysis on how these different expressions of culture can influence applications design.

As for the procedures to be done with the users´ accounts after their death, we realize that most respondents in this research chose options that imply moving the data to another location, such as sending data to a digital cemetery, or handing them to an heir. This choice is closely related to religiosity, since these users also defined afterlife by means of a locative change. By this choice, they expressed a desire to change either the way the data are displayed in the SN or the person in charge for managing them, in the case of a digital will.

Regardless of the options chosen by the users in this study, one notices that there are different desires concerning the fate of the digital legacy, which indicates that systems should not only provide functionalities to handle these assets, but also give users flexibility to configure their accounts according to their volition. Of course, these features must be designed according to the laws of each country, but one must not forget that laws are created to meet the demands of society. At this point, this research shows some possibilities to discuss and develop new norms and laws concerning digital legacy.

The discussion on symbols representing death may suggest some guidelines for icons that can be used along with functionalities in applications that deal with post mortem digital legacy, in order to permit the expression of different understandings of death in graphical interfaces.

One of the limitations of this research is its representativity for the whole Internet Generation. According to Tapscott [22], this generation comprises those who were born between 1977 and 1997. As this research addresses interviewees who were born between 1993 and 1998, this study represents only part of the Internet Generation time span.

For future works, it is necessary to have a larger and more representative sample of different religions, also considering users from other generations. Nevertheless, the method herein employed and the objectives of this research make clear that this study is significant for the design of more universal solutions. Considering religion as a lens to see the world and the interactive processes among people is a possibility that should not be ignored by designers and the HCI community.

References

1. Boukharaeva, L.M.: Os valores educacionais dos estudantes nos limites da educação universitária – Questionário, Ijuí, Rio Grande do Sul, Brazil (1994) (in Portuguese)
2. Brasil: Conselho Nacional de Saúde. Resolução 196/96. Diretrizes e Normas Regulamentadoras de Pesquisas Envolvendo Seres Humanos (1996) (in Portuguese), http://conselho.saude.gov.br/resolucoes/reso_96.htm (accessed January 15, 2013)
3. Brubaker, J.R., Hayes, G.R.: "We will never forget you [online]": an empirical investigation of post-mortem Myspace comments. In: Hinds, P., Tang, J.C., Wang, J., Bardram, J., Ducheneuat, N. (eds.) Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW 2011), pp. 123–132. ACM, New York (2011), doi:10.1145/1958824
4. Carroll, E., Romano, J.: Your Digital Afterlife: When Facebook, Fickr and Twitter Are Your Estate, What's Your Legacy? New Riders Pub., Berkeley (2010)
5. Coelho, J.F.C., Falção, E.B.M.: Ensino científico e representações sociais de morte humana. In: Revista Iberoamericana de Educación, 3rd edn., vol. 39 (2006) (in Portuguese)

6. Comitê Gestor da Internet. Brasil. Survey on the use of Information and Communication Technologies - ICT Households and Enterprises 2011.NIC.br/CETIC.br (2011), http://op.ceptro.br/cgi-bin/cetic/tic-domicilios-e-empresas-2011.pdf (accessed January 18, 2013)
7. Costa, H.: Pesquisando a escatologia da morte no imaginário católico e protestante. In: Pereira, A.P., et al. (eds.) Anais do XV Encontro Regional de História – ANPUH-Rio (2012), http://www.encontro2012.rj.anpuh.org/resources/anais/15/1338400274_ARQUIVO_artigoparaseminarioanpuh.pdf (accessed January 6, 2012)
8. Durkheim, É.: As formas elementares da vida religiosa, Martins Fontes, São Paulo, Brazil (2000) (in Portuguese)
9. Geertz, C.: A Religião como Sistema Cultural. In: A Interpretação das Culturas. LTC, Rio de Janeiro, Brazil (2008) (in Portuguese)
10. Hallam, E., Hockey, J.: Death, Memory and Material Culture. Berg Publishers, Oxford (2001)
11. IBGE: Censo Demográfico (2010) (in Portuguese), ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/tab1_4.pdf (accessed January 6, 2012)
12. Locasto, M.E., Massimi, M., DePasquale, P.J.: Security and privacy considerations in digital death. In: Peisert, S., Ford, R., Gates, C., Herley, C. (eds.) Proceedings of the 2011 Workshop on New Security Paradigms Workshop, pp. 1–10. ACM, New York (2011), doi:10.1145/2073276.2073278
13. Maciel, C.: Issues of the social web interaction project faced with afterlife digital legacy. In: Gomes, A.S., Souza, C.S., de, S.J.A. (eds.) Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction (IHC+CLIHC 2011), pp. 3–12. Brazilian Computer Society, Porto Alegre (2011)
14. Maciel, C., Pereira, V.C.: The influence of beliefs and death taboos in modeling the fate of digital legacy under the software developers' view. In: Workshop Memento Mori: Technology Design for the End of Life, CHI 2012, Austin, TX (May 2012), https://sites.google.com/site/chi2012eol/accepted-papers (accessed November 18, 2012)
15. Maciel, C., Pereira, V.C.: The internet generation and its representations of death: considerations for posthumous interaction projects. In: Maciel, C., Souza, P.C., de, C.J.A., Neris, V. (eds.) Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems (IHC 2012), pp. 85–94. Brazilian Computer Society, Porto Alegre (2012)
16. Massimi, M., Odom, W., Banks, R., Kirk, D.: Matters of life and death: locating the end of life in lifespan-oriented HCI research. In: Tan, D., Fitzpatrick, G., Gutwin, C., et al. (eds.) Proceedings of CHI 2011, pp. 987–996. ACM Press, New York (2011), doi:10.1145/1978942.1979090
17. Massimi, M.: Technology and the human lifespan: learning from the bereaved. Interactions 18, 26–29 (2011)
18. Pew Research Center's Internet & American Life Project. Photos and Videos as Social Currency (2012), http://pewinternet.org/Reports/2012/Online-Pictures.aspx (accessed January 18, 2013)
19. Rodrigues, J.C.: O Tabu da morte. Fiocruz, Rio de Janeiro (2011) (in Portuguese)

20. Salgado, L.C., de, C., Leitão, C.F., de Souza, C.S.: A Journey Through Cultures. Springer, New York (2013)
21. Santos, C.A.F.: Os profissionais de Saúde Enfrentam – Negam a Morte. In: Martins, J., de, S. (eds.) A Morte e os Mortos na Sociedade Brasileira. Hucitec, São Paulo (1983) (in Portuguese)
22. Tapscott, D.: A hora da geração digital: como os jovens que cresceram usando a internet estão mudando tudo, das empresas aos governos. Agir Negócios, Brazil (2010) (in Portuguese)
23. Walter, T.: On Bereavement: The culture of grief. Open University Press, Philadelphia (1999)
24. Wyche, S.P., Grinter, R.E.: Extraordinary computing: religion as a lens for reconsidering the home. In: Junior, D.R.O., Arthur, R.B., Hinckley, K., et al. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2009), pp. 749–758. ACM, New York (2009), doi:10.1145/1518701.1518817
25. Wyche, S.P., Grinter, R.E.: Using sketching to support design research in new ways: a case study investigating design and charismatic Pentecostalism in São Paulo, Brazil. In: Mai, J.E. (ed.) Proceedings of the 2012 iConference (iConference 2012), pp. 63–71. ACM, New York (2012), doi:10.1145/2132176.2132185
26. Wyche, S.P., Hayes, G.R., Harvel, L.D., Grinter, R.E.: Technology in spiritual formation: an exploratory study of computer mediated religious communications. In: Hinds, P., Martin, D. (eds.) Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW 2006), pp. 199–208. ACM, New York (2006), doi:10.1145/1180875.1180908
27. Zhang, M., Jennett, C., Malheiros, M., Sasse, M.A.: Data After Death: User Requirements and Design Challenges for SNSs and Email Providers. In: Workshop Memento Mori: Technology Design for the End of Life, CHI 2012 (2012), https://sites.google.com/site/chi2012eol/accepted-papers (accessed January 6, 2012)

A Tale of Two Puppets, Two Avatars and Two Countries

Yoram I. Chisik, Monchu Chen, and M. Clara Correia Martins

University of Madeira, Caminho da Penteada, 9020-105 Funchal, Portugal
{ychisik,monchu}@uma.pt, mccmzz@gmail.com

Abstract. We describe observations from studies conducted with school children in Nepal and Portugal with the aim of exploring two questions: 1. Can we develop an interactive game that will continue to engage the child in gameplay when the electricity supply fails? 2. What are the discernable differences in the ways children in Nepal and Portugal interact with the game given the huge socio-economic and cultural differences between them? The studies highlight the potential of the design in fostering continued engagement in the game when the electricity supply fails. They also show surprising little difference between the children in Nepal and Portugal. The observations suggest game mechanism and basic interactions can be regarded as universal across cultures and levels of income and can thus be designed and tested in any location and be expected to work and be understood by children from any culture and background.

Keywords: Games, Play, Children, Electricity, Design, Interaction, HCI4D.

1 Introduction

Designing for children is notoriously hard as adults are indelibly separated from children by a generation gap and thus cannot possibly imagine the world from their perspective [6,7]. Designing for children in developing countries is doubly hard, as designers have to bridge a set of environmental, cultural and socio-economic perceptual gaps in addition to the aforementioned generation gap [8,4]. To confound matters further the physical distance and other logistical difficulties make it hard to employ the participatory and emphatic design practices that enable designers to bridge generational and environmental gaps in their perception [5,9].

The work presented in this paper was born out of an invitation to organize a children entertainment workshop in Nepal, a country with severe infrastructure limitations where even in the capital city electricity is only available for a few hours a day.

Our goal in this study was two fold. First we wanted to see if we could transcend Nepal resource limitation by developing a game that would be equally engaging with or without electrical power and will thus always have the power to engage and entertain its players. Second, we wanted to see whether the socioeconomic and technology use differences between Nepal and Portugal (Our home country) has any impact on the ways in which children in these countries engaged with the computer and played the game as suggested by recent studies [14,13].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 658–665, 2013.

2 The Game

For the purpose of the study we developed a simple game consisting of two screen avatars, a singer and a conductor (Fig. 1) that are controlled by two sock puppets operated by two players (Fig. 2), and governed by two self-discoverable rules.

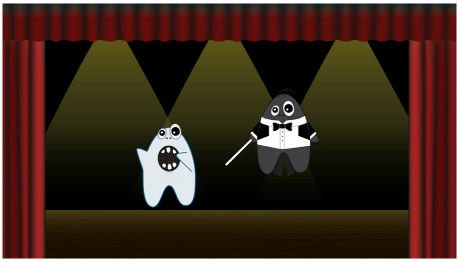


Fig. 1. The avatars

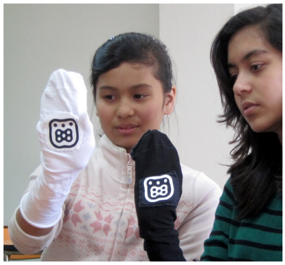


Fig. 2. The sock puppets

We based our design on the natural curiosity and inventiveness of children [12,3] and on the malleability of flexible tangible objects such as sock puppets [10,2]. We strove to make the visual design of the avatars and the activities they perform as universal as possible so any child will be able to relate to them.

The game has a simple set of mechanics. When the singer sock puppet opens his mouth the singer avatar starts singing, and when the conductor puppet opens his mouth the conductor avatar waves his baton. In addition if the conductor avatar waves his baton at the same time the singer avatar is singing the singer will change his tune.

The game is governed by a set of two rules:

The first rule is that the puppet (operated by a player) controls the avatar appearing on the screen. To enact this rule the player has to interact with the computer.

The second rule is that the player/puppet/avatar control each other. To enact this rule the players have to interact with each other.

We hypothesized that if the children can discover/learn these rules in the digital game they would be more likely to enact them when the electricity supply fails and the computer ceases to function and that the sock puppets coupled with the learned rules and a sense of play would provide an engaging (and hopefully an equally engaging) experience.

The game was developed using Quartz Composer [1] and reacTIVision [11]. The motion of the sock puppets and the evaluation of the open/close mouth events were detected using a webcam connected to a computer and fiducial markers affixed to the sock puppets (Fig. 2).

3 Putting the Game to the Test

To put the game to the test we conducted three field studies in schools. Our aim in these studies was not to conduct a formal study of the game but to gauge the overall concept. We primarily wanted to see whether the children found the overall experience fun and engaging, how they related to the avatars, whether they would be able to discover the rules of the game and what they would do when the computer ceased to function due to a simulated power failure.

Initially we had planned to conduct two field studies, one in Nepal and one in Portugal so that we can not only investigate the ways in which the game is being played but also see if there are any differences in the ways Nepalese and Portuguese children engaged with the game given the huge disparity in terms of culture, geography and economic wellbeing between them. However in light of the great socio-economic disparity between different schools in Nepal we decided to conduct two field studies in Nepal, one in a private school catering to the children of the well to do and the other in a public school catering to village children so that we can gain a more representative image of the child population of Nepal.

3.1 The Schools

The first trial was conducted with 14 students (10 Females and 4 Males) between the ages of 13-14 at the Ullens School in Kathmandu, Nepal (Fig. 3, left), a private school offering well-equipped classrooms to students coming from well endowed families capable of paying the high fees charged by the school.

The second trial was conducted with 12 students (8 Females and 4 Males) between the ages of 10-13 at the Shree Rudrayanee School (Fig. 3, middle) located on the outskirts of Kathmandu, a government run school offering free public education in bare bones facilities to the general population.

The third trial was conducted with 24 children (12 boys and 12 girls) between the ages of 10 to 11 in the Escola Basica dos 2º e 3º Ciclos dos Louros (Fig. 3, right) located in Funchal, Madeira, Portugal, a government run school offering free public education in well equipped classrooms to the general population.

In all schools the school selected the participating students and no remuneration was provided to either the school or the students.

Fig. 3. The three schools in which the studies were conducted

3.2 Setup and Method

We aimed to keep the setup of the environment and the procedure followed by the facilitators the same in each of the schools as much as possible and for the most part we were successful.

In each of the schools the field study was conducted using a laptop computer set up in one of the classrooms. In Kathmandu the study was conducted in English by two facilitators one interacting with the children and the other serving as an observer, note taker and cameraman. The Funchal study was conducted in Portuguese by three facilitators two interacting with the children and the third serving as an observer, note taker and cameraman.

Gameplay was conducted with a pair of children at a time. The pairs were always composed of two girls or two boys (in Kathmandu this was a result of self-organizing by the children themselves, in Funchal the pairs were matched by one of the facilitators). Each pair spent about 10-15 minutes with the game.

The children were not provided with any information about the game or the rule set. They were only instructed to choose one of the two puppets and to point their puppet towards the camera embedded at the top of the laptop screen.

Each field test session consisted of two conditions, a power condition in which the computer was functioning and the participants could interact with it and a no power condition in which the computer was not functioning and the players had to solely rely on the sock puppets and their own intuition in order to play the game. The power condition always preceded the no power condition so that the children would have a chance to play the game and discover the rules. The children were not told about the no power condition, the screen simply went black.

The facilitators not only observed the reactions of the children to the game they also engaged in a conversation with the children to see whether they recognized the avatars and made any correlations between their actions and the actions of the avatars. All the gameplay sessions were video recorded so that they could be further analyzed at a later date. In the Funchal study the children were debriefed after the gameplay session by one of the facilitators to gauge their reactions to the game and their understanding of the elements and rules of the game.

4 Observations

Perhaps one of the most interesting observations of all is the striking lack of difference in the reactions of the children to the game and its constituent elements despite the immense differences in terms of culture, language and environment between the Kathmandu and Funchal and the wide socio-economic disparity between the two schools in Kathmandu. Without exception all of the children found the game engaging and fun, they immediately took to moving the characters around the screen, recognizing who controls which avatar and making the singer sing. We therefore report our observations of the children reactions to the game in a unified form marking the differences between the two cities and three schools where necessary.

4.1 Discovering the Rules of the Game

Discovering the First Rule. All the children were able to map their sock puppet to the corresponding avatar (white sock puppet to the singer, black sock puppet to the conductor) and recognized the relationship between the actions they performed with the sock puppet and the activity of the avatar on the screen. Thus all of the children easily discovered the first rule (the puppet controls the avatar) although there were some minor differences in the speed in which the children came to these realizations all did so within the first two minutes of playing the game.

Discovering the Second Rule. In both schools in Kathmandu only a small number of the children discovered the second rule (the player/puppet/avatar control each other). In the game the singer avatar sang nonsense words. When the singer sang without the direction of the conductor he repeated the word "oolalah", when the conductor started twirling his baton he changed the word to "oogagah". We assume the difference in tone between the two words was not very distinct so some of the children did not notice the difference in singing and thus only saw the singer and the conductor as independent actors sharing the same stage. For the test in Funchal we changed the sounds so when the singer sang without the direction of the conductor he warmed his voice by singing la..la..la..la…and when the conductor started twirling his baton the singer started humming the tune of Beethoven's *Ode to Joy* in order to make the tonal shift between the two scenarios very distinct. This change had the desired effect with all of the children in Funchal discovering the second rule.

4.2 Interacting with the Technology

The reacTIVision Fiducial Markers used to control the avatars were both a prominent and highly visible feature of the sock puppets and a source technical difficulties as the reacTIVision software has difficulties with fast moving objects. Although the children were not given any information about the game mechanics or the technology behind the game none of them had any difficulties in figuring out the mechanics and none were fazed by technical glitches instead they incorporated a form of experimentation into their gameplay for example:

• Exploring the line of sight of the camera and seeing at what point the avatar stopped responding to their hand (puppet) movement.
• Playing with the open and close mouth actions. Some of the children constantly held the mouth open so that the open mouth fiducial marker was visible as this achieved the desired result. Others experimented with hiding the fiducial marker with their other hand to test whether the visibility of the marker was the controlling element
• When the system did not respond as expected such as when the movement of a fast moving hand was not effectively reflected in the movement of the avatar the children most frequently responded by repeating the same action over and over to validate whether the avatar would respond as expected. In face of a non-compliant avatar response, the children would attempt at either moving the hand faster or/and moving the

puppet closer to the camera – at times within a few millimetres of it – to try to get the desired response.

4.3 Interacting with the Avatars

In some cases the relationship between the child and the avatar was more personal in both Kathmandu and Funchal we observed children talking to the avatars by saying "hello" to the conductor at the beginning of the game or telling the singer to "shut up" when he was singing his warm up song. Some of the children also tried to guide the avatars by placing their sock puppet next to the avatar on the screen and trying to nudge him in the direction they wanted him to go.

As was hoped the children also brought their own meanings and interpretation into the game. In both the Funchal and Kathmandu studies some of the children referred to the conductor as the as the magician and to the singer as the ghost. One pair in Funchal thought the objective of the game was for the characters to run after each other another pair suggested that the conductor has a baton so he could hit the ghost with it.

4.4 The No Power Condition

The initial reaction of most children to the no power scenario was to try and get help pointing out that the computer is no longer working, when no help was forthcoming they adopted a wide range of approaches:

- Some of the children adapted the same approach they used in case of a slow response, i.e. continue to play and position the sock puppets in different ways to try and get a response from the system despite the fact that in the no power condition the screen was entirely black.
- A few tried to fix the system by pressing some of the keys or looking around the computer for the reset button. In these cases it was usually one of the children who was the adventurous one trying to fix the system while the other child was the cautious one, either telling his partner not to touch the system or physically preventing him from touching anything.
- The inability to get the system back to work did not, in most cases, deter the children from continuing to play-act with the socks for a while. In some cases the children followed the narrative of the game and even sang the singing avatar's tune, in others they followed their own narratives some of which were created from elements of the game. For example, In both Funchal and Kathmandu a pair of boys actually play-acted a scuffle with the sock puppets explaining that that is how they see the game as they assumed the objective of the game is for the conductor to catch the singing ghost and that the conductor baton was a weapon to be used on the ghost.

Most of the children play acted for a short period of time as they were self conscious someone was watching them and some simply waited patiently for the facilitators to help them when all other attempts at either getting attention from the system or the facilitators failed.

5 Concluding Remarks

In this study we set out to examine two questions: 1. Can we create an interactive game that would prove to be engaging with and without electrical power? 2. Can we discern any differences in the ways children in Nepal and Portugal interacted with the game given the huge socio-economic and cultural differences between them?

The first question remains to be fully answered as the no-power condition proved to be too contrived to provide a definitive answer to the question. However there were clear signs of engagement highlighting the potential of the approach. We intend to incorporate the lessons learned in this study into the next version of the game.

The second question yielded a fascinating set of observations. The first and perhaps the most significant one is the striking lack of difference in the reactions of the children to the game despite the immense differences in terms of culture, language, socio-economics, and exposure to technology between Nepal and Portugal and the socio-economic and technology exposure disparities between the schools in Nepal.

This finding is of primary importance as it implies that the design of mechanism and low-level basic interactions of technologies can be regarded as universal across cultures and levels of income and can thus be designed and tested in any location and be expected to work and be understood by kids of any culture and background. Furthermore, having done and tested their initial designs at home field study teams can move beyond basic interactions and mechanics and focus on aspects of environment, culture and language where participatory design truly comes into its own thereby maximizing the benefits of the field study and minimizing the overall cost of development.

The second significant observation is that the children often did not regard the failure of the technology as a failure but as an opportunity. Although the fast hand movements of the children led at times to a slow response by the game as the software struggled with the fast motion, the children were not frustrated by these failures but actively sought out ways to get around these difficulties by experimenting with different hand motions, speeds, positions and collaborations to get the system to perform as they wished. Thus the children saw what we perceived as technical failures as a sort of challenge to be solved. This perception can be effectively integrated into the game design process as a means of dealing with the foreseen and unforeseen challenges of developing technologies for developing and resource challenged countries.

Our future work will follow up on the insights gleaned in this study in a number of tracks: 1. We intend to continue with the development of the game towards the ultimate goal of creating an equally engaging game. 2. We intend to explore how children perceive and engage with the hidden “magic” inside the machine and what are the different methodologies they adopt in their approaches to problem solving. 3. We will continue to replicate the follow up studies in a number of countries so as to further explore whether ideas and observations gleaned using participatory and emphatic design methodologies can be applied across the board and if not where do the boundaries lie.

Acknowledgements. We would like to express our outmost thanks and extreme gratitude to Ishwor Hamal, Veronika Halupka, Monique Park, Joana Gomes, Sophia Gomes and Rui Branco for the help and assistance they provided us during the field studies. We thank Diana Xu and Janet Read for comments they made on an early draft of this paper. Last but by no means least we would like to thank all the children and teachers who participated in the study for playing our game and sharing their insights with us.

References

1. Apple Quartz Composer (2013), `https://developer.apple.com/technologies/mac/graphics-and-animation.html`
2. Bakker, S., Antle, A.N., van den Hoven, E.: Identifying embodied metaphors in children's sound-action mappings. In: Proceedings of the 8th International Conference on Interaction Design and Children (IDC 2009), pp. 140–149. ACM, New York (2009)
3. Caillois, R.: Man, Play and Games. University of Illinois Press, Champaign (1961)
4. Chetty, M., Grinter, R.E.: HCI4D: HCI challenges in the global south. In: CHI 2007 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2007), pp. 2327–2332. ACM, New York (2007)
5. Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., Thies, W.: "Yours is better!": participant response bias in HCI. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems (CHI 2012), pp. 1321–1330. ACM, New York (2012)
6. Gilutz, S., Black, J.B.: Child and Design Factors interacting in Children's HCI Helping children focus on the content, not the interface. In: Designing for Children, Mumbai, India, February 2-6 (2010)
7. Hourcade, J.P.: Interaction Design and Children. Foundations and Trends in Human-Computer Interaction 1(4), 277–392 (2008)
8. Hussain, S., Sanders, E., Steinert, M.: Participatory Design with Marginalized People in Developing Countries: Challenges and Opportunities Experienced in a Field Study in Cambodia. International Journal of Design 6(2) (2012)
9. Kolko, B., Putnam, C., Rose, E., Johnson, E.: Reflection on research methodologies for ubicomp in developing contexts. Personal Ubiquitous Computing 15(6), 575–583 (2011)
10. Marco, J., Cerezo, E., Baldassarri, S., Mazzone, E., Read, J.C.: Bringing tabletop technologies to kindergarten children. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology (BCS-HCI 2009), pp. 103–111. British Computer Society, Swinton (2009)
11. Reactivision, `http://reactivision.sourceforge.net/`
12. Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (2001)
13. Toyama, K.: Technology as amplifier in international development. In: Proceedings of the 2011 iConference (iConference 2011), pp. 75–82. ACM, New York (2011)
14. Yardi, S., Bruckman, A.: Income, Race, and Class: Exploring Socioeconomic Differences in Family Technology Use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2012), pp. 3041–3050. ACM, New York (2012)

Avalanche! Reanimating Multiple Roles in Child Computer Interaction Design

Timo Göttel

University of Hamburg, Department of Informatics, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
tgoettel@acm.org

Abstract. This paper values children as design partners in Child Computer Interaction to ensure building environments highly suitable for the target group. However, it appears beneficial to address underlying roles, align design processes on school activities or give considerations to the voluntary nature of such projects. We conducted eight projects with 75 pupils using initial learning environments. We found that starting design processes with pupils as users helps to identify further design steps that can be consequently tackled with the former users as informants and design partners. Thus, we suggest the Avalanche Design Cycle to engage children as fresh users at several stages asking them to become informants and design partners consecutively. The Avalanche Design Cycle is an iterative design process complying with school life allowing pupils to join/leave design teams. Thereby, it is aiming at sustained team sizes and more representative groups.

Keywords: Child-centered design, users, informants, design partners.

1 Introduction

Child Computer Interaction (CCI) is a subcategory of Human Computer Interaction with constant growth over the last decade[1]. Development of computer environments for children needs specific design methodologies because adult designers often hold misconceptions of child users based on fading individual childhood memories [2]. Additionally, children have specific requirements that for example cannot be associated with the goal to complete a task as it is done when designing for adult users [2]. Because of these profound differences current CCI research is focused on engaging children in design processes and accordant success stories. In this paper we highlight the need to open design processes for external user tests with children. Thus, we examine current CCI design methodologies based on a literature review and project observations to illustrate shortcomings in using these methodologies. Consequently,

[1] In the past eleven years, the CCI community has established the annual Conference on Interaction Design and Children (IDC), an IFIP TC13 group, and is "featured community" at the CHI Conference on Human Factors in Computing Systems for the 3rd time in 2013.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 666–673, 2013.

we present a prototyped design process allowing design teams to provide entry points, explicitly address roles, and consider school life.

2 Related Work

Harel [5] first presented the idea of children designers. She enabled children to design learning environments on given educational objectives for younger children using a computer. Consequentially, Scaife et al. [13] pointed out that designing for or with children is an essential question. In this regard, Scaife et al. propose a hybrid form of User-Centered Design (UCD) and Participatory Design (PD) to respond to child characteristics. Yet, they identify issues in seeing children as adequate design partners because children cannot define their own learning goals. Therefore, Scaife et al. understand children as informants testing prototypes, verbalizing needs or ideas, and designing through scenarios or games.

Later Druin [3] reviewed PD with children and presents four roles to involve children in design (Figure 1): Users are observed while utilizing technology to get an understanding on how to improve it. Testers are requested to comment on prototypes while using them. Informants are involved at the design process at certain stages. Informants contextually switch roles between user and tester. Design partners are full members of a design team with adults. Druin does not see a universal role for every context but she states that these roles are building upon each other. In this connection, many reports on successful implementations of children as design partners appeared in recent years (e.g. [8, 12]).

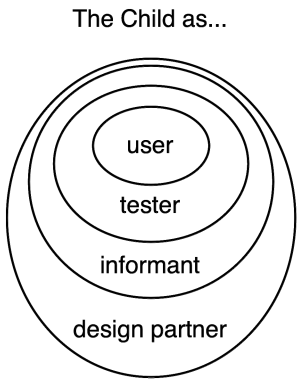


Fig. 1. Design roles according to Druin [3]

Accordingly, a recent review on CCI research acknowledges a highly regarded value in relying on design partners amongst the research community. Yet, the paper also reveals that in contradiction to these statements most papers presenting a system relied on testers to inform the design (59%) whereas only 31% were build upon projects with design partners [14]. This is surprising because design partners can theoretically be addressed as informants. The mismatch may be explained by a lack of reports in CCI literature on how to activate multiple roles with design partners. This may be in fact a large barrier considering that even experienced adult designers need strong and comprehensive tools as for example Personas to impersonate conventional users for evaluation sessions [1]. Accordingly, CCI researchers may need a model that allows to activate multiple roles when designing with design partners because they do not want to miss the chance to benefit from the other roles (e.g. informants).

When looking at CCI literature, iterative evaluations with external users are rarely presented. This is astonishing because UCD for example strongly relies on external testers throughout the entire process. An article by Kelly et al. [7] is a good example for CCI papers that present user tests with pupils without prior knowledge of the system or the design process: The evaluations mark the end of their project. Kelly et al. state that "the evaluation produced useful results for improvements, which were completely met at the second evaluation stage". This suggests that the tested environments can be seen as high-fidelity prototypes. Yet, literature shows that evaluating low-fi prototypes with external users during the whole design processes can provide more insights than revealing minor usability issues at the end of a project (cp. [10]). Hence, current CCI design processes do not fully consider user tests with fresh users as recurring milestones. Thereby they miss the chance to share prototypes with external users to establish feedback loops that may produce more universal results.

In a different context, Resnick [11] promotes a "creative thinking spiral" that highlights the possibility to share preliminary outcomes with others as an essential component in technology that supports learning and user engagement (Figure 2). We confirm such requests by adding that sharing should also be present in design processes to facilitate a coherent experience for pupils. In fact, many CCI processes rely on sharing ideas amongst team members but fall short in sharing prototypes with externals (e.g. [8]). Hence, we want to expand sharing in CCI processes and understand this as a good motivation for recurring user tests in project layouts.

Additionally, some authors mention that the findings of projects with design partners are hard to transfer to other contexts because of rather specific and small sample design teams [12, 8]. In this connection, Moraveji et al. [9] present many contradicting results in literature on PD with children. They ascribe that to processes of "finding" qualified children raising the question: Does selecting the "right" participants really provide representative groups?

Furthermore, while many projects are conducted in school settings (e.g. [7, 13]), CCI literature hardly describes the integration of design projects in school environments or obligations of the child participants. Only Horton et al. [6] acknowledge that sound descriptions and methodologies promoting the course of action are needed to help pupils, teachers, and researchers building sustainable relationships by choice.

In summary, we understand design partners that produce good results. Yet, we argue that most CCI environments are build with informants because with design partners it is hard to benefit from insightful findings usually provided by the underlying roles of design partners mentioned by Druin [3]. Hence, we want to bring forward a model that understands design partners as a consecutive process from users to informants to (experienced) design partners. Furthermore, we reason that sharing with externals as proposed in the "creative thinking spiral" is often neglected during CCI design processes. Hence, we see a strong need to establish design processes that build upon multiple roles and can be embedded in school contexts. Thereby, we want to establish a design process that is constantly refreshed by new team members to ensure sustained team sizes in unsolicited projects and assemble representative groups.

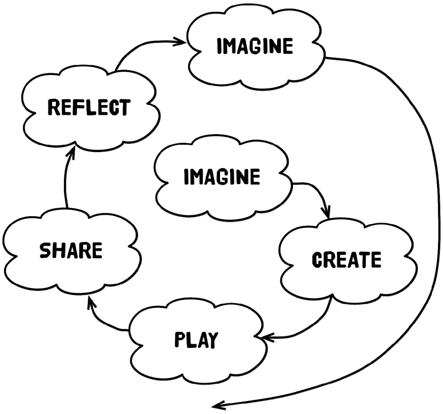


Fig. 2. "Creative thinking spiral" according to Resnick [11]

3 Project Observations

In research on the Virtual Sandbox [4], we conducted seven projects using initial learning environments (ILEs[2]) together with a total of 75 pupils (age 10–17). The participants were asked to design interactive stories acting as users on an ILE. During the projects external pupils playtested the scenarios. It helped the team members to get new perspectives and ideas. Moreover, it allowed the testers to become interested in the project and get into contact with the experienced designers. Although we did not find such remarks in CCI literature, we became aware that working with pupils on a voluntary basis often comes with uncertainties. Although the projects generally lasted only few days (1–5 days) we encountered pupils who simply did not show up after some sessions, or others who wanted to join later.

The arrangement of the projects showed that pupils have different motivation or engagement levels during a school year. For example, test periods or some days prior to vacations can hardly be used to host intense design sessions. More likely, such times were used to play around or relax. It is stunning that established CCI methodologies do not reflect in detail on everyday life of pupils that turns out to be strongly taken by school and leisure activities. Hence, to some extent this everyday life is competing against additional design projects conducted by the CCI community. The article by Horton et al. [6] confirms this point of view.

[2] For definition and instances of ILEs refer to correspondent special issue of ACM Transactions on Computing Education 10, 4 (November 2010).

To study role switching processes we asked 12 former users in a following eighth project to comment as informants on our observations on issues with ILEs when using them in teams. After an extensive discussion, we successfully engaged them as design partners in four sessions that were focused on supporting group work in ILEs applying brainstorming techniques, paper prototypes, and a variant of comicboarding (cp. [9]). Switching roles from users to informants to design partners appeared to be natural to the pupils. They were able to build their arguments on previous user sessions.

One topic clearly showed issues when designing with pupils. When asked how to document scripts for other team members they did not see sense in discussing documentation features at all. They were sure that everyone would understand their scripts. Furthermore, they did not consider approaches presented by adult team members. In such a situation introducing new test users to share current outcomes may highlight group thinking issues. Consequently, observing their peers instead of discussing with senior designers may help child designers to rethink their design or approach.

In conclusion, engaging child users as informants and design partners afterwards provided a wide range of approaches and ways of thinking not influenced by prior design decisions. Nevertheless, based on our experiences and concerns of Scaife et al. [13] we argue that design partners do accurately support developing environments but cannot sufficiently assist in identifying important following steps. In spite of compliance with the idea of design partners, such situations require input by experienced (adult) designers based on shared observations available to all team members. Accordingly, we claim for a clear process that addresses users, informants and design partners at different stages of the design process. Furthermore, we see a need for aligning such approaches with school activities and offer chances to join or leave design projects.

4 Avalanche Design Cycle

Project observations showed that pupils find it easy to switch roles when offered clear reference points during design processes. Therefore, we argue that CCI design processes should be arranged in iterative phases including pupils in four consecutive steps as users, informants, design partners, and experienced design partners. Users allow researchers and designers to observe the actual usage of environments without prompting the users to evaluate their activities. Consequential, hypotheses can be developed and discussed with the users afterwards allowing them to switch from users to informants. This transition from user to informant offers many insights and reflections allowing to omit the tester role presented by Druin [3]. In a following step, the former informants should be encouraged to support the upcoming design process as design partners. Assuring an iterative character milestones should be arranged to get new users involved in evaluation sessions. New users should then be encouraged to join the existing design team to become informants and design partners themselves. Introducing new users is important because they provide new unbiased perspectives and represent various approaches in design and implementation. Fresh team members should be tutored by experienced design partners. Based on the results of Harel [5], it is feasible to provide entry points during a design process allowing experienced pupils to tutor new design partners.

To meet these claims, we propose the Avalanche Design Cycle (ADC) which is incremental and open for new participants as drafted in Figure 3. It consists of recurring user tests to share outcomes with pupils who do not have prior knowledge on the system or the design process. After these stages the users are asked to become informants to discuss their ideas on the used environment and to suggest design ideas. Furthermore, they are encouraged to comment on observations and implications of the user tests and join the team as design partners. Thus, user tests enable design teams to set or shift focuses according to observations on user behaviors and provide opportunities to reflect on one's own activities. Additionally, the user test stages represent sound opportunities to motivate new pupils to become involved in a design process.

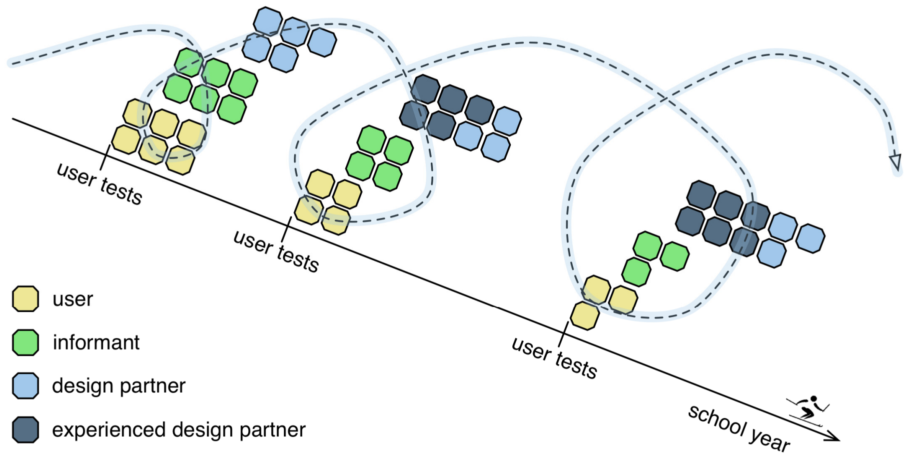


Fig. 3. The Avalanche Design Cycle (ADC) aiming at a sustained growth of design teams by offering recurring new incentives. User tests provide entry/exit points for pupils and can be aligned to schools' test periods offering playful aspects with minimal time and effort for pupils during these crunchtimes.

From our point of view, ADC should initially start with user tests on prebuilt prototypes or commercial systems. It helps building a common ground and allows pupils to influence further research directions by switching roles thereafter. After the initial user test all upcoming evaluations should be done on prototypes build by the design team.

An ADC process engages former users in succeeding sessions by asking them to comment on observations of the design team. This extensive discussion supports users in realizing that they actually could contribute design ideas build upon their experiences as users. This marks the point to ask the informants if they are interested in becoming a part of the design team. We suggest to apply brainstorming techniques, paper prototyping and further variants of PD that can be found in CCI literature (e.g. [7,8,9,12]).

For design partners it is important to observe their peers in using the prototypes. These insights help to rethink design approaches and to support experienced design partners in tutoring informants or freshly recruited design partners.

Additionally, ADC allows to conduct long term and large scale-projects fitting into dynamic scenarios at school life: By constantly introducing new team members during user tests a sustained team size is guaranteed. User tests mark a point in time that allows former members to conveniently leave a team at the end of a school year or when focusing on something else. Consequently, the concerns of Moraveji et al. [9] could be addressed because constantly attracting new users who can become design partners prevents from selecting only the "right" participants. Thus, it would give access to a wide range of pupils to act as users and let them decide afterwards if they want to become involved in the design of the respective system. Besides, school life has predefined times building on strong commitment by pupils. As we understand user tests they should be playful and fun with minimal time and effort for pupils. Thus, user tests should be aligned with known crunchtimes to keep on track in the design process while offering compensation to pupils during these times.

5 Discussion

In this paper we highlight a need for multiple consecutive roles in CCI design processes based on literature review and project observations implementing established CCI methodologies and roles. Yet, we derived ADC as a premise to identify issues, explicitly address roles, and rethink design processes with children as well as considering surrounding settings. In presenting ADC, we want to encourage other community members to reflect on their activities and connect it with ADC. Hence, we promote ADC as an ongoing "open source" prototyping process rather than being a final methodology. To open up this process, we put forward hypotheses on ADC that we hope will be challenged from different CCI community perspectives – in addition to our own research agenda on evaluating and refining the Avalanche Design Cycle:
ADC may offer...

- lightweight project takeoffs by engaging children as users
- several entry/exit points to consider unsolicited project layouts
- recurrent milestones to share outcomes of design teams with externals
- activation of insightful design roles (users, informants, design partners)
- alignment with school events and teachers projects
- constant/growing design teams by constantly introducing fresh users

Accordingly, ADC has to be comprehensively implemented and evaluated in long term studies at schools. Further adjustments are expected in alignment to school life and the transitions from user to informant to (experienced) design partner. Consequently, more concrete activities supporting the role transitions and interactions should emerge. Additionally, observations on a broad implementation of ADC should give insights on how to engage teachers, principles, and IT administrators to foster a smooth, fun, and engaging design cycle fully embedded in school life.

References

1. Adlin, T., Pruitt, J.: The Essential Persona Lifecycle: Your Guide to Building and Using Personas. Morgan Kaufmann (2010)
2. Antle, A.N.: Child-Based Personas: Need, Ability and Experience. Cognition, Technology and Work 10(2), 155–166 (2008)
3. Druin, A.: The Role of Children in the Design of New Technology. Behaviour & Information Technology 21(1), 1–25 (2002)
4. Göttel, T.: Virtual Sandbox: Adding Groupware Abilities to Scratch. In: 8th International Conference on Interaction Design and Children, pp. 158–161. ACM, New York (2009)
5. Harel, I.: Children Designers: Interdisciplinary Constructions for Learning and Knowing Mathematics in a Computer-Rich School. Ablex Publishing (1991)
6. Horton, M., Read, J.C., Mazzone, E., Sim, G., Fitton, F.: School Friendly Participatory Research Activities with Children. In: CHI 2012 Extended Abstracts on Human Factors in Computing Systems, pp. 2099–2104. ACM, New York (2012)
7. Kelly, S.R., Mazzone, E., Horton, M., Read, J.C.: Bluebells: a Design Method for Child-Centred Product Development. In: 4th Nordic Conference on Human-Computer Interaction, pp. 361–368. ACM, New York (2006)
8. Laughnan, J.: The Organization of Inventing and Prototyping Activities with Children as Design Partners. In: 2004 Conference on Interaction Design and Children, pp. 127–128. ACM, New York (2004)
9. Moraveji, N., Li, J., Ding, J., O'Kelley, P., Woolf, S.: Comicboarding: Using Comics as Proxies for Participatory Design with Children. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1371–1374. ACM, New York (2007)
10. Nielsen, J.: Paper Versus Computer Implementations as Mockup Scenarios for Heuristic Evaluation. In: IFIP TC13 3rd International Conference on Human-Computer Interaction, pp. 315–320. North-Holland Publishing Co., Amsterdam (1990)
11. Resnick, M.: Sowing the Seeds for a More Creative Society. Learning & Leading with Technology, 18–22 (2007)
12. Roussou, M., Kavalieratou, E., Doulgeridis, M.: Children Designers in the Museum: Applying Participatory Design for the Development of an Art Education Program. In: 6th International Conference on Interaction Design and Children, pp. 77–80. ACM, New York (2007)
13. Scaife, M., Rogers, Y., Aldrich, F., Davies, M.: Designing for or Designing With? Informant Design for Interactive Learning Environments. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 343–350. ACM, New York (1997)
14. Yarosh, S., Radu, I., Hunter, S., Rosenbaum, E.: Examining Values: an Analysis of Nine Years of IDC Research. In: 10th International Conference on Interaction Design and Children, pp. 136–144. ACM, New York (2011)

Likeness and Dealbreakers: Interpreting Interpersonal Compatibility from Online Music Profiles

Mo Kudeki and Karrie Karahalios

University of Illinois
kudeki@uiuc.edu, kkarahal@cs.uiuc.edu

Abstract. How much would a stranger know about you if they knew only the last song you listened to? Quite a bit, at least according to the stranger. Music taste often acts as proxy for our personalities, and we constantly perform this taste perception online. In our study, we find that good first impressions are common, bad first impressions are extreme, and that people pick up on many subtle cues about personality and whether they share common values based on data found in an online music profile. Similar motivations for listening to music and the perception of individuality are more highly valued than similar taste in genres or artists, suggesting that social media applications could benefit from incorporating these motivations in predicting compatibility between users.

Keywords: Music, listening history, taste performance, impression formation, personality, social networking service.

1 Introduction

Describing one's music taste is typically a difficult task. However, we do so indirectly every day on social networking profiles, either by explicitly listing favorite bands, or simply by listening to music on a computer or mobile device and having its song data automatically uploaded ("scrobbled") to social network sites such as Last.fm.

Up-to-the-minute information about what your friends are listening to is available on Facebook, or within music listening applications such as Spotify and Rdio. As music listening habits and histories become more public, people have the almost constant opportunity to pass judgments – good or bad – about their friends' and acquaintances' music taste.

People feel that knowing someone's music taste provides insight into them as a person [11]. Likewise, users of Last.fm feel that their music taste "signifies aspects of their identity that their community of peers may read," [2]. It is not surprising then that careful impression management and profile work is pervasive on Last.fm; people deliberately manipulate their own profiles, deleting embarrassing songs, even going so far as to queue up a list of music, hit play, and then leave the room [10]. The purpose of this curation is to communicate to their peers on social networks a particular statement, often of prestige/expertise, association with a particular subculture, or to differentiate oneself from their friends [4].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 674–681, 2013.
IFIP International Federation for Information Processing 2013

Perhaps surprisingly, stereotypes about fans of different musical genres are quite accurate, down to predicting specific personality traits and drug preferences [7]. Listening to a CD of someone's favorite songs even provides a more accurate depiction of their personality than other zero-acquaintance methods [8].

There are many studies about impression management on social networks, and about music-based stereotypes (offline). People care about how others perceive their musical taste, and put significant time and effort into making sure their profiles signal the right impression. Less is known about how people interpret these curated music profiles online. This is where we focus our study.

2 Methods

Our goal in this study was to uncover how people understand each other's music taste and personalities based on the limited music preference information available on a social networking site profile.

We chose Last.fm as the social media platform for our study for several reasons. First, the central focus of Last.fm is music taste and music discovery, so its users are familiar with viewing the music listening histories of other users. Second, Last.fm maintains a long listening history (it was founded in 2003) and collects a large amount of data about each user (every music listen scrobbled from a user's iTunes library, for example); we could easily collect months or years of listening history from actual users. Third, Last.fm's API and open privacy policy make user listening data publicly accessible [6].

We created an online survey where subjects viewed and rated others' Last.fm profiles, in addition to answering a few questions about their own music-listening habits. We then showed the subject a series of five profiles filled with other, anonymous Last.fm users' listening histories. The data visible to the subject rating a particular profile is that person's: last 15 song listens, top artists over the past 3 months with playcounts, and top artists overall with playcounts.

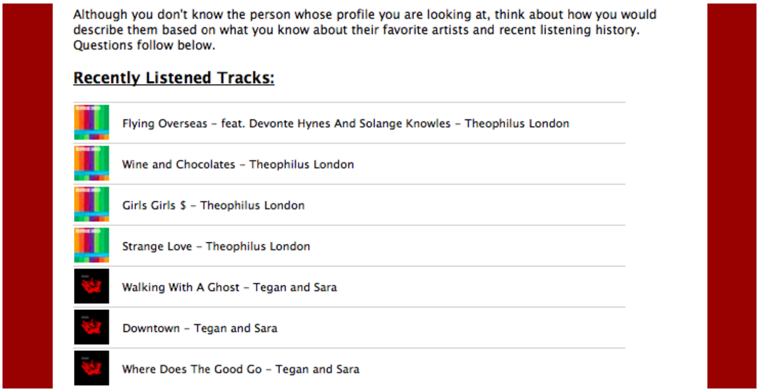


Fig. 1. Sample profile view in the survey (recent listens)

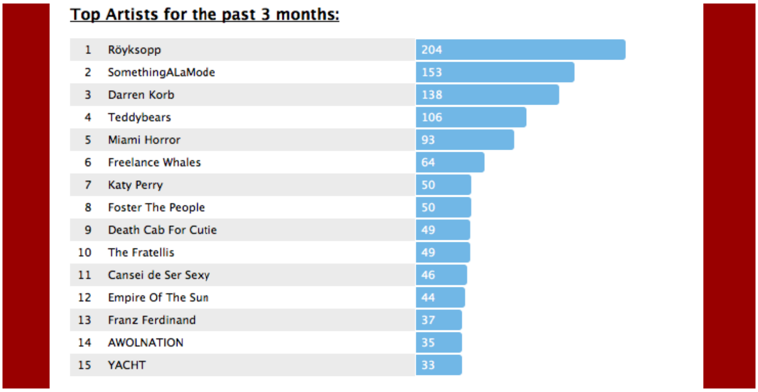


Fig. 2. Sample profile view in the survey (top artists)

We chose to use the last 15 song listens because it gave the subject a slice-of-time view of actual listening history – what songs are listened to together in one sitting, while the other two sections contain aggregate data. The intention was to provide insight to a profile owner's current taste (the 3-month window) alongside their long-term favorite artists. If the profile owner has only used Last.fm for a short time or has very constant taste, these two lists will be similar, whereas users with a longer account history or rapidly changing taste may have little to no overlap.

For each of the five profiles, the subject filled out a questionnaire where they described their perceived music taste and personality traits of the profile owner. Most questions were open-ended in short-answer format. Subjects did not have access to certain information that is typically present on a real Last.fm profile page, such as an avatar, username, age, location, and gender. We wanted to see what guesses subjects made about demographic information based solely on music listening history. Upon completion of the survey, the subject's own Last.fm profile information (the three sections listed above) was stored and rated later by four other study participants.

2.1 Design of Study Interface

Participants logged into the survey by authenticating with their Last.fm accounts. They first answered a few questions about themselves: their age, occupation, how often they log into Last.fm, and how many hours per day they spend listening to music. The subject then proceeded to rate and answer questions about each profile.

We made the design decision to show only a very limited amount of music listening history (up to 30 different artists and 10 song/artist pairs) in order to expose the subject to roughly the same amount of information one can glean by glancing at a real Last.fm profile. We limited the information presented to help simulate forming an immediate first impression as one would on Last.fm.

The survey interface (Figures 1-2) is designed to look as much like an actual Last.fm page as possible. The profile information is displayed using the same layout, styling, and images used on the Last.fm site, down to the rendering of the playcount bars' lengths. We wanted subjects to have a familiar browsing experience.

2.2 Survey Question Content

The questions about the profiles subjects rated focused on two main areas: the profile owner's music taste (three questions), and how they perceive the profile owner's personality, demographics, lifestyle, and whether they think they would get along with the profile owner (three more questions). Given the small number of artists present in a profile, it may be more difficult to make statements about a profile owner whose artists were unfamiliar. To provide context for this, we asked, "How many of the artists in this user's profile are you familiar with?"

2.3 Implementation

We built the survey using PHP with MySQL and handled the Last.fm API calls through a third-party library, PHP Last.fm API [5]. When a subject logged into the survey with their Last.fm credentials, their profile was checked to verify they had enough data/activity to qualify as a sample profile for future survey participants.

The survey requested that the subject rate the anonymous profiles of the five most recent survey participants. This assignment strategy wouldn't work unless there were already five profiles in the database, so seed profiles (selected by browsing Last.fm) were used at the beginning of the study. Subjects' responses about these profiles were discarded. When the survey was completed, we stored the five answer sets, as well as the current subject's listening history to show to future participants. We discarded the identifying username, making all the responses anonymous.

The survey ran for approximately one week. Advertising was done entirely online, on Last.fm forum pages, the Facebook page and SoundCloud group for Last.fm users, on Tastebuds.fm forums (a Last.fm-based dating website), and via email and Facebook amongst student groups at a large university. A total of 48 participants completed 225 profile ratings.

2.4 Response Coding

We used mixed methods to analyze our responses. We counted word frequencies for the listed genres and adjectives used to describe profile owners, and took averages of the questions with numerical responses (recognized artists, time per day spent listening to music). We qualitatively examined all 225 responses and performed an open-ended coding on the short-answer questions, sorting recurring ideas and patterns into categories. We focused specifically on comments about the profile owner's personality and presumed compatibility with the subject. The coding approach we used was similar to a University of Kansas study exploring stereotypes of Chinese international students, compiling an open-ended list of descriptors into a collection of stereotypes. This study used these participant-generated stereotypes as content for a later part of the study [9]. The common genres, stereotypes, and descriptors we found could be used in a follow-up study investigating the accuracy of perceived personality traits based on music profiles.

3 Results

Participants ranged in age from 18-55, averaging 23.6 years old, and around 40% were students. They were avid music listeners and active Last.fm users, self-reporting an average of 4.85 hours spent listening to music per day. 40 out of 48 log into Last.fm multiple times per month[1]. Musical recognition of bands in others' profiles was high, with subjects recognizing at least 10 artists in a particular profile more than 60% of the time.

The four most common genres listed were "rock", "indie", "pop", and "electronic", which accounted for 117 out of 374 total genre mentions. However, out of 120 unique genres mentioned, 69 of them are only mentioned a single time, meaning just once by a single subject. Many of these genres were clearly chosen to sound either very unique ("dadrock", "library ghetto"), or to fill a gap where traditional genre names lacked sufficient nuance ("foreign synthpop", "post-rock ambienty"). If we count combination genres like these that include at least one of the top four genres (e.g. "indie-folk", "electronic ambient"), this accounts for roughly another third of genre mentions (103 out of 374). The wide range of these genre names suggests that participants may have been trying to display their own musical expertise and individuality in their evaluations of others.

Responses about the profile owners clustered into two major themes: whether or not the subject thought they would get along with the profile owner (compatibility), and a more holistic analysis of the profile owner's life experience, personality, and psyche, based on cues in the listening data.

Could We Be Friends? Subjects believed they would get along with an arbitrary profile owner 53% of the time, thought they wouldn't 32% of the time, and had a more lukewarm response ("maybe" or "unsure") just 15% of the time according to our coding. However, they tended to rate others' music taste as more different than similar to their own (an average rating of 2.41 out of 5). Subjects thought liking the same bands would give them something to talk about, or that they could relate to someone with similar taste to their friends:

- "Most likely, at the very least we can bond over Muse."
- "Yep. Sigur Ros, Andrew Bird and Radiohead? The bands I don't know sound pretty fun too."
- "Yes. I know a few people that enjoy this genre; they're pretty wonderful."

Some subjects believed that even if a friendship were unlikely, they may get along as casual friends and thought they could identify what drugs they could do together:

- "They seem outgoing, bubbly, easy to converse with. They might move on before anything gets meaningful, though."
- "I do like old school rap and can maybe get high with him"

[1] Our sample may have been self-selecting, since the study was advertised as a survey about music taste. Our subjects were passionate about music, but this may not be representative of all Last.fm users.

The tendency to give others the benefit of the doubt is found often in Computer Mediated Communication studies. People rate others higher when fewer cues are present, assuming the person in question has similar attitudes to their own in the absence of more information [12]. However, even though the profiles in our study contain very few cues, certain "dealbreaker" bands evoke an intensely negative reaction. This happened in roughly 1/3 of cases, and the subject typically deemed these "dealbreakers" so repulsive that a friendship would be impossible:

- "foster the people is unforgivable"
- "No. This guy likes Kanye West and Ani DiFranco. Not even Ratatat saves that."
- "He actually enjoys listening to James Blake whereas I just jump out the building whenever I hear his pretentious dickery."

Although negative impressions were less frequent, the character descriptions of these profiles were much stronger and harsher than the positive ones were positive:

- "This is a load of mopey almost-emo indie bullshit with a couple older (much more decent) bands thrown in to help this listener justify their continued existence."
- "this seems like the sort of person who gets their opinions from reddit and people like that are generally obnoxious."

Negative social stereotypes came up in these dealbreaker ratings, most frequently the "hipster" persona, often in conjunction with indie music. Comments often mentioned the gratuitous wearing of unnecessary apparel:

- "Guy wears glasses, maybe a scarf when not completely necessary"
- "Probably at college for film studies or some bullshit, works in the college coffee shop, likes taking shitty desaturated photographs on their DSLR camera and updating their microblog."

You Are What You Listen to. Cues in music profiles were used to make conclusions about a wide variety of traits, from organizational skills, to recent activities they'd done, to which relatives played an influential role on music taste during childhood:

- "College student who recently has been studying for finals. the reason for the classical music."
- "College age female whose parents had the common sense to introduce her to The Beatles before her identity was completely whitewashed."

One particularly important value to discern from a profile was whether or not this person valued seeking out new music for themselves. Participants in our study tended to be avid music listeners, and since music is an important part of their lives and identities, they reacted negatively when they saw profiles where that didn't seem to be the case. This was also assumed to be a shared value of members of the Last.fm community, and subjects appeared to prefer people who strive for individuality. They criticized profiles they saw as "mainstream" or only included the more well-known songs/artists of a particular genre:

- "To this person, music is about being entertained, to me music is an art that is appreciated."
- "But seriously, a good indicator that I'll get along with someone is the amount of music they listen to that isn't on the billboard top 100. Nothing personal, but in general, liking lesser known music means one puts more thought into what they listen to, has a deeper taste other than 'catchy', and in general appreciates music more."

Similar motivations were more important than having bands in common, and lack of overlap can be an opportunity to learn from one another. Likewise, one could like the right band for the "wrong" reason:

- "Totally different music-listening taste – either this person would be my best friend after opening my mind up to new and wonderful things, or we will never get along because we'd both strongly believe the other should listen to our music."
- "I really like arcade fire, so I guess we like all of these things for different reasons? my aversion to muse may stem from how mainstream it is? although arcade fire is preeeetty mainstream at this point? (OH gosh I hate that word)"

Subjects were comfortable discerning specific personality traits from listening histories. Only a small minority (three responses) claimed that there would be no correlation between music taste and whether they would get along. The vast majority thought that personality came through loud and clear in the profiles:

- "Organized; they appear to prefer to listen to music by albums based on their recent tracks."
- "I don't like the bands I recognize, they sounds [sic] like a person from my same age group that I choose to not hang out with because there are too many cooler people out there"
- "Fun; musical diversity tends to be an indicator of this."

4 Conclusion

We found several identifiable trends about how people interpret each other based on music listening histories. People were bimodal about whether they could get along with someone, being either cautiously optimistic, or having an extremely negative reaction triggered by a single "dealbreaker" artist. Second, people detect behaviors, habits, motivations, and shared/unshared values from music profiles. These results give validity to the curation efforts people put towards their online music profiles. In particular, individuality and uniqueness were valued amongst our participants, warranting further study of whether this applies to a broader population.

Recent work [1] shows a number of factors that play into music listening decisions, and found, as we did, a difference "between conservative listeners and those more keen on experimenting." Another study [3] points out that what makes a mix tape good is the collection of songs chosen and the relationships between those songs. Yet there are few systems that recommend collections rather than individual items. In our study, we found that having shared artists was only one facet of music taste compatibility, and perhaps not the most important. Suggesting similar users is a common feature of social media, whether it be “neighbors” on Last.fm or matches on a dating website. We believe representing music taste as a collection of features (including dealbreakers, breadth of taste, attitudes towards finding new music, etc.) could help in suggesting better matches. We can also imagine using these findings in the reverse direction to build new music exploration interfaces to surface music enjoyed by like-minded people. The social traces we’ve found in this study could be leveraged to improve discoverability of both music and of people who are truly compatible, not just similar.

References

1. Baur, D., Büttgen, J., Butz, A.: A Large-Scale Principal Components Analysis of Long-Term Music Listening Histories. In: CHI 2012: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 1273–1276. ACM, New York (2012)
2. Fitzpatrick, C.: Scrobbling identity: Impression management on last. fm. Technomusicology: A Sand- Box Journal 1(2) (2008)
3. Hansen, D.L., Golbeck, J.: Mixing it up: recommending collections of items. In: CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1217–1226. ACM, New York (2009)
4. Liu, H.: Social network profiles as taste performances. Journal of Computer-Mediated Communication 13, article 13 (2008)
5. Oakes, M.: PHP Last.fm API (updated September 6, 2011), `https://github.com/matto1990/PHP-Last.fm-API/` (retrieved November 25, 2011)
6. Privacy Policy - Last.fm (February 11, 2009), `http://www.last.fm/legal/privacy` (retrieved December 17, 2011)
7. Rentfrow, P.J., Gosling, S.D.: The content and validity of music-genre stereotypes among college students. Psychology of Music 35, 306–325 (2007)
8. Rentfrow, P.J., Gosling, S.D.: Message in a Ballad: The role of music preferences in interpersonal perception. Psychological Science 17, 236–242 (2006)
9. Ruble, R.: Making Ourselves Understood: The Role of Previous Experience, Stereotypes, Communication Accommodation, and Anxiety in Americans’ Perceptions of Communication with Chinese Students. Thesis, University of Kansas (2011)
10. Silfverberg, S., Liikkanen, L.A., Lampinen, A.: “I’ll” press play, but I won’t listen”: profile work in a music-focused social network service. In: CSCW, pp. 207–216 (2011)
11. Voida, A., Grinter, R.E., Ducheneaut, N., Edwards, W.K., Newman, M.W.: Listening in: Practices surrounding iTunes music sharing. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2005, pp. 191–200. ACM Press, New York (2005)
12. Walther, J.B.: Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Communication Research 23, 3–43 (1996)

Paraplay: Exploring Playfulness Around Physical Console Gaming

John Downs, Frank Vetere, and Steve Howard

Department of Computing and Information Systems, The University of Melbourne, Australia
jdowns@student.unimelb.edu.au,
{f.vetere,showard}@unimelb.edu.au

Abstract. We present the concept of *paraplay*: playful activities that take place within the context of an interactive game or other play activity, but outside the activity itself. By critically examining work related to gaming and play goals and motivations we argue that the concept of playfulness should have a stronger role in our understanding of gaming sessions, and particularly social gaming sessions. In order to further understand the role of playfulness in social gaming we conducted an empirical field study of physical console gaming. Six families with a total of 32 participants were provided with an Xbox 360 console, Kinect sensor, and three casual physical video games to play together for a period of approximately two weeks. Participants were instructed to record their social gaming sessions. We conducted video analysis on these recordings as well as interviews with many of the participants. We found numerous types and examples of playfulness within the gaming session even from those who were not actively participating in the game. Drawing on the results of this study we present a taxonomy of paraplay and discuss the ways that playfulness can be exhibited in a social play session. We show that participants in a game situation act within a wider context of playfulness, according to a variety of significant roles ranging from active player through to audience member. We explore these roles and their attributes to provide a rich account of paraplay and its importance in understanding playful activities broadly.

Keywords: Games, context of gaming, metagames, physical console gaming, play, social gaming, videogames.

1 Introduction

Play has recently received increased interest and attention from the HCI (Human Computer Interaction) and CSCW (Computer Supported Cooperative Work) communities. Two separate streams of work seem to be emerging: one focusing on the ways in which games, particularly videogames, can be designed, evaluated, and understood [e.g. 1–3]; and another looking at the ways play and playfulness are to be understood and used within HCI and CSCW, including their application beyond games and the sociability of different types of playful systems [e.g. 4, 5].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 682–699, 2013.

In this paper, we examine an intersection between these trends: the playfulness that occurs on the periphery of games – and, by extension, also on the periphery of other types of play. We term this *paraplay*.

The play activity itself is just one part of a playful situation. Unlike other researchers in this area [e.g. 1], we do not see the game as existing within a thin, fixed boundary or 'magic circle'. Indeed, the periphery of the activity is a place where many different types of activities and playfulness can occur. This periphery provides an interesting site for research given the variety and richness of the interactions that take place. In the case of a physical console videogame, for example, players move in and out of the game in various ways, and audience members can indirectly participate to varying degrees by watching, commenting, and mimicking the player's movements.

Paraplay is concerned with the multiple types of playfulness that occur within the context of a wider play session. Play sessions can take the form of structured activities such as board games, card games, videogames, or sports, or unstructured activities like running around aimlessly, play fighting, or playing make-believe. In this study, however, our concern is primarily with paraplay around physical console videogames.

Our motivation for studying paraplay is twofold. Firstly, there is a theoretical gap: the concept of playfulness at the periphery of games and other activities has not been well described or theorised, particularly within HCI. Secondly, there are practical implications: designers of games and other playful systems could incorporate additional forms of play into their systems and thereby enhance the experience for players and audience members.

By studying the periphery of playful activities we can begin to understand more about their contexts and the nuanced relationships between different types of systems and their users. As entertainment systems like gaming consoles become more prevalent and more social it is important to understand the broader ramifications of their use, and to ensure that their designs do not unintentionally limit participation to a few active players while ignoring the opportunities for wider involvement. Additionally, the relationship of paraplay to play – as a peripheral activity worthy of attention in its own right – is a relationship that might also be transferrable to other domains, where situational context is critical to the activity – for example, mobile technologies and communication.

In the following sections we first discuss related literature within HCI, CSCW, play studies, and game studies. The concept of paraplay is then introduced and described in general terms, before an empirical study is presented that explores the nature of paraplay within the context of physical console videogames.

2 Related Work

This section introduces a number of foundational concepts and working definitions. First, the concept of play is described and a definition of play is provided that captures the types of play we are interested in exploring. Next, the role of play within games is considered, as is the role of the context of gaming. Finally, the concept of paraplay is more fully described.

2.1 Play

Play, and the relationship between play and technology, has attracted the attention of HCI researchers for a variety of reasons. Play has intensely social aspects [4, 5] and provides a mechanism and medium for interaction between participants irrespective of physical proximity, age differences, and technological abilities [6].

In HCI and CSCW, play is often discussed and examined in the context of games, or of other systems that are specifically designed to encourage playfulness. However, the process of defining play is fraught with difficulty. 'Play' is expected to incorporate a variety of activities: games and sports, as well as less structured activities like playing make-believe, and even jokes, word play, and rhyming [1]. The disparate nature of these activities makes the construction of a unifying definition of play and 'being playful' a difficult yet important endeavour.

From our review of the definitions of play, playfulness and games, we argue there are three key components:

Play Is Fun. The primary reason a person engages in play (whether that be in the form of a game or otherwise) is to have fun. Indeed, players of videogames tend to cite 'having fun' as one of their primary motivators [7]. Unfortunately, of all of the commonly used definitions of play and games in HCI [see 1 for a review], not one includes the idea that play should be fun.

The concept of fun is inherently subjective and contentious. What one person finds fun another may find utterly boring. Indeed, in our analysis we take an ethnocentric approach: rather than providing a single, objective definition we instead argue that an individual may stipulate what is and is not fun to them at any particular moment. In general, however, fun activities are not merely pleasurable [8]. The notion of fun suggests a light-heartedness to the activity and the way it is conducted.

Play Is Conducted for Its Own Sake. Individual games and activities can be conducted in a very serious way (for example, a high-stakes poker match), but the sort of play we explore in this paper is not serious; the previous element, fun, suggests a whimsical, non-serious nature to the activity in question. Because play is not serious, it is therefore not performed for any instrumental end, but purely for its own pleasure; it is autotelic [9].

Play Has Goals. Suits [10], in his discussion of games and play, describes the goals of play activities. In a competitive game, the ultimate goal – what Suits terms the lusory goal – is typically to win: each player is competing against one another, or in the case of a solo game, against a computer or their own high score. Players perform game actions with this lusory goal in mind. In a less structured play session, such as make-believe, the lusory goal shifts from winning to merely perpetuating the activity.

We argue that there is another category of higher-order goal that can be added to Suits' model: an overall goal for the entire session. We term this the *superlusory goal.* This superlusory goal transcends any specific game that a player might find themselves in, irrespective of the lusory goal.

The superlusory goal of a play session is usually to have fun. In a free play session, such as playing with toy trains, the lusory goal might be to perpetuate the play, but the superlusory goal of fun provides the reason and impetus to do so. In a game of chess the lusory goal might be competitive in nature, but the superlusory goal to have fun still enables the players to enjoy the game irrespective of the outcome. In a social gameplay setting, the lusory goal of one player might be to win while another player's lusory goal is to continue the game as long as possible; despite these different lusory goals, both players can share the same superlusory goal of having fun together.

Based on these characteristics, for the purposes of this paper we present the following working definition of play:

> **Play**: An individual or collective activity, conducted for its own sake, in which at least one player's superlusory goal is to have fun.

2.2 Games

As structured systems that are generally intended to be fun, games are a clear example of designed playfulness. While games can be conducted in a non-playful manner, in general HCI is concerned with games that provide the elements of play described in the previous section: they are typically fun and light-hearted activities conducted purely for the sake of pleasure. The definition of a game is also controversial and difficult to pin down, but as a working definition for our purposes we might consider a game as [cf. 11]:

> **Game:** A rule-based system that is designed to be playful.

Games include board games, card games, and so forth, as well as videogames.

Juul [3] notes that a genre has emerged of casual games designed to appeal to a variety of types of players through the use of techniques like cartoon-style graphics, simple controls, and episodic, mini-game style play. Many of these games are intended to be played socially, either while collocated or using an online multiplayer system. Physical game controllers such as Microsoft's Kinect for Xbox 360 are often employed to enable gesture-based or mimetic interfaces in these games. The coincidence of these trends – casual gaming and mimetic interfaces – seems to be partly due to the increased sociality afforded by physical gaming [12], and partly because it is typically simpler to design and develop interaction paradigms for the controls required to play these games. As a result of this shift toward casual gaming, gaming devices are now being purchased and used by a variety of users beyond the 'traditional gamer', including a high proportion of home users [7]. This has led to a corresponding interest in exploring the ways that gaming contexts are created, both in the home and elsewhere [e.g. 13, 14].

Gaming Context: The Metagame. The concept of the metagame, popularised by Garfield [15], describes the ways in which games intersect with 'real life'. Players bring their own experience, knowledge, social relationships, and goals to a gaming situation, and the metagame describes how these components interact with one another. Carter et al. [16] argued that the term 'metagame' has become somewhat muddled, with different but overlapping definitions being adopted by various game scholars. In this paper, we take the 'metagame' to refer to the wider context of a game. The game forms part of this context, and provides an activity that is a focus for attention – but there is far more happening in the environment than just the game itself.

Mueller et al. [17] noted the importance that metagames hold for the sociality of physical videogames due to the interrelationship between the game, the player, and the physical space. Players have opportunities for exploiting the ambiguity and anticipation that result from physical exertion through strategies such as false starts and the intentional use and misuse of other physical cues. These metagame elements, while outside the game proper, are nevertheless a key part of the game experience.

Play within Games. Within traditional videogames, it is clear that 'the game' (as a piece of software, a narrative, or an interaction between players and the gaming system) is typically the focus of the experience. However, a player might choose to play in a different manner than the designer intended; as such, they are acting within the metagame, but the types of actions they perform might be completely different to that envisioned by the game designer.

Voida et al. [18] identified the 'incorrect' playing of games as an intentional play type. Using an example of a racing game player who intentionally crashes their cars in order to watch the ensuing carnage, they argue that 'the primary form of engagement during those special effects seemed to be between each individual gamer and the game and not between two members of the gaming. This type of 'trifling' [10] with the rules is framed as disruptive to the central purpose of gameplay, rather than as an opportunity for a game to be appropriated for playful purposes outside its core rule set. In contrast, Consalvo [2] notes that players can gain enjoyment from a game in many ways, such as playfully modifying the game so that characters have an absurd appearance.

We can also consider the example from Mueller et al.'s [17] discussion of the creation of meaning in an exertion game. In a game of table tennis, a particularly powerful shot might cause the opposing player to have to run across the room to retrieve the ball. Although acting within the rules of the game, the player is essentially appropriating the game in order to derive additional pleasure and playfulness.

2.3 Paraplay

The examples in the previous section provide a foundation to explore the concept of paraplay. Each example includes the idea of exploiting the game to achieve playful purposes beyond the prescribed constraints, and of a higher-level (superlusory) goal being more important than the specific game activity underway. This idea of playing beyond the game itself is what we term paraplay, and can be extended beyond games to any types of play. We define paraplay as:

Paraplay: Play that occurs on the periphery of other types of play.

Within this definition, 'play' refers to any playful activity as described above, whether structured or unstructured.

'Other types of play' refers to the canonical playful activity: the primary activity that is ostensibly the reason for the interaction. In the case of a board game, the canonical playful activity is the set of actions and moves that form the play of the game. In the case of a videogame, the canonical playful activity is the game activity itself: the actual race, battle, or sequence of moves; what most players would consider to be the game itself. Although other elements of the game system such as splash screens, menus, and so forth are part of the software and have been incorporated by the designer, they are not part of the canonical playful activity.

Audience members – those participants who are co-present but not actively playing – are also considered to be within the play context but outside of the canonical playful activity. These audience members might have different levels of interaction in the situation. O'Hara et al. [19] described differentiated participation by audience members in an alternate reality game: some audience members become tightly involved in a game session by adding commentary, making suggestions, and participating in other ways, while other audience members remain detached from the game, only becoming involved peripherally and sporadically. At both extremes – and at all levels in between – audience members can be involved in various playful ways.

Finally, 'the periphery' of the play is the play context excluding the canonical playful activity. In the case of a game, the play context is referred to as the metagame; this includes activities that are peripheral to the game, which may be playful (paraplay – for example, using a board game's pieces to act out a playful narrative) or non-playful (for example, conversations and social interactions between players). Therefore, paraplay is playfulness at the periphery of the game itself, but still within the metagame.

When discussing the periphery of play it is tempting to incorporate or assume the concept of the 'magic circle'. Originally described by Huizinga [20], the magic circle is considered a hard boundary of time and space in which play occurs; inside this boundary special rules apply. However, the concept of the magic circle is fraught [21, 22], and the implication that there is a hard boundary between 'Play' and 'Not-Play' is heavily contested, particularly when applied to digital games and play. Indeed, in this work we are specifically interested in exploring the nuanced forms of interaction at the periphery of games and play, so the magic circle is not an appropriate metaphor for our analysis.

Of course, there are many activities that occur on the periphery of the canonical playful activity that are not playful, and therefore not considered paraplay. For example, an audience member making a player a cup of tea would be acting at the periphery of the play, but unless the tea-making was performed in a particularly playful manner – say, by the tea-maker juggling the teapot – it would not be classified as paraplay.

The superlusory goal concept provides a lens through which to consider paraplay. Because the superlusory goal applies to the entire play session, it also follows that individual canonical playful activities are not the only elements covered by this goal. Indeed, in any given gaming session, players might participate in multiple distinct games in sequence. The same superlusory goal applies irrespective of the specific game in progress. By extension, it also follows that other activities within the play session – even if they are not part of the canonical playful activity itself – can be covered by the superlusory goal. These activities might be peripheral or even irrelevant to the game underway, but they do take place within the gaming context. All of these other playful activities are paraplay.

3 Investigating Paraplay: Approach

Although metagame elements have been described by others, as have the different types of play within a wider play session, these descriptions have not been well integrated. Additionally, there has not been any work specifically examining these alternative forms of play. As such, we report a study designed to specifically address the question: What is the nature of paraplay? The following sections describe the study and lead to a discussion of paraplay in the context of physical videogame play.

3.1 Aim

Our introduction of paraplay noted that paraplay is broader than just games. Playfulness can occur at the periphery of any type of play. However, in this study we were specifically interested in exploring the nature of paraplay around physical console gaming. These gaming systems (such as Nintendo's Wii and Microsoft's Kinect for Xbox 360) enable a rich set of physical and social interactions, including many that are playful in nature. In order to study this we conducted an observational study.

3.2 Method

We conducted a study with six families, comprising a total of 32 participants. Each family participated for approximately two weeks. Families were provided with an Xbox 360, Kinect sensor, and three Kinect-enabled videogames:

- **Kinect Adventures.** A number of mini-games of varying levels of physicality and complexity, most of which support two simultaneous players.
- **Kinect Sports.** Various sport-themed mini-games (including sprinting, bowling, and volleyball), also with multiplayer support.
- **Dance Central.** Music-based videogame in which players perform specified dance moves and receive feedback about their accuracy. Includes rudimentary multiplayer support.

Each family was also provided with a fixed digital video camera to record themselves playing games. All studies were conducted in family homes.

3.3 Participants

Based on previous work [e.g. 4, 23], we expected that familial environments would be particularly conducive to rich forms of play as well as other forms of social interaction. Participants were recruited through snowball sampling and through advertisements in local newsletters. All families who volunteered to participate, and who met the criteria, participated in the study. Table 1 lists the participating families and the demographics of these family members.

Table 1. List of participants

	Child (<10 years)		Youth (10-18 years)		Adult	
Family	Female	Male	Female	Male	Female	Male
1			P1		P2, P3	
2				P4, P5	P6	P7, P8
3		P9, P10			P11, P12	P13, P14
4		P15	P16		P17, P18, P19, P20	P21, P22, P23
5	P24				P25, P26, P27	P28
6				P29	P30, P31	P32

3.4 Data Collection and Analysis

Family members were asked to record all gameplay sessions for the duration of the study. It is conceivable that participants' behaviour was affected by the presence of our recording equipment; however this limitation applies to all observational research. In total, approximately 19 hours of gameplay video data was collected and analysed. While video data were the primary source of data for this study, interviews were also conducted with the families at the beginning and end of each study. Interviews were primarily intended to collect contextual information about the family's broader social interactions, gameplay habits, and level of physical activity.

Interview data were analysed using a Grounded Theory approach, and are presented here only when they provide relevant contextual information to add to the video data. Video data were catalogued and transcribed based on the Interaction Analysis [24] approach. The transcripts were then analysed using a Grounded Theory approach in order to identify high-level themes and concepts of relevance. These themes were manipulated (combined and separated) until distinct themes were isolated.

We then applied our definition of paraplay as a selective lens to sample this wider set of themes. Specifically, we looked for activities that were playful in nature and that were within the game context, but at the periphery of the game itself.

One complexity was dealing with the ambiguous nature of the periphery of play; this was accentuated by the physicality of the games that we used in our study. In the case of videogames, we argue that elements like the menu system, pre-game setup, and so forth are not part of 'the game' even though they have been designed and are included in the game software. Instead, for our purposes, 'the game' is the canonical playful activity: the actual game event that players participate in (e.g. the act of dancing in Dance Central, or of bowling or jogging in Kinect Sports).

During our analysis we also observed a distinction between the player and audience member roles. For the remainder of this paper we use the following terminology: *participants* are all people within the gaming environment; *players* are the participants who are currently playing the game; and *audience members* are participants who are not actively playing the game.

4 Investigating Paraplay: Findings

The following sections discuss the multiple ways in which playfulness was created and exhibited within the gaming environment. The findings presented in this section highlight the fact that within a videogaming session there are a number of ways the game can become a resource for play. The gameplay itself is only one of these: the design of the game, the gaming technology, and the game rules all provide mechanisms for play to emerge. Similarly, the social context in which the games are played, the existing social roles of the participants, and the physical environment all affect the types of play that can emerge around a game.

Our findings both support the existence of paraplay and suggest that there are multiple forms of paraplay. The following sections describe the types of paraplay in more detail. Themes were drawn almost exclusively from the video analysis; interview data were primarily used to add context.

4.1 Paraplay and the Player

The majority of gaming sessions we observed were social in nature – they included both players and audience members.

Some players would use the game as a way to deliberately create a performance and to add playfulness and humour to the situation. This was often done for the benefit of audience members, but was also occasionally observed during solo gaming sessions when no audience was present. These performances took a number of different forms.

Sound Effects. Younger players often made audible sound effects to accompany their actions. These were sometimes associated with the real-world analogue of the game action – for example, grunting and whooshing while serving a tennis ball. Audience members also added sound effects in an attempt to participate in the game without playing directly.

Singing and Dancing Along with the Game. Kinect Sports and Dance Central both used popular songs in various parts of the games. Players, and occasionally audience members, sang along and encouraged others to sing along, particularly while they were waiting for the game to load or for other players to join in.

Taking a Bow. At the conclusion of a successful level, players would sometimes take a bow, playing along with the performance characteristic of the situation. Occasionally audience members would cheer or applaud the player, participating in the make-believe theatrical aspect of the game. Sometimes players also exploited particularly bad game performances for humorous purposes by bowing ironically.

4.2 Paraplay and the Audience

Rather than simply sitting back and watching, audience members became involved through laughing along, teasing, providing commentary, and even indirectly participating in the game.

Teasing and Trash Talking. Audience members and players used the game as an opportunity for paraplay in the form of playful banter, teasing, and trash talking. For example, the following exchange took place as a mother (P11) and son (P10) watched the father (P13) play the Kinect Adventures mini-game 20,000 Leaks:

P11: "watch Daddy do it by himself, OK?"
P10: "maybe Daddy's very bad"
P11: "maybe Daddy's really bad, we'll see"
P10: "cos he can't get any higher than us"
P11: "yeah, break that glass!" (to the game)
P10: "make it sink! Make Daddy sink! "

In this example, the father was playing a game in which he used his body to plug holes in an underwater aquarium as fish continued to create new holes.

Audience Emulation. Audience members 'played along' with the game by performing many of the game actions – leaning, jumping, moving their head or arms, and so forth. This was often done even when the player was facing away from the audience member and therefore could not see them; likewise, the sensor was not able to detect the audience members and so they were not actual 'players' in the game.

Commentary. Audience members frequently provided commentary of the game, describing players' movements and actions, and assessing whether they were contributing to the player's success. Sometimes these commentaries were provided in a form that suited the game – for example, in a track and field event, commentary might be in the stereotypical 'sports announcer' tone.

In this example, P15 is watching his father (P21) play bowling on Kinect Sports and providing advice and commentary about his performance, as well as playing along with the game actions.

4.3 Paraplay and Transitions

Any individual game's context had a number of transition points. One transition type was when players moved into a play session, or changed between games or forms of play. Another transition type was when participants shifted in roles, from player to audience or from audience to player. Paraplay was seen to emerge at each of these transition points.

Transitioning the Space. The preparation of the play space was often a source of amusement. Depending on the size of the space and the amount of ambient light, players sometimes had to stand on top of furniture, perform strange contortions, or even move furniture to other parts of the room in order for the sensor to work correctly. These peculiar actions were seen as amusing and playful by the players and audience members.

Transitioning to Other Forms of Play. One family commented that their gaming provided an opportunity to participate in other forms of play together. In one example, at the conclusion of an evening playing Kinect games, this family played 'hide and seek' throughout their house.

Transitioning Between Roles. Participants shifted between roles throughout the play session, and often used these natural inflection points as times for paraplay. For example, the following extract from family 3 shows the grandmother (P12) engaging in a short play session with her grandson (P10) based around a role transition point and a silly hat she found:

P10: (is wearing a baseball cap)
P12: picks up silly hat, puts on
"can I play with this on my head?", dances around
P12: "OK, you've got a hat, I need this one"
moves into play area to begin bowling
balances hat precariously as she bowls ball
hat falls off, dances with hands on hip, turns around 360 degrees
Game announces strike
P12: laughs, "see, it was the hat, because the hat, that's why! You have to put the hat on and you will... (come) first"

In this example, while she was an audience member she insisted on wearing the silly hat so that she would be no different to her grandson (P10). When she scored a strike, she jokingly attributed her success to her wearing the hat.

4.4 Paraplay and the Technology

The game itself, and the Xbox and Kinect technology, provided opportunities and affordances for paraplay to emerge. Game features, such as virtual game characters and the game's use of the Kinect camera for taking photos and video, were sometimes designed and sometimes appropriated for playful purposes. Additionally, the mechanism for controlling one's avatar – 'avateering' – provided sufficient flexibility that players could move their avatars in playful and interesting ways.

Virtual Characters. In some games there were elements designed to provide opportunities for serendipitous and planned forms of playfulness. These forms of paraplay were enabled by the game software itself. In the case of Kinect Sports, these took the form of virtual game characters.

For example, during a game of Kinect Sports bowling, P15 (the son in family 4) accidentally threw the ball into the virtual crowd, eliciting a bemused reaction from his father (P21). P15 was unaware the game included this particular 'feature', and found it extremely entertaining - so much so that he went to get his sister (P16) and attempted to reconstruct the manoeuvre for her.

Similarly, before the starting gun fired in some of the track events, players could move into other characters' lanes and elicit humourous responses from these characters. Players also discovered that in the main menu of the game, waving and clapping caused the virtual stadium crowd to respond by cheering or clapping along. Typically a player would discover these game elements by accident and then, amused with the results, would attempt to demonstrate them to others.

Game Photos and Replays. Many games showed short video replays or mid-game photographs of players at the end of each level.

As players became used to the games, they would learn the places in the game levels where the system took photographs, and would often prepare playful, amusing, or silly poses in preparation for these points. For example, during the Kinect Adventures mini-game River Rush, P5 performed a characteristic Michael Jackson-style pose at the end of the game and then waited for the photographic replays so he could show others.

These points were a common place for interaction between players and audience members, and for paraplay to emerge through playfulness and humour. Audience members would laugh at the players' photographs (whether intentionally posed or not), and sometimes imitate or tease them.

Avatars. Appearance of players' avatars was a common source of amusement, particularly when their appearance was incongruous with players' real appearance. In the following example, family 4 note the gender and hairstyle of the father's (P21) avatar, teasing him playfully:

P17: "Daddy's got pigtails too"
P21: "what?"
P23: "I like your hair mate"
P21: "thanks mate, it took me a long time to do it, you know that?"
P17 and 20 laugh

Additionally, in one of the clearest examples of paraplay, players exploited the games' mimetic interface controls to make their avatars behave in amusing ways for their own benefit or for the benefit of the audience. However, this ability was only present in some games and at certain points in the game, For example, P16 (the daughter of P21) took advantage of the fact that the medal ceremony in Kinect Sports provides full mimetic control over her avatar. She used the opportunity to perform an impromptu dance routine. Similarly, players sometimes experimented with the boundaries of the sensor's capabilities, such as when two participants tried hugging each other to see if their avatars would mirror their movements:

P1: "now cuddle"
P1 and P3 hug
P3: (looking at avatars on screen) "they're not cuddling"

Technical Problems. Although the sensor technology was reasonably robust, it occasionally malfunctioned in interesting and amusing ways, and sometimes these malfunctions were appropriated for playful purposes by players and audience members.

Generally technical issues were exhibited through players' avatars moving in unusual ways and performing strange contortions. After some initial confusion about why their avatars were behaving in this way, participants began to treat these situations as opportunities for play. Some participants attempted to emulate the avatar's contorted position while others simply laughed, teased, and commented.

5 Discussion

Section 4 introduced a number of themes that emerged from our study. A common thread throughout each of these themes was the idea of playfulness. Participants of all types – players and audience members alike – used the game session as a means for play, and this play was not restricted to the gameplay itself. In this section we discuss some of the overarching themes and important concepts that came from the study, and relate these to our understanding of paraplay and to other literature in the area of play and HCI.

5.1 Paraplay and the Player

There are a variety of types of play engaged in by the active player. The most obvious of these is the play associated with the game itself, but this is excluded from paraplay by definition. We encountered a number of other ways in which players 'played' within the gaming context, particularly around creating a deliberately staged 'performance' for audience members, and by using competitiveness as a means of play.

Performance. Performance was a clear example of playful behaviour. Frequently, players would show off for audience members, pose for in-game photos, take a bow (to acknowledge their good performance, or to ironically highlight their bad performance), and sing and dance along with game music. These activities were particularly prevalent when players had an interested and engaged audience, but some players found amusement in these performing activities even when there was no audience (or when the audience was not paying attention).

Performance-like playfulness was also highlighted by Lindley et al. [5] in their discussion of the Wayve messaging device. They argued that the device 'became a legitimate means of drawing attention to oneself, in a way that can be understood as "broadcasting identity", both of oneself and of the family as a whole' (p. 2359). Similarly, the types of performance observed in our study were playful means of drawing attention, showing off, and promoting an individual and group identity.

Importance of Physicality. As players moved around the play space, adopted game poses, performed physical gestures, and interacted with the game and other participants, the physicality of their actions made their playfulness all the more salient. In a discussion of physical game controllers, [25] note that they 'expect that (physical game controllers) may enhance audience enjoyment' and suggest that empirical studies be designed to explore this phenomenon. Our study indicates that the physicality of Kinect gaming does indeed affect enjoyment, and provides a richer and more engaging platform for play to occur.

Rhetoric of Competitiveness. The levels of competitiveness in a game varied depending on specific players, player combinations, and specific game sessions. In general, however, players became or pretended to become particularly competitive, and would engage in teasing and disparaging comments about other players' game performance. Occasionally this was not intended playfully, but generally it was, and the game's announcements of scores, leaderboards, and unlocked achievements provided focal points for this type of playful interaction and bantering. In these situations, the rhetoric of competitiveness was used as a mechanism for playfulness [cf. 18].

5.2 Paraplay and the Audience

Similarly to [23], we found a role distinction between audience members and players. Far from being passive observers, though, audience members were active participants in the gaming sessions, having and enabling fun in a variety of ways. Throughout our study, some clear examples of this phenomenon were observed, including audience members providing commentary on the game and emulating the players.

Commentary. Some audience members – even those who never played the games at all – participated in the gaming experience by acting as commentators. In some games, the role of commentator suited the theme of the game (for example, sports announcements in many of the Kinect Sports events) and audience members in turn fit their commentary into the style. In other games, general commentary was provided on player performance and on the game design and aesthetic. Irrespective of the type of commentary and who was providing it, banter and playful teasing often featured, as it did in the analysis of messaging through the Wayve device [5].

Emulation, Mimicry, and Playing Along. Audience members also emulated and mimicked players, particularly the physical movements they made as part of game actions. Mimicry was often intended to playfully highlight how ridiculous players looked as they performed the game movements. Occasionally it was used for audience members to rehearse for an upcoming turn, to simply try out the movements themselves, or to feel like they were 'playing along' – similar to group karaoke games [26].

5.3 Paraplay and Transitions

The fluidity of the game sessions meant that there were frequent transitions in the activities underway, and in the roles held by participants.

Transition into Play. At the beginning of each gaming session, there was frequently a marked transition of the space as it was prepared and configured for gaming. These initial transitions served as points of demarcation for a gaming session to begin.

In their study of social board gaming, Xu et al. [27] found that game 'chores' – setting up, scorekeeping, managing turn taking, and so forth – were often considered to be fun and part of the game experience. Rather than being perceived as undesirable 'make work', chores presented points for reflection, socialisation, and transition. Similarly, the transitioning of a space into a play area for Kinect gaming – through the reconfiguration of furniture, setup of game options, and so forth – all presented points at which participants engaged in playfulness.

Transitions between Activities. Additionally, there were transition points that occurred when participants decided to switch games, or even to switch activities entirely. One of the clearest examples of paraplay was the transition from playing videogames to other forms of play. In some cases we were told by participants that at the conclusion of their videogaming sessions they started playing other games such as hide and seek throughout the house, as well as unstructured forms of play such as running around outside together.

Transitions between Roles. The roles of 'player' and 'audience member' were not fixed. As in Voida & Greenberg's study [26], players would take turns, swap roles mid-session, or take over another player's turn due to a variety of reasons: a telephone ringing, a player giving up, or a player becoming too tired to continue. Davis et al. [4] argued that intergenerational play has both episodic and flexible characteristics, and we suggest that transitions between roles are one key way in which this flexibility can be introduced. Indeed, the Kinect game technology supported these – sometimes abrupt – changes in role and enabled players to change or take over each other's turn as necessary. As such, there were two clearly identifiable transition points to demarcate a particular participant's membership in the roles.

When audience members became players, this was frequently within a multiplayer turn-based game, and so the incoming player was a competitor to the outgoing player. In these cases there were often incidents of teasing and trash-talking between these participants (for example, 'I can easily beat that score!', 'you're useless, let me show you how it's done', etc.).

There were also times when audience members became players to permanently take over another player's turn. In these cases the incoming player would sometimes reconfigure the game by selecting a new avatar or changing the game entirely.

As players moved into the audience member role, they frequently engaged in parallel forms of playful behaviour. When playing a turn-based competitive game, they would use the opportunity to tease and trash-talk the incoming player; they would also highlight their own performance and suggest that their scores could not be beaten.

In other cases, as players transitioned into audience members they intentionally engaged in playful self-deprecation. This was sometimes intended ironically, and sometimes merely as 'soft' commentary on their own performance.

The frequency and ease of these transitions highlights the fluidity of the boundaries between the game, the gaming context, and the rest of the environment. Again, we argue that the magic circle metaphor places too much emphasis on the distinction between the game (or play) and the rest of the environment, when in our analysis we are concerned with the varied ways in which this boundary is morphed and moulded.

5.4 Paraplay and the Technology

The gaming technology used in our study – the Xbox 360, Kinect sensor, and the game software – enabled play in a variety of ways, specifically with regards to the physicality of the control system and the designed playfulness in the game software.

Mimetic Controls. The Kinect's mimetic interface controls, combined with (in some games) a direct mapping between a player's physical movements and their avatar's movements, also provided the freedom for players to 'avateer' – to control their avatar in a playful and amusing way. Players were able to simply strike amusing poses, or to control their avatar to perform (or pretend to perform) actions in the virtual world.

Designed Game Elements. In some cases, the game software also provided elements outside of 'the game' (as it is defined above) that were adopted for playful purposes. The menu system in Kinect Sports provides one example: players were able to add fireworks and encourage the virtual crowd to clap and cheer by moving their body in specific ways. Similarly, in some of the track events in Kinect Sports, players could interact with non-player characters by pretending to shove and push them, provoking amusing reactions. In these examples players would often discover these features accidentally and would be either confused or amused by the results until they were able to successfully reproduce them.

Voida et al. [18] discussed a similar phenomenon, which they termed 'falling prey to the computer's holding power' (p. 377). They noted that in these situations the engagement was between the player and the game software, rather than between players. Consalvo [2] described these types of play as ways of 'gaining more enjoyment from the game, in a variety of ways'. We argue these playful game software elements are just some of a great many factors that enable a variety of playful behaviours surrounding a game.

Technical Problems. The complexity of the Kinect's sensor technology meant that there were a number of technical issues observed throughout the study. In some cases the players and audience members were able to appropriate these technical problems and incorporate them into their play session. For example, when the sensor malfunctioned and caused players' avatars to move in unusual ways, audience members found this an amusing source of conversation and banter.

This is similar to the notion of seamful design. Seams – technical limitations or barriers in a technology – are typically considered flaws, and a traditional perspective argues that 'getting rid of them is just a matter of time and improved service' [28]. Challenging this view, [28] argue that seams can and should be deliberately used by designers 'in ways that users (might) find useful or interesting'. With regards to Kinect games, designing improved sensors and recognition algorithms is certainly one approach, but given the casual and fun nature of the games we used, another design approach would be to embrace these technical issues and use them to enhance playfulness and ambiguity.

6 Conclusions

In this paper, we have introduced the concept of *paraplay*. We define paraplay as play that occurs on the periphery of other types of play. In our study, the canonical playful activities were Kinect-enabled videogames, and our data support the idea that within the gaming session – the metagame – multiple types of play are present. The gameplay itself is just one of these.

From our study we identified two distinct roles that participants adopt: active player and audience member. As participants take these roles, and as they transition between them, various types of playful behaviour emerge. These behaviours are both enabled and mediated by factors such as the game's design, the gaming technology, pre-existing social relationships, and the superlusory goal of the session.

Although we have a relatively small data set to draw from, the findings and analysis are nevertheless interesting enough to suggest a number of implications for the designers of playful technologies and physical videogames. These include:

1. Be sensitive to the fact that the game may be a spectacle, observed by audience members. Attempt to make games engaging for audience members.

2. Build up a sense of playfulness throughout the entire gaming session, including within elements such as menus and splash screen. This will help to provide explicit hooks to other games and other types of play, even non-digital play.
3. Where possible, include some unstructured game elements that do not have specific goals or requirements. These provide opportunities for creative play and paraplay to emerge.
4. Sensor-based gaming technologies often include 'seams'. Rather than treating these as bugs to be fixed, consider how these seams might be appropriated by users (for example, as physical spaces where audience members can 'play along' without fear of the technology directly including them into the game).
5. When designing an avatar-based control system, provide full mimetic control if possible, even if it is not necessary for the game itself. This will allow players to 'avateer' in playful ways.

The concept of paraplay also leads to a number of open issues for future research to address. These include identifying the types of play activities that have paraplay around them, the ways in which games and other activities can be designed to enhance or manipulate paraplay, and a further description of the elements of our taxonomy of paraplay.

Our research contributes to the theoretical understanding of social play within HCI and CSCW. It also provides design implications based on the different roles, types of play, and superlusory goals that players adopt during a videogaming session.

Acknowledgments. We wish to acknowledge funding from the Australian Research Council (ARC) grant DP110101999. We also wish to acknowledge support from the Institute for a Broadband-Enabled Society (IBES) at The University of Melbourne.

References

1. Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. MIT Press, London (2004)
2. Consalvo, M.: Cheating: Gaining Advantage in Videogames. MIT Press (2007)
3. Juul, J.: A Casual Revolution: Reinventing Video Games and Their Players. MIT Press, Cambridge (2009)
4. Davis, H., Vetere, F., Gibbs, M., Francis, P.: Come play with me: Designing technologies for intergenerational play. UAIS 11, 17–29 (2012)
5. Lindley, S.E., Harper, R., Sellen, A.: Designing a technological playground: A field study of the emergence of play in household messaging. In: Proc. CHI 2010, Atlanta, Georgia, USA, p. 2360 (2010)
6. Davis, H., Vetere, F., Francis, P., Gibbs, M., Howard, S.: "I wish we could get together": Exploring intergenerational play across a distance via a "Magic Box". J. Intergen. Relationships 6, 191 (2008)
7. Interactive Games & Entertainment Association: Digital Australia 2012. Bond University, Evanleigh, Australia (2011)

8. Blythe, M., Hassenzahl, M.: The semantics of fun: Differentiating enjoyable experiences. In: Blythe, M.A., Overbeeke, K., Monk, A.F., Wright, P.C. (eds.) Funology: From Usability to Enjoyment, pp. 91–100. Kluwer Academic Publishers (2004)
9. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper Perennial Modern Classics, New York (1990)
10. Suits, B.: The Grasshopper: Games, Life and Utopia. Broadview Press (2005)
11. Brynskov, M., Ludvigsen, M.: Mock games: A new genre of pervasive play. In: Proc. DIS 2006, pp. 169–178. ACM, University Park (2006)
12. Whitson, J., Eaket, C., Greenspan, B., Tran, M.Q., King, N.: Neo-immersion: Awareness and engagement in gameplay. In: Proc. Future Play 2008, Toronto, Canada, p. 220 (2008)
13. Flynn, B.: Geography of the digital hearth. Information, Communication and Society 6, 551–576 (2003)
14. Sall, A., Grinter, R.E.: Let's get physical! In, out and around the gaming circle of physical gaming at home. JCSCW 16, 199–229 (2007)
15. Garfield, R.: Metagames. In: Dietz, J. (ed.) Horsemen of the Apocalypse: Essays on Roleplaying, pp. 16–21. Jolly Roger Games, Charleston (2000)
16. Carter, M., Gibbs, M., Harrop, M.: Metagames, paragames and orthogames: A new vocabulary. In: Proc. FDG 2012, Raleigh, NC (2012)
17. Mueller, F., Gibbs, M.R., Vetere, F.: Design influence on social play in distributed exertion games. In: Proc. CHI 2009, Boston, MA, pp. 1539–1548 (2009)
18. Voida, A., Carpendale, S., Greenberg, S.: The individual and the group in console gaming. In: Proc. CSCW 2010, Savannah, GA, pp. 371–380 (2010)
19. O'Hara, K., Grian, H., Williams, J.: Participation, collaboration and spectatorship in an alternate reality game. In: Proc. OzCHI 2008, pp. 130–139. ACM, Cairns (2008)
20. Huizinga, J.: Homo Ludens: A Study of the Play Element in Culture. J. & J. Harper Editions, New York (1970)
21. Consalvo, M.: There is no magic circle. Games and Culture 4, 408–417 (2009)
22. Juul, J.: The magic circle and the puzzle piece. In: Proc. Phil. Computer Games (2008)
23. Voida, A., Greenberg, S.: Console gaming across generations: Exploring intergenerational interactions in collocated console gaming. JUAIS 11, 45–56 (2012)
24. Jordan, B., Henderson, A.: Interaction analysis: Foundations and practice. The Journal of the Learning Sciences 4, 39–103 (1995)
25. De Kort, Y.A.W., IJsselsteijn, A., Gajadhar, B.J.: People, places, and play: A research framework for digital game experience in a socio-spatial context. Social Psychology 6, 823–830 (2007)
26. Voida, A., Greenberg, S.: Wii all play: The console game as a computational meeting place. In: Proc. CHI 2009, Boston, MA, pp. 1559–1568 (2009)
27. Xu, Y., Barba, E., Radu, I., Gandy, M., MacIntyre, B.: Chores are fun: Understanding social play in board games for digital tabletop game design. In: Proc. DiGRA 2011, Utrecht, Holland (2011)
28. Chalmers, M., Dieberger, A., Höök, K., Rudström, Å.: Social navigation and seamful design. Cognitive Studies 11, 1–11 (2004)

Exploring Twitter Interactions through Visualization Techniques: Users Impressions and New Possibilities

Guilherme Coletto Rotta, Vinícius Silva de Lemos, Ana Luiza Moura da Cunha, Isabel Harb Manssour, Milene Selbach Silveira, and André Fagundes Pase

Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga 6681, Prédio 32, 90619-900, Porto Alegre, RS, Brazil
{guilherme.rotta,viniciusdelemos,analza.cunha}@gmail.com, {isabel.manssour,milene.silveira,afpase}@pucrs.br

Abstract. Social networking websites have been taking a huge space on the Internet. With millions of members spread across the world, these services connect their users through "friendship" relations, and within a short period of time the information is disseminated in the network. Considering Twitter, we investigated how standard techniques of information visualization can improve the analysis of its users' behavior, i.e. the group of people who most closely relate, and the frequency and type of messages they send. This investigation was done through user's studies with a set of visualization techniques implemented in an interactive system prototype.

Keywords: visualization techniques, social networks, twitter users' behavior.

1 Introduction

The advent of Web 2.0 has aroused the emersion of new services for the Internet over recent years, giving users a higher level of – user-to-user – interaction in the network. Within this context, several applications focused on the relationship among users were developed, created to bring together people with specific interests in common, to facilitate the contact between distant friends and, for instance, the contact between a company and its customers. Social networks are, mostly, websites created with these purposes, and have millions of members spread around the world [1]. With the ever-increasing number of members, a large amount of data is created on these networks and flows over the Internet. Making an analysis of this information can be difficult, since the complexity and the way it is organized do not assist in this task.

Taking into account the amount of information published in these networks, and the growing speed in which this information is updated and exchanged, the research herein presented focuses on how standard and interactive techniques of information visualization can help in the relation and extraction of new information based on what has already been published.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 700–707, 2013.

Considering Twitter, it differs from other social networking sites because a friendship link doesn't need to be accepted by both parts: in order to receive tweets from a user, a person must follow him/her; however, the user is not obliged to follow this person back. Since Twitter just allows you to view the friends of one member at a time, and through a simple list that appears on every profile page, information visualization techniques can help to better view the network of contacts, and to analyze the frequency that individuals or groups send tweets and what kind of tweets they send – the Twitter Users' Behavior.

In order to investigate this, we made user studies using a system prototype developed in our University, which provides visualization techniques that allow the analysis of users' connections, and the frequency of tweets sent by one or a group of users, and, mainly, it provides ways to classify these tweets, a feature difficult to find in most existing visualization tools for Twitter. Thus, as our main contribution, we demonstrate how interesting for the user point of view (given different user profiles) is to allow an integrated manner to explore and analyze the interactions of one user or a group of users in a social network.

This next section presents the visualization techniques available in the prototype. The user studies and a discussion on how they can help to analyze users' behavior are following presented. Some conclusions and future works are presented in the last section.

2 Visualization Techniques to Analyze Users' Behavior

There are many tools that use information visualization techniques to help in the analysis of users' behavior, as presented in some previous works [2-5]. In the work herein presented, we used the prototype system developed by Rotta et al. [6], called DeepTwitter.

DeepTwitter provides standard visualization techniques, selected according to some related work [7-9], to allow the understanding of how they can help in the users' behavior research in social networks. Through them, users' connections and the frequency of tweets sent by one or a group of users can be easily analyzed. Tag cloud visualization, most popular users and ways to classify the tweets are also provided, enabling the learning about the social network and the behavior of its members, through a constant collection of generated data.

Figure 1 shows DeepTwitter's main window, which displays the graph of the **user's social network**, expanded as he/she selects members of the network to include their friends or followers to the network. Two groups of users can be also observed in Figure 1, as well as a window that quickly shows some information regarding the user on which the mouse is over, such as his/her last tweet, description, number of friends and followers, etc. This visualization plays an important role, given that it has the ability to display, at once, a large number of members and connections, a task not possible to be done through the website. The discovery of other members that the user might have interest in also occurs in a very efficient way, through the verification of which users related to him are already following these members, for example.

The prototype also enables the **creation of categories**, into which different tags can be inserted. Through these tags tweets could be classified in both, timeline and tag cloud visualizations. Thus, it is possible to differentiate tweets by subject, clustering all the tweets that contain a given set of tags inside the same category that holds these tags.

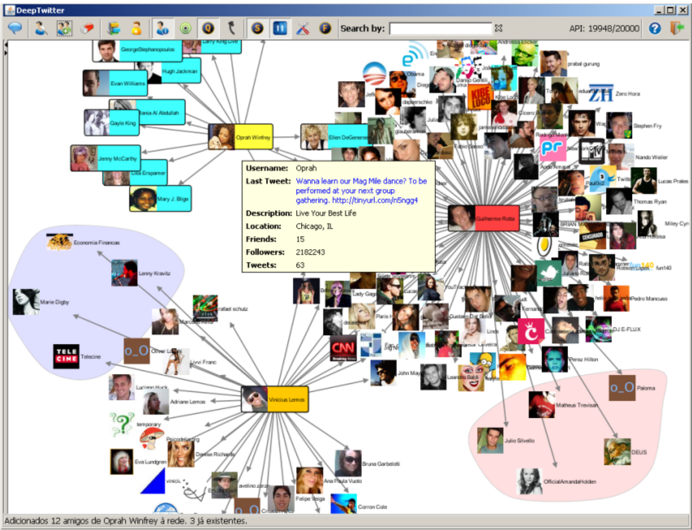


Fig. 1. The user social network with two groups of users and a window containing information about the selected user

When visualizing tweets from individuals or groups in DeepTwitter's main window it is possible to open the **timeline** visualization. The timeline organizes the tweets on a chart, based on the day (x axis) and time (y axis) in which they were sent, allowing the observation of a user or group of users behavior, such as the frequency in which the tweets are sent and what times and subjects are predominantly devoted to it. For timeline visualization, users can choose: the option to display the users' pictures or ellipses for each tweet; and the option to classify or not tweets within the categories already defined. By doing this, the pictures are outlined or the ellipses are filled with different colors, in order to facilitate the identification of which tweets belong to each category. Moreover, the tags are highlighted when the mouse moves across the tweet, as shown in Figure 2.

DeepTwitter also provides a **tag cloud** visualization, which shows the tags with different font sizes to allow the user to quickly and easily find information [10]. The tags could be the ones specified by the user when creating categories, or the ones that are currently the Trending Topics at the Twitter homepage.

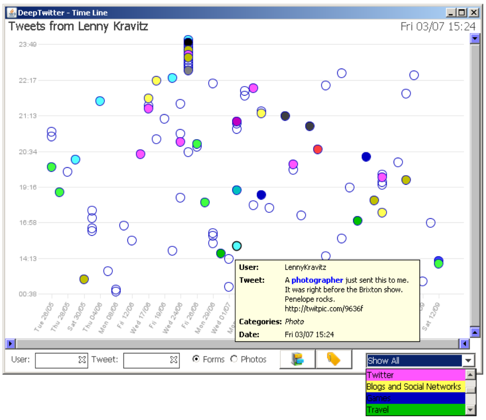


Fig. 2. Example of a Timeline with tweets classification

Finally, the **most popular user** visualization can be applied to a group of selected members or to all users in the contact network. The classification can be made by the number of friends, followers, favorite tweets, or tweets that have already been written. When one of these options is selected through a button interface, the picture size is appropriately changed in relation to the selected option and to other users. According to the selected option, the user that has more friends, e.g., appears with the biggest picture, and the one that has fewer friends, appears with the smallest picture. Through this visualization, with just a quick look the user can identify the most active or important members from his network of contacts, which would be very difficult to do using the Twitter website.

3 Getting User Impressions about the Visualization Techniques

In order to verify users' impressions about the use of visualization techniques, and to know which public would be more interested in this kind of tool, an user study was conducted, in two phases: the first one focusing on the general public, characterized by people who use Twitter for personal use (without professional goals), and the second focused on professionals who use Twitter as support for their work or even as their main activity (as professionals who monitor the impact of their companies on social networks and the Internet, for example).

In both studies, aiming to allow the users to understand the visualization techniques and their possibilities of use in a real environment, they were invited to use the prototype in their day-to-day activities. To better understand and analyze this use, a qualitative approach [11], with a restricted number of users was chosen: first, the users were invited to use the prototype and, then, to answer a set of questions related to this use.

3.1 First Phase: General Public

The methodology used in the first phase and its results are presented as follows.

Used Methodology. In this phase, typical Twitter's users from the general public were invited to participate in the study. Following the example of the evaluations of the related work [2-4], which were performed respectively with 10, 12 and 12 users, we invited 11 users to use DeepTwitter and then answer an online questionnaire about that use. The online questionnaire contained 21 open and closed questions about the user profile and his experience with Twitter and DeepTwitter.

By agreeing to participate, the users received a copy of the prototype and were free to use any visualization technique available from the tool for about a week and at any time they so wished. It is worthwhile mentioning that this use was made in an individual way and it was not accompanied by any observer. After this use they were instructed to answer the online questionnaire. The questionnaire contained open and closed questions, about the user profile, the Twitter and the prototype.

Results. Most users that answered the questionnaire were students (82%) and from the Social Communication area (55% of the students).

Regarding the use of Twitter, we found that most users (91%) had used it for two years or more and 64% used it every day. When questioned about why they used this social network, the main choices were (the users could choose more than one option): personal purposes (64%), work (36%) and information (36%).

Regarding the use of DeepTwitter, 64% used it for one week, mostly using the prototype less than once a day (64%). In relation to the visualization tools available in it, the Social Networking Visualization was the most popular tool, chosen by 64% of the users when asked about the main advantage of using DeepTwitter and by 73% when asked about the visualization tool they considered the most interesting. The possibility to classify the tweets was also pointed out by 27% of users.

The preference for these two visualization tools – social networking and tweets classification – was also highlighted in the users comments: "*This visualization facilitates the verification of users who I follow that also follow me, which is very interesting*", "*The social networking is cool because you can check nearby connection; the tweets per group also facilitate some 'conversations'*", and "*tweets classification per tag is a good function for organizing content*".

Beyond the scope of personal use, one possible application was highlighted by two distinct users: the use by companies. One of the users pointed out that for a company this Social Networking visualization tool would be very useful for the connection with potential new followers: "*Especially for a business twitter, this tool helps you connect to new followers that have to do with the company subject. For example, through the social networking visualization I can see who are the followers in common among my followers, it helps me follow relevant people in my twitter*". Another user proposes the use of tags classification for this purpose: "*for a company, for example, launching a promotional tag*".

3.2 Second Phase: Professionals

The methodology used in the second phase and its results are presented as follows.

Used Methodology. In line with the results of the first phase that pointed to a professional use of this kind of tool, we developed a second phase with professionals who use Twitter as a working tool, inviting them to freely use the prototype in their day-to-day work, and after this use, participate in an interview to report their impressions.

As occurred in the previous phase, the participants received a copy of the prototype to use, by themselves and without the help of an observer, for about a week. The three groups of participants in this phase were professional users, as will be described in the next section. After using the prototype, they participated in a semi-structured interview, conducted by members of the research team. The interview began with 14 questions about the user profile and his experience with Twitter and DeepTwitter.

Results. The participants interviewed represent three distinct profiles: (1) employees of a private company of web hosting that were responsible for the company's public relations (group composed by 3 participants, that answered the questions in group and used the tool with the company account; they will be identified here as User 1); (2) the team responsible for the Twitter account of the Faculty of Social Communication of a large private university in the country (group composed by about 4-6 participants, students of the University, who also responded in group and used the institutional account, so they will be identified here as User 2); and (3) a University professor of Journalism (User 3), who used the program with his Twitter account. The academic profiles of users covered the areas of Public Relations, Journalism and Administration, and their professional profiles are about Social Communication and Customer Relations.

Focusing specifically on the last part of the interview, the users connected to the area of relationships (Users 1 and 2), when asked about the main feature of the prototype, highlighted the Visualization of their followers. User 1 highlights that with the timelines they can "*identify over time followers of our activities in order to be aligned with our audience*". User 3, working with Journalism, believes that the main feature is "*the ability to follow the conversations of specific groups of people*". These points were confirmed when all users of this phase highlighted that the visualization tool they would use more often would be the timeline analysis. User 3 said: "*For journalism or marketing, for example, the analysis of the timeline is a very useful feature*".

3.3 Discussion

Some of the issues presented in the results will be discussed in accordance with the visualization techniques available.

Social Network Visualization. From our studies, we found that for users of the general public the Social Network Visualization was considered the most attractive feature. On the other hand, users from companies (professionals), who often focus on their followers' behavior, highlighted that this feature should be removed from the prototype because the visualization gets very confusing when showing a great number of followers. That is, to the general public this feature allows a very interesting view of their entire network of contacts (and the relationships between their followers and who they follow), while for users interested mainly in their followers' activity, with hundreds (and even thousands) of them, this way of viewing the network seemed useless, because the amount of users shown did not allow a clear view of these connections. In this case, alternative techniques for the visual analysis of large graphs [12] could be presented, but that was out of the scope of this research.

Timeline and Tweets Classification. To users concerned with the relationship area, the timelines allow them to monitor the activities of their followers and thus find a better way to interact with them, knowing the best time of the day to send them an important notification or greeting message, for instance. From a journalistic point of view, the possibility of viewing the timelines from user groups or the classification of tweets permits focusing on specific issues (or conversations), enabling the analysis of new perspectives (how a particular post was discussed, how a particular group reacted to certain news, and so on), thus generating new information ("news").

Tag Cloud and Most Popular Users. Since Tag Cloud and Most Popular Users are visualizations techniques very common nowadays, these were the least mentioned in the evaluation. Only one user highlighted the possibility of using the tag cloud for mapping the subjects most in vogue at the moment. Another one mentioned that the most popular users could be used in order to evaluate the influence of their followers.

4 Conclusion and Future Works

The increase in use of social networks has been intense in recent years, as intense as the activity of users on these networks. Understanding the behavior of users on these networks, e.g. to companies' public relations identify a better way to get in touch with them, can be a hard task, given their high level of interactivity. In this context, we discussed how information visualization techniques could be helpful in supporting this task, through a set of standard visualization techniques and the discussion of these visualizations with its potential users.

From the analyses carried out we can clearly see that different profiles of users are interested in different features (and in different visualization ways), as noted with the preference of users in general (general public) for the social network visualization, and professional users for the timelines. In addition, a comment from one of the users (professional profile) that believes that everything the Twitter already offers should be excluded from the prototype led us to reflect.

Adding to this reflection the trend of users to simultaneously use several small applications to monitor the use of social networks, we now are designing distinct small visualization applications, focusing on specific market niches. Our first application is aimed specifically at professionals working in the area of relationships and their needs, exploring features as visualization of feelings and statistics about mentions.

Focusing on the information visualization area, the interest of users in the visual exploration of the available information opens the opportunity to research and develop novel interactive visualization techniques, focused specifically on social networks, and, also, in multi touch devices that are widely used nowadays and could provide new interactive approaches to visualization.

Acknowledgments. This work was partially supported by PUCRS (Edital 07/2012 – Programa de Apoio a Integração Entre Áreas/PRAIAS).

References

1. Boyd, D.M., Ellison, N.B.: Social Network Sites: definition, history, and scholarship. J. Computer-Mediated Communication 13(1), 11 (2007)
2. Barata, G., Nicolau, H., Gonçalves, D.: AppInsight: What have I been doing? In: International Working Conference on Advanced Visual Interfaces, pp. 465–472 (2012)
3. Leung, C.K.-S., Carmichael, C.L.: Exploring Social Networks: a frequent pattern visualization approach. In: 2nd International Conference on Social Computing, pp. 419–424 (2010)
4. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C.: TwitInfo: aggregating and visualizing microblogs for event exploration. In: Conference on Human Factors in Computing Systems, pp. 227–236 (2011)
5. Kaye, J., Lillie, A., Jagdish, D., Walkup, J., Parada, R., Mori, K.: Nokia Internet Pulse: a long term deployment and iteration of a Twitter visualization. In: Conference on Human Factors in Computing Systems, pp. 829–844 (2012)
6. Rotta, G., Lemos, V., Lammel, F., Manssour, I., Silveira, M., Pase, A.: Visualization Techniques for the Analysis of Twitter Users' Behavior. In: 7th International AAAI Conference on Weblogs and Social Media (2013)
7. Heer, J., Boyd, D.: Vizster: Visualizing online social networks. In: IEEE Symposium on Information Visualization, pp. 32–39 (2005)
8. Perer, A., Shneiderman, B.: Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 265–274 (2008)
9. Kwak, H., Hong, Y., You, J., Moon, S.: Visualizing a Personal Timeline By Adding Multiple Social Contexts. In: International AAAI Conference on Weblogs and Social Media, pp. 14–17 (2012)
10. Halvey, M.J., Keane, M.T.: An assessment of tag presentation techniques. In: 16th International Conference on World Wide Web, pp. 1313–1314 (2007)
11. Denzin, N.K., Lincoln, Y.S.: Collecting and Interpreting Qualitative Materials. Sage Publications, Inc., London (2007)
12. von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., Wijk, J.J., van, F.J.D., Fellner, D.W.: Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. Computer Graphics Forum 30(6), 1719–1749 (2011)

Helping Me Helping You: Designing to Influence Health Behaviour through Social Connections

Noreen Kamal[1], Sidney Fels[1], Joanna McGrenere[2], and Kara Nance[3]

[1] Electrical & Computer Engineering, University of British Columbia, Vancouver, BC, Canada
[2] Department of Computer Science, University of British Columbia, Vancouver, BC, Canada
[3] How To Be Healthy MD, Rolling Meadows, IL, USA
{noreenk,ssfels}@ece.ubc.ca, joanna@cs.ubc.ca,
karajnance@gmail.com

Abstract. Of key importance to avoiding significant health problems such as cardiac disease and stroke is eating nutritious foods and leading an active lifestyle. However, leading a healthy lifestyle remains elusive and obesity continues to increase in North America. We investigate how online social networks (OSN) can change health behaviour by blending theories from health behaviour and participation in OSNs. We designed, prototyped and evaluated an OSN, called *VivoSpace*, using factors derived from the blend of theoretical models with the intention to change health behaviour. Our results from field studies with 35 participants from two clinical and one non-clinical social network groups revealed that designing these factors into the online social network lead to positive health behaviour change, namely, improved diet for the clinical group. The use of theoretical models in the evaluation provided a means to understand how the design can be improved to see further health behaviour change.

Keywords: Health, behaviour change, social networks, diet, physical activity.

1 Introduction

The importance of a nutritious diet and an active lifestyle has been found to be central to a healthy population, which leads to lower healthcare costs for a nation and a better quality of life for its citizens. Specifically, a healthy diet and physical activity can reduce advancement of (and prevent occurrence of) chronic diseases [2], prevent vascular diseases such as cardiac disease and stroke [22,7,15], and result in overall better health [31]. Despite greater medical and scientific knowledge, North Americans continue to be more obese due to poor diet and a sedentary lifestyle. We address this problem through the design and evaluation of an online social network (OSN), grounded in theoretical models, to improve nutrition and increase physical activity.

In addition to diet and physical activity, a significant factor in health outcomes is one's close and distant social networks [8,23]. Social supports can promote better health by providing a sense of belonging and building self-efficacy [6]. Social technology has been shown to increase social support for individuals afflicted with a

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 708–725, 2013.

particular condition, such as ACL (Anterior Cruciate Ligament) [25], ALS (Amyotrophic Lateral Sclerosis) [13] and menopause [36]. Furthermore, the social dynamics for peer involvement in weight management include supportive relations and passive involvement (such as social norms), and observed patterns of interaction such as (un)disclosure [24]. Studies have realized the benefits of social connection by designing technologies for health and weight management, which we review in detail in the *Related Work* section.

Recently, social connection through websites and other web enabled technology have had an explosion of engagement and use, namely in use of OSN services such as Facebook®. We leverage this engagement as a means to build social support in the design of our OSN for health behaviour change. We also take heed of studies that have shown we need to be cautious in using large OSN services such as Facebook® for our health needs, as issues of sharing private health information on such a large social network service does not make sense [29,33]. We build on these works by designing an OSN for small social network groups rather than using existing OSN such as Facebook®. This is done to allow users to share their nutrition and physical activity data with individuals from a known social network group rather than a large social network that would broadcast their data to a diverse range of individuals known from various sources. A small social network group allows users to have control over who sees the information that is disclosed.

The main contribution of this work is to apply two theories to the design, prototyping and evaluation of an OSN, *VivoSpace*, for health behaviour change. We combine the *Social Cognitive Theory* from health behaviour that underpins the importance of social aspects of healthy living with the *Uses and Gratifications Theory* that provides motives for participation in OSNs. The justification for this approach is that our OSN needs to be engaging in order for it to be used, so we look to the *Uses and Gratification Theory* to provide the factors that motivate use of OSNs [17,19]. Similarly, the factors for health behaviour change from the *Social Cognitive Theory* [4] are also applied to the design with an aim to change health behaviour. This theoretical approach is undertaken to create better design principles that help people to both start and maintain healthy habits. Others have used health behaviour theories to design technologies for motivating increased physical activity, improved nutrition, or both; however, the factors from the *Social Cognitive Theory* have not been directly used to design and evaluate technology to promote positive health behaviour. Furthermore, the *Social Cognitive Theory* has never been combined with the *Uses and Gratification Theory* to ensure engagement in the technology. Finally, OSNs have not been explicitly designed for the purpose of health behaviour change before. This work informs research in both HCI and Persuasive Technologies [12], and guides the design of the multitude of applications that are coming to market for fitness and improved diet.

In our design of an OSN, *VivoSpace*, we chose to focus on healthy eating and physical activity as both are closely linked to health outcomes such as ideal weight, prevention of illness and management of chronic diseases [7,2,15]. We are *not* focusing on *weight loss* specifically, but rather nutritious eating habits and physical activity over sedentary lifestyle. We deployed *VivoSpace* to three separate social network groups in both Canada and the United States in a 4-week study with 35 participants.

We use the factors from the theories to create design elements that ensure engagement in *VivoSpace* and changes in health behaviour. In providing this contribution, the main outcomes of this study are: 1) an understanding of different social dynamics in small social network group through use of an OSN for healthy; and 2) details of design mechanics in an OSN that influence health behaviour and motivation to continue to use the website.

2 Foundational Theories

Theoretical models assist us to understand human behaviour. The theories used in our approach to design, prototype and evaluate *VivoSpace* are the *Social Cognitive Theory* and the *Uses and Gratification Theory.* Various HCI researchers have considered the *Social Cognitive Theory* in designing social technologies to influence health behaviour change (as described in the *Related Work* Section). We build upon previous works by utilizing the *Social Cognitive Theory* as a foundation for our design and evaluation. We did not directly use other health behaviour theories in our approach, as we wanted to provide clarity in linking factors for health behaviour change from the theory to the design. Further, as *engagement* in the technology is necessary for the system to be used, we apply the *Uses and Gratifications Theory* to derive design strategies for motivating the use of the OSN. In other words, the OSN needs to be used in order for the design elements for health behaviour change from the *Social Cognitive Theory* to be effective.

2.1 Social Cognitive Theory

The *Social Cognitive Theory* holds that behaviour is determined through expectancies and incentives, and of key importance is self-efficacy to changing health behaviour [3,4]. The *Social Cognitive Theory* [4] was first drawn from the *Social Learning Theory* by [3], and it was further developed into five core determinants of health behaviour change:

1. *Knowledge* is an individual's knowledge of and *expectations of outcomes* related to the health risks and benefits of different health practices.
2. *Perceived Self-Efficacy* is an individual's own competence to perform the behaviour needed to influence outcomes and the individual's ability to exercise control over one's health habits. *Perceived Self-Efficacy* can be influenced vicariously through others, which allows for social modeling and the building of social norms.
3. *Goals* include proximal and distal goals, and they need to be concrete plans and include strategies to realize them. This factor builds accountability.
4. *Perceived Facilitators* include social and structural factors, and social approvals and social supports can be facilitators. They incorporate the value of the outcome of health behaviour change, which can be health, appearance, social approval, or economic gain.
5. *Perceived Impediments* is the opposite of perceived facilitators, and they also include both social and structural factors.

2.2 Uses and Gratification Theory

The *Uses and Gratifications Theory* originated from communications research and is generally recognized to be a sub-tradition of media effects research [28]. Initially, it provided a theoretical approach in defining the motivations for using traditional mass media such as newspaper, radio and television [35]. It has received renewed interest because of its applicability to telecommunications, computer-mediated-communication [34,11] and the Internet [30]. It has also been applied to online communities [11,19], and more recently, uses and gratifications has been studied to understand the motivations for using Facebook® [17].

We consider five key works based that present motivational factors based on the Uses and Gratification Theory on their application to the Internet, online communities, and OSNs [11,19,30,35,17]. By reviewing these works, we find that there are eight main motivational factors that influence use of OSNs based on the Uses and Gratification Theory:

1. *Entertainment* refers to a motivation that stems from enjoyment of playing or interacting with others [30,11,19], and also encompasses *Diversion* [35] and the need to *Pass Time* [30].
2. *Social Enhancement* is a value that one ties to their status within social peers [11,17], and is also linked to *Social Utility* [35] and *Interpersonal Utility* [30].
3. *Maintaining Interpersonal Connectivity* pertains to sustaining contact with one's social networks [11,19], and it also relates to *Social Utility* [35] and *Interpersonal Utility* [30], or as Joinson describes as *Social Connection* [17].
4. *Self-Discovery* occurs when self-knowledge is obtained through social interaction [11,19], and it includes *Personal Identity*, which is a reinforcement of ones values, attitudes and beliefs [35].
5. *Get Information* is an instrumental purpose like receiving information [11,19], and it can also overlap with *Information Seeking* [30] and *Surveillance* [35]. It also suggests social sharing, which includes factors such as *Content Creation and Curation*, *Social Network Surfing*, and *Social Investigation* [17].
6. *Provide Information* is democratized with online communities and OSNs compared to previous media, and refers to a motivational factor that is unique from *Get Information* [19]. In OSNs, we see *Status Updating* [17] as a specific means of *Providing Information*.
7. *Convenience* provides an understanding of why people would choose to use the Internet as opposed to read the newspaper, or socialize via an OSN service as opposed to meet in person. This factor was only described by [30].
8. *Shared Identity* occurs as one constructs their identities through the use of an OSN and relates to others through shared interests and values [17].

3 Related Work

Although OSNs have not been studied specifically for health behaviour change, technologies with social components designed for health behaviour change have been explored by HCI and Persuasive Technology communities. We build on this work by specifically looking at the design of OSNs through the use of theoretical models. Some of the works that have incorporated social aspects in their design have considered the *Social Cognitive Theory*. We also review studies that investigated social aspects of health behaviour without consideration of the *Social Cognitive Theory.* Finally, we present some works that did not investigate social aspects for health behavour change as a central component of their study, but their results informed our work. A description of all these studies is presented in **Table 1**. There are no works that combined the *Social Cognitive Theory* with the *Uses and Gratification Theory* to design of an OSN for health behaviour change.

Table 1. List of related works showing the name of the application developed, description of the application, and the number of participants (p) and duration of the evaluation

Application Name	Description	Evaluation
Stepping Up for Health (SUH) [34]	Internet mediated walking program that measured the impact of adding an online community on the walking habits of individuals.	324 p ,16 wks
VERA [5]	Users take pictures at moments when they make health decisions and provide annotation about the picture's relation to health.	45 p, 2 wks; 44 p, 4 wks
Shakra [1]	Mobile phone application infers whether the user is still, walking or traveling in a car, and provides and shares the amount of time that the user was active with some competition features.	9 p, 1 wk
Fish'n'Steps [21]	Links users' step counts to the growth and happiness of a virtual fish.	19 p, 14 wks
Houston [9]	Pedometer and mobile phone application that allows users to set goals. It includes a sharing version and a personal version.	13 p, 3 wks
Lifestyle Coaching Application [14]	Smart phone and website version where users diary their food and physical activity, set goals and provide a points system for a game experience.	40 p, 4 wks
EatWell [16]	Social mobile phone application designed to assist African American communities make better food choices, where users share audio recordings of experiences related to food.	12 p, 4 wks
Chick-Clique [37]	Mobile application provides sharing of step counts, visibility of group averages, and text messaging capability.	7 girls, 4 days
MAHI [27]	Mobile application to communicate with one's diabetic educator.	49 p, 6 mo.
UbiFit Garden [10]	Wallpaper of the users cell phone shows a garden. Flowers appear with more physical activity, and it shows butterflies as goals are met.	12 p, 3 wks; 28 p, 3 mo
IMPACT [20]	A step-count personal informatics system that is evaluated to understand the influence of context on step count.	49 p, 8 wks
Fit4Life [33]	A mock design of a health system that collects dietary and activity level, shares on OSN sites, and provides advice.	N/A

Since the *Social Cognitive Theory* postulates that health is a social matter and not just an individual one, we look at three works that have considered this theory in the design of their technologies. First, the *SUH* [34] study considered *social support* and *social modeling* from the *Social Cognitive Theory* to explain how the online community experimental group can promote increased step count and lower attrition. The study found that individuals were more engaged and participant attrition was lower with those that used the online community version. However, the *Social Cognitive Theory* was not central in the design and evaluation of *SUH*. Second, the *VERA* system [5] showed that the social aspects allowed participants to mimic others, and that the system built group identity and allowed for greater accountability. Although the authors did mention that the design of *VERA* was consistent with *social modeling* and *accountability* aspects from the *Social Cognitive Theory*, the use of the theory was not central in the design and evaluation. Third, the *Shakra* [1] system loosely based the social aspects of their design on the *Social Cognitive Theory* to show that community was important to health behaviour. *Shakra* did not show any increase in activity, but the participants enjoyed the sharing and competition aspects. We expand the use of the *Social Cognitive Theory* by anchoring the design and evaluation of our OSN on factors from the *Social Cognitive Theory* as well as the *Uses and Gratification Theory*.

Other HCI researchers have done significant work in considering social aspects when designing health technologies without the use of the *Social Cognitive Theory*. The results from these studies are mixed; however, the social aspects did lead to building community empowerment and social supports. The *Fish'n'Steps* [21] used the *Transtheoretical Model* to evaluate the stage or level of health behaviour change of the study participants, but did not use the theory in the design of the system. The *Houston* [9] system showed those that used the sharing version met their goals more often than those that used the personal version. The *Lifestyle Coaching Application* [14] revealed no significant difference between mobile and web versions, or individual and team player groups. The qualitative analysis for *EatWell* [16] revealed that the use of the system led to a sense of empowerment for the community, as users were from the same geographic location. *Chick-Clique* [37] showed an increased step count for one group, but not the other, due to confounding reasons. Finally, the *MAHI* system allowed for social support for newly diagnosed diabetics [27], and a means to construct one's identity for those that have been living with diabetes for more than one year [26]. The variable results can be better understood with the use of theoretical models in the design and evaluation of these systems, which can provide insight into how these systems were successful in achieving health behaviour change and how they can be improved.

The HCI community produced other works related to health behaviour change that inform our study, where social aspects did play a role in their work although it was not a central part of their study. First, the *UbiFit Garden* [10] took a theoretical approach to develop design strategies for health behaviour change; however, different theoretical models were used and the *UbiFit Garden* was not an OSN, so our work extends this approach to the design of an OSN to change health behaviour. Secondly, the *IMPACT* system was created to better understand the influence of context on personal health informatics [20], where social influence is a critical part of context. Finally, another work that deserves mention is a mock parody design called *Fit4Life*, which shows the importance of mindfulness in understanding our behaviour, and the privacy concerns of posting our health behaviours on large OSN services such as Facebook® [33].

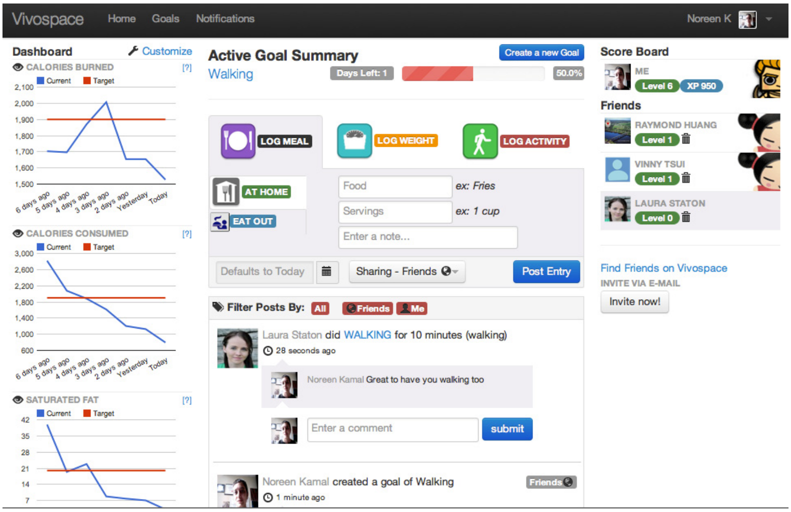


Fig. 1. Main home page of VivoSpace showing the Dashboard on the left; the goals summary, log entry and newsfeed in the middle; and friends with game progress on the right, each level has a character associated with it

4 Designing and Prototyping VivoSpace

By blending the five key factors that influence health behaviour change from the *Social Cognitive Theory* and the eight factors that motivate the use of OSNs from the *Uses and Gratifications Theory*, we designed *VivoSpace*. The *VivoSpace* system provides features that promote eating healthy foods and engaging in physical activity because both are important to good health behaviour. The design of *VivoSpace* was developed through a *User-Centered Design* process, initially with paper prototypes that were evaluated using one-on-one interviews with 11 people, and then with a medium-fidelity prototype that was evaluated in a laboratory with 36 people [18]. These studies revealed that the *VivoSpace* design allowed for greater understanding of one's health behaviour, but social gaming features were needed to build motivation to use the system. Finally, this high-fidelity prototype was tested with seven individuals using a cognitive walk-through prior to the field study described in this paper. The high fidelity prototype was developed using PHP and MySQL within the CodeIgniter application framework. Twitter's® Bootstrap was used for the user interface elements and Doctrine PHP libraries were used for the database abstraction layer. Furthermore, *VivoSpace* incorporates Wolfram Alpha's® Application Programming Interface (API) to obtain nutritional information for meals and caloric expenditure for physical activities logged. When the API cannot return the nutritional information, a site administrator enters the correct information based on web searches, which we anticipate will eventually be crowd-sourced.

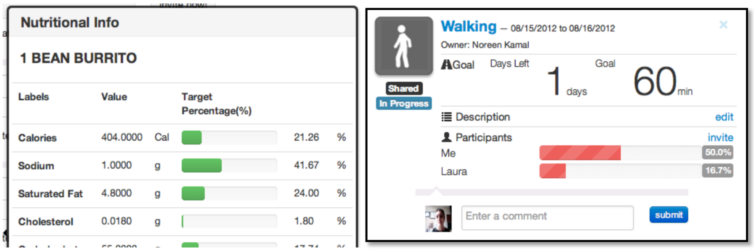


Fig. 2. Left: the nutritional content of a meal is displayed when the logged meal is clicked. Right: goal details show the user's and participating friend's progress towards the goal target.

VivoSpace has three main components: logging, goals, and a personal dashboard. The main home page of *VivoSpace* is shown in **Fig. 1**. Users have the ability to log their meals, physical activity and weight. After users log their meals, the nutritional content of the meals is available when they click on the name of the food (**Fig. 2.** left). Similarly, when they log a physical activity, the caloric expenditure is available. Users have the ability to make a log private or to share it with their social network. The newsfeed shows the log entries for themselves and their friends (we will use the word "friend" to refer to social connections on *VivoSpace).* The shared log entries allow for commenting between the individuals in the user's social network. The historical trends for calories consumed, calories burned, and each nutrient is charted on the dashboard, which is always visible on the left side of the home page. User's can customize their targets for each of the nutrients or use the defaults, which are based on US and Canadian national standards. Users also have the ability to create concrete proximal goals from a library of goals such as walk for 60 minutes over 2 days, or eat 35 g of fiber over 3 days. They can invite friends to participate in their goals with them. The goal progress is shown on the home page, and details are shown when the user clicks on the goal (**Fig. 2.** right). *VivoSpace* also provides the following gamification features: users earn small experience points (XP) by using *VivoSpace* and major experience points for successfully completing goals. Users progress through 10 levels with each level requiring more points to advance to the next level. Furthermore, each level reveals a new character. These design elements are developed based on the factors from the *Social Cognitive Theory* and the *Uses and Gratification Theory,* which are mapped on **Table 2**.

5 The Study

The *VivoSpace* OSN provides design mechanics that should promote motivation to use the system, as well as promote health behaviour change by applying factors from the *Uses and Gratifications Theory* and the *Social Cognitive Theory* (respectively).

We evaluated *VivoSpace* in a field study with three separate social networks in both Canada and the United States. This study's contribution is the application of theories to design and evaluate an OSN for health behaviour change. Since we look at three distinct social networks, we provide a discussion on how different social dynamics within a social network contribute to healthy behaviour through the use of *VivoSpace*. We provide details for which design elements contribute to positive health behaviour change within *VivoSpace*, and how the design can be modified to further contribute to health behaviour change.

Table 2. Mapping of the five factors from the *Social Cognitive Theory* and eight factors from the *Uses and Gratification Theory* to the design elements provided on *VivoSpace*

	Factor	*VivoSpace Design Mechanic*
Social Cognitive Theory	Knowledge	Knowledge is obtained through the nutritional value for logged meals and caloric expenditure for logged physical activity. Information about the definition of each nutrient is available through a help link beside the name of the nutrient.
	Perceived Self-Efficacy	Seeing historical log entries and charts of nutrients consumed on the dashboard shows one's capability to eat healthy and be physically active. Also seeing the log entries of others builds self-efficacy vicariously through others by social modeling.
	Goals	Proximal goals can be created and they are always visible at the top of the homepage. Distal goals are achieved through the charts (on dashboard) displaying historical trends and targets.
	Perceived Facilitators	Visibility of others' activities allows users to mimic behaviour, and the commenting on shared entries provides dialogue for encouragement. The gaming features provide motivation.
	Perceived Impediments	The commenting feature on shared log entries can provide dialogue with one's friends to overcome barriers.
Uses and Gratification Theory	Entertainment	The gaming features include the accumulation of experience points and advancement through 10 levels.
	Social Enhancement	Social enhancement is provided by: 1) visibility into the level obtained by one's friends; 2) visibility into the shared meals of one's friends; and 3) participating with a friend on a goal and seeing their progress towards the target.
	Maintaining Interpersonal Connectivity	Social connection is provided through the ability to comment on friends' log entries and participate with friends on goals.
	Self-Discovery	Ability to see the nutritional value of foods eaten and caloric expenditure for physical activity and the historical trends on the dashboard allow users to discover their health behaviour.
	Get Information	Nutritional information is obtained for each log entry, and historical information is available on the dashboard.
	Provide Information	Information is provided through logging of meals, physical activity, and weight, and through commenting on friend's entries.
	Convenience	Obtaining the nutritional value of foods and caloric expenditure of physical activity, and having a historical storage of one's nutritional information provided convenience of access to information.
	Shared Identities	Through the sharing of one's meals and physical activity, and allowing friends to comment on these entries.

5.1 Methods

Recruitment and Participant Overview. We recruited a total of 35 participants from three different settings including clinical and non-clinical settings. For the clinical groups, we considered patients that are associated with the same clinic, share the same geographic location, and are generally from the same socioeconomic status, as was found to be important by Grimes et al. [16].

The first group of participants from a clinical setting was recruited from a primary care clinic in suburban Chicago, which has a focus on wellness and weight loss. A total of 22 participants were recruited by sending an email through the clinic's patient portal. Half of the participants were randomly assigned to the control group and the other half to the experimental arm. The participants in the experimental group used *VivoSpace* during the 4-week study period, while participants in the control group did not use *VivoSpace*. The reason for having a control group was to control for newly acquired health behaviours from association with the clinic.

The second group of participants from a clinical setting was recruited from an integrative medical center in Vancouver, Canada. This clinic provides a focus on prevention and partnership between physician and patient. Participants were recruited from their medical-group-visit patients, where patients meet with a physician as a group, so they already take advantage of social aspects to good health. A total of 3 participants were recruited from this clinic. The small number reflects the limited number of patients that take part in the medical group visits, and it means that we could not have a control group for this clinic. The participants were recruited through direct email contact from the clinic.

The non-clinical group of participants was a close-knit group of colleagues at a small software gaming company in Vancouver Canada, where all the employees have good knowledge of each other. A total of 10 participants were recruited by directly asking the employees of the company. The company has a total of 35 employees. There was no control group for the non-clinical group, as they were not associated with a wellness or integrative medical clinic, and we used the results from the pre-questionnaire (see *Measures* section below) to evaluation their health without *VivoSpace*.

Table 3. An overview of the participants recruited for the field study showing gender distribution, age, obesity, and their rank of their healthiness (1-6 Likert Scale, 1=very unhealthy)

Group	Male (n)	Female (n)	Mean Age (range)	SD Age	Over-weight (n)	Obese (n)	Mean Health Rank
Chicago Clinic (exp)	2	8	48.15 (24-76)	17.3	2	5	2.9 (SD=1.3)
Chicago Clinic (cntl)	0	9	58.22 (29-68)	11.8	4	3	3.9 (SD=1.0)
Vancouver Clinic	0	3	48.0 (44-57)	N/A*	1	0	3.0 (SD N/A)
Non-Clinical Group	4	6	32.6 (20-46)	7.4	3	0	4.6 (SD=0.9)

*Number of participants is too small (n=3) for standard deviation to be meaningful.

A total of 35 participants were recruited from three groups: Chicago clinic (11 in experimental group and 11 in control group), Vancouver clinic (3), and the non-clinical group from the Vancouver gaming company (10). Out of the 22 participants recruited from the Chicago clinic, 3 dropped out prior to the end of the study. One was in the experimental group, and the other two were in the control group. **Table 3** shows an overview of the remaining participants from all three groups. Overall, there is a bias towards women. Furthermore, the participants from the non-clinical group were younger than those from the clinics, and this group ranked their health better than those from the clinics (health was ranked on a 6-point Likert scale, where 1=very unhealthy and 6=very healthy).

Participants from each group were asked to friend each other, but there was no friending across the social networks, so that we could evaluate the social dynamic within each group. All participants received a $50 honorarium.

Measures. All participants completed a pre-questionnaire prior to the start of the study and post-questionnaire after the 4-week study period. The questionnaires both had the following: 1) current health behaviour; 2) knowledge test of nutrients; and 3) 5-point Likert questions that linked back to the factors from the *Social Cognitive Theory*. The post-questionnaire for those participants that used *VivoSpace* also included the following: 1) 5-point Likert questions that linked back to the factors from the *Uses and Gratification Theory*; and 2) open-ended qualitative responses about their thought on *VivoSpace* and how it contributed to health behaviour change. Change in health behaviour was analyzed through comparison between the pre- and post-questionnaires.

The first set of questions that inquired about health behaviour included: questions about height and weight to obtain a body-mass-index in order to assess if the participant was overweight or obese; the number of salads, vegetables, fruits, French fries and potato chips that they ate in the past week (no other foods were enquired about); and how often they walked, performed cardio exercise, and other physical activity in the past week. The current stage in health behaviour change from the Transtheoretical model was also inquired, as was done by [9,21,23]. The Transtheoretical Model defines the stages of change as pre-contemplation, contemplation, preparation, action, and maintenance [32].

The second set of questions was a 10-question multiple-choice test on the meaning of nutrients to assess knowledge. They included question such as, “What foods have the highest fiber content?” The answers included: a) whole wheat breads, beans, and vegetables; b) any breads and cereal grains; c) whole wheat bread and meats; and d) all of the above.

The third set of questions was 5-point Likert style questions that asked for the level of concurrence on statements that mapped to the factors from the *Social Cognitive Theory*. For example, the statement for *perceived self-efficacy* for eating nutritious foods stated, “I am capable of eating highly nutritious foods and resisting unhealthy foods.” The 5-point Likert questions ranged from “strongly disagree” (1) to “strongly agree” (5).

The post-questionnaires for those participants in the experimental group also included 5-point Likert questions for the eight factors from the *Uses and Gratification*

Theory. The post-questionnaire also inquired about which design features of *VivoSpace* contributed to the factors from the *Social Cognitive Theory*. For example, after the afore mentioned Likert question on *perceived self-efficacy*, another 5-point Likert statements inquires concurrence with, “I was able to learn about my capabilities by logging my meals on VivoSpace”

The open-ended qualitative responses included general thoughts on *VivoSpace* and how it can be improved to motivate use and health behaviour change, and specific responses for how it can provide factors from both theories more effectively.

6 Results

The results cover the following: 1) whether the incorporation of the *Uses and Gratification Theory* into the design encouraged use of *VivoSpace*; 2) whether the incorporation of the *Social Cognitive Theory* led to an increase in the self-report on the five factors from this theory; 3) and if there was any observed change in health behaviour at the end of the 4 weeks. Analysis includes statistical comparison of pre- and post-questions as well as qualitative analysis of open-ended free form questions.

6.1 Overview of System Use

The use of *VivoSpace* over the 4 weeks varied substantially. All 23 participants that were using *VivoSpace* created an account and friended others from their network group. The participants from the non-clinical group (Vancouver gaming company) used it more than those from the Chicago clinic, which can likely be attributed to this group’s ease with technology. One participant from the Vancouver clinic used the system the most. **Table 4** shows an overview for usage of *VivoSpace* for all three groups. The non-clinical group was less inclined to share their entries than those from the clinical groups. The comments included: encouragement about the post, “good job!” and “Woot!”; additional information about the food or activity entered, “Santa Cruz organic added to Perrier”; or their feeling about the activity, “I love Thai food, yum”.

Table 4. Overview for usage of *VivoSpace* for all three groups, includes the mean and range of log entries made, disclosure of log entries, and total comments made for each group

Group	Mean # of log entries	Min-max # of log entries	% private log entries	Total # of comments
Chicago Clinic	18.3 (SD=27.4)	0-83	2.2%	20
Vancouver Clinic	93 (SD=161.1)	0-279	1.1%	19
Non-Clinical Group	54.6 (SD=63.3)	2-184	59.2%	24

The questionnaire also inquired about each participant's perception of use based on the eight factors from the *Uses and Gratifications Theory*. The results of the 5-point Likert responses for all three groups are shown in **Table 5**. The strongest factors for motivating use were to *Provide Information* and *Self-Discovery* for all groups;

however, interestingly *Shared Identity* was strong only for the loose social connections of the clinical groups. Although the Likert response for to *Get Information* was a bit lower with mean of 2.67 for the non-clinical group, the qualitative responses showed that they did use it to receive information; for example, "*I liked watching my vitamin consumption over time*" (P9). The qualitative responses for the Chicago clinic revealed sentiment for the *Shared Identity* motivational factor; such as, "*I liked being supportive to like minded people achieving goals*"(P32). The Vancouver clinic had one active user, who was the heaviest user of *VivoSpace*. She continued to use it on a daily basis despite the fact that the she did not have the full social experience of *VivoSpace*. This user had some mistrust in the nutritional and energy expenditure information that was provided by *VivoSpace*, "*I'm not sure how accurate the nutritional and calorie burning info was, and I think that more detailed information would need to be made available for it to be useful*" (P31). With respect to *Self-Discovery*, she said, "*It was interesting to be accountable; it made me think more about it*" (P31).

Table 5. Mean values for results of 5-point Likert Scale responses for the Uses and Gratification Theroy for all 3 groups

Group	Entertainment	Social Enhancement	Maintain Connectivity	Self Discovery	Get Information	Provide Information	Convenience	Shared Identity
Chicago Clinic	2.14	1.71	2.0	2.42	1.57	2.85	2.17	2.85
Vancouver Clinic	3.5	3.0	2.0	3.5	2.5	3.5	2.5	3.0
Non-Clinical Group	2.44	2.78	2.67	3.11	2.67	3.0	2.33	2.22
Overall Mean	**2.44**	**2.39**	**2.33**	**2.89**	**2.22**	**3.00**	**2.29**	**2.56**

6.2 Change in Key Factors from Social Cognitive Theory

The responses from pre- and post-questionnaires were compared using paired t-tests for both the non-clinical and clinical groups. Paired t-tests were done for the Chicago clinical group and combined clinical experimental group (both Vancouver clinic and Chicago clinic). Due to the small number of participants from this clinic, independent t-test with only the Vancouver clinic could not be done. The results from the Chicago clinical experimental group and combined clinical experimental group were compared with the Chicago control group.

Knowledge. *Knowledge* was measured in two ways: 1) through a 10-question multiple choice test; and 2) through direct inquiry via a 5-point Likert scale question. There was no statistically significant difference between any of the groups: non-clinical group; all clinical groups; Chicago clinic experimental group; and Chicago clinical control group. The participants did not gain any knowledge of nutritional content of food consumed and physical activity. Users felt that there should have been other ways to obtain health information, as in the following qualitative response, "*It would be nice if I could ask* VivoSpace *the nutritional information for a Big Mac or something that is for food that is not eaten*" (P7). Others wished that *VivoSpace* provided more prescriptive advice on how to change behaviour, "*How can I have more vitamin C in my diet? I would like a recommendation on what I should eat*" (P6).

Perceived Self-Efficacy. A central concept to the *Social Cognitive Theory* is *Perceived Self-Efficacy*. The paired t-tests for all groups showed a statistically significant increase between the pre- and post-questionnaires' 5-point Likert scale inquiry into one's *Self-Efficacy* **(Fig. 3).** The greatest increase was observed for the Chicago clinical experimental group and the smallest increase was observed for the non-clinical group. However, these results need to be taken with caution, as the control group from the Chicago clinic also showed an increase in *Perceived Self-Efficacy*, which perhaps means that there were other factors contributing to the increase in self-efficacy such as association with the Chicago clinic.

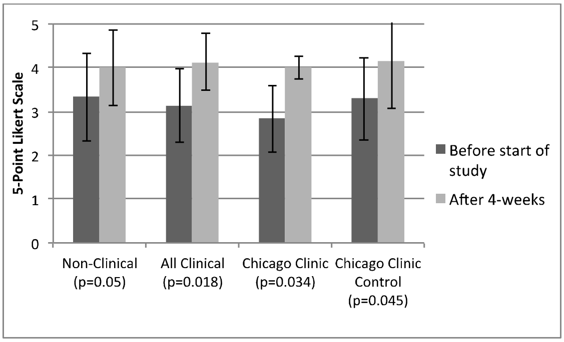


Fig. 3. The pre and post results of self-reported *perceived self-efficacy* for eating nutritious foods using a 5-point Likert scale. Mean values are represented with the error bars showing SD.

For those using *VivoSpace*, we also inquired about the aspects of *VivoSpace* that contributed to their self-efficacy. Generally, the participants felt that logging their meals and physical activity can lead to increasing their understanding of their capability.

Goals. The statistical analysis revealed no significant difference. The qualitative responses showed that users wanted to be able to customize their goals, "*most of the goals currently in the system were not appropriate for me so there was nothing for me to do with other members*" (P7). Making the goals more social and more central to the system also can contribute to more creation and completion of health goals, as one participant says, "*putting* [the goals] *out there for people to see*" (P32).

Perceived Impediments and Facilitators. No statistically significant difference was observed for either decreasing perceived impediments or increasing perceived facilitators. Most users of *VivoSpace* felt that the system's ability to provide their nutritional information does assist them to reduce impediments or increase facilitators. However, as one participant points out, time constraints are the biggest impediment to healthy behaviour, and logging one's information on *VivoSpace* takes time (P28). This perhaps shows the need for the system to have reminders, as suggested by P1. There were some features that some participants found to be facilitators: "*I like having the charts to show my progress; especially seeing my weight line go down*" (P28); and "*nutritional information regarding foods I was eating helped push me to keep cooking those foods or avoiding other foods*" (P9).

6.3 Changes in Health Behaviour

There was health behaviour change observed between the pre- and post-questionnaires for the Chicago clinic experimental group and a strong trend toward health behaviour change for the other two experimental groups. The Chicago clinic experimental group increased the number of salads they ate in a week with a mean difference of 1.0 ($t=2.65$, $p<0.05$). For the non-clinical group, the number of servings of potato chips (and similar foods) consumed in a week went down with a mean difference of -0.56 ($t=1.89$, $p=0.095$). For the clinical group (combination of both clinics), we saw a strong trend toward a positive move in the stages of change from the Transtheoretical Model [32] with a Wilcoxon Signed Ranks test ($Z=1.84$, $p=0.066$). By comparison, there was no health behaviour change observed in the Chicago clinic control group.

7 Discussion

An OSN can change health behaviour if it is designed with features that influence factors from theoretical models. The results for health behaviour change do not show large changes, but the small steps towards healthier living are encouraging because positive health behaviour change did occur in some way for each of the experimental groups, and did not occur for the control group. The maintenance of these new behaviours is not known at this time, but would be important to investigate further. By evaluating *VivoSpace* with the use of the *Social Cognitive Theory* and *Uses and Gratification Theory*, we can better understand which design elements contributed to health behaviour change; furthermore, the results reveal the factors from the theories that were not provided in the design of *VivoSpace*, which provide possibilities for design improvement.

Logging meals and physical activities to reveal the nutritional value of meals and caloric expenditure of activities combined with charts showing trends of calories and nutrients consumed build ***perceived self-efficacy***. The greatest increase in self-efficacy was observed in the Chicago clinical experimental group, which suggests that the process of logging meals and physical activity does build self-awareness. By logging health information, users become more accountable and aware of their capabilities.

The design can be improved by also *providing nutritional information for foods not eaten) and having tailored messages for how to improve health behaviour to build* ***knowledge***. The qualitative results show that the design can be modified to promote an increase in knowledge of nutritional value of food and energy expenditure of physical activities. The participants suggested that *VivoSpace* could assist them to make decisions about food, so it should have the ability to obtain nutritional information for foods that were not eaten. Further, *VivoSpace* can be more prescriptive and provide tailored messages for how the user can make healthier choices.

Providing reminders for ***goals*** *and allowing for customizable* ***goals*** *will increase user engagement.* The goals feature on *VivoSpace* should provide reminders to the users after they are set, and goals should be more central in the design. A key change to the goals feature would be to allow users to create customized goals, as the rigidity of the goals provided on *VivoSpace* led to fewer people creating goals. These changes will increase *Social Enhancement* and make the system more *Entertaining*; both of which are factors for motivating use from the *Uses and Gratification Theory.*

Email notifications should be provided when comments are made to one's log entries, and a link should be provided to the log entry and the comment. VivoSpace should provide notifications when comments are made to their posts, or when one responds to one's existing comment. This will increase dialogue between participants, which will build social supports to allow for an increase in the motivation to *Maintain Connectivity* from the *Uses and Gratification Theory* and *Perceived Facilitators* from the *Social Cognitive Theory.*

Design should include a mobile version and other ***conveniences*** *such as automatic sensing to ease the time required to use it*. The time that was required to enter the data was a contributing factor for some participants to ask for easier means of entering their data. A mobile version and linking to devices such as fitbit® to automatically sense steps taken can ease the burden of data entry.

The differences observed between the clinical and non-clinical groups show the potential to use an OSN such as *VivoSapce* with clinical groups. Overall, both clinical groups disclosed their log entries more than the non-clinical group, and the group from the Chicago clinic saw observable changes in health behaviour. Furthermore, both clinical experimental groups saw the greatest increase in self-reported self-efficacy. The Chicago clinic experimental group showed greater social motivation to use *VivoSpace* compared with the non-clinical group.

8 Conclusions

An OSN called *VivoSpace* was designed and prototyped based on factors derived from the *Uses and Gratifications Theory* for use of the OSN, and the *Social Cognitive Theory* for health behaviour change. Through a 4-week field trial with three social network groups, we used these theories combined with usage data and health behaviour data to evaluate *VivoSpace*. We were able to determine the design elements through the application of these theories that will lead to use of the system and changes in health behaviour. The study also revealed opportunities to improve the design, so more significant health behaviour change can be observed. We found that a close-knit non-clinical social network group interacts with the system differently than a group of strangers from the same clinic. An OSN such as *VivoSpace* can be applied to clinical groups, who are associated with the same primary care clinic to improve health behaviour. OSNs such as *VivoSpace* can provide a health intervention to populations with the greatest need to change health behaviour leading to a healthier population. The use of theoretical models provides an effective means to design and evaluate technologies for health behaviour change.

Acknowledgements. We are thankful for funding from the Canadian Networked Centres of Excellence on Graphics, Animation and New Media (GRAND NCE) and the Institute for Computing, Information and Cognitive Systems (ICICS) at UBC. We thank Raymond Huang and Vincent Tsui for programming VivoSpace. We would also like to thank Michelle Mica Aragon for assisting in contacting participants from the Chicago clinic, and Dr. Lawrence Cheng at Connect Health, Michael Fergusson at Ayogo Games, and Wolfram Alpha ® for their support in the study.

References

1. Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., Brown, B., Muller, H.: Shakra: Tracking and Sharing Daily Activity Levels with Unaugmented Mobile Phones. Mobile Netw. Appl. 12, 185–199 (2007)
2. Anderson, J.J., Polombo, R.D., Earl, R.: Position of the American Dietetic Association: The Role of Nutrition in Health Promotion and Disease Prevention Programs. J. Am. Diet. Assoc. 98, 205–208 (1998)
3. Bandura, A.: Social Learning Theory. Prentice Hall, Englewood Cliff (1977)
4. Bandura, A.: Health Promotion by Social Cognitive Means. Health Educ. Behav. 31, 143–164 (2004)
5. Baumer, E.P., Katz, S.J., Freeman, J.E., Adams, P., Gonzales, A.L., Pollak, J., Retelney, D., Neiderdeppe, J., Olson, C.M., Gay, G.K.: Prescriptive Persuasion and Open-Ended Social Awareness: Expanding the Design Space of Mobile Health. In: Proc. CSCW 2012, pp. 475–484. ACM Press (2012)
6. Berkman, L.F.: The Role of Social Relations in Health Promotion. Psychosomatic Medicine 57, 245–256 (1995)
7. Blair, S.N., Morris, J.N.: Health Hearts and the Universal Benefit of Being Physically Active: Physical Activity and Health. Ann. Epidemiol. 9, 253–256 (2009)
8. Cohen, S.: Social Relationships and Health. Am. Psychol., 676–684 (2004)
9. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design Requirements for Technologies that Encourage Physical Activity. In: Proc. CHI 2006, pp. 457–466. ACM Press (2006)
10. 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. ACM Press (2009)
11. Dholakia, U.M., Bagozzi, R.P., Pearo, L.K.: A Social Influence Model of Consumer Participation in Network- and Small-Group-Based Virtual Communities. Int. J. Res. Mark. 21, 241–263 (2004)
12. Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufman (2003)
13. Frost, J.H., Massagli, M.P.: Social Uses of Personal Health Information Within PatientsLikeMe, and Online Patient Communities: What Can Happen When Patients Have Access to One Another's Data. JMIR 10, e15 (2008)
14. Gasser, R., Brodbeck, D., Degen, M., Luthiger, J., Wyss, R., Reichlin, S.: Persuasiveness of a Mobile Lifestyle Coaching Application Using Social Facilitation. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 27–38. Springer, Heidelberg (2006)
15. Goldstein, L.B., Adams, R., Becker, K., Furberg, C.D., Gorelick, P.B., del Zoppa, G.J.: Primary Prevention of Ischemic Stroke: A Statement for Healthcare Professionals from the Stroke Council of the American Heart Association. Circulation 103, 163–182 (2001)
16. Grimes, A., Bednar, M., Bolter, J.D., Grinter, R.E.: Eatwell: Sharing Nutrition-Related Memories in a Low-Income Community. In: Proc. CSCW 2008, pp. 87–96. ACM Press (2008)
17. Joinson, A.N.: Looking at, Looking Up or Keeping Up with People?: Motives and Use of Facebook. In: Proc. CHI 2008, pp. 1027–1036. ACM Press (2008)
18. Kamal, N., Fels, S.: Determining the Determinants of Health Behaviour Change through an Online Social Network. In: Bang, M., Ragnemalm, E.L. (eds.) PERSUASIVE 2012. LNCS, vol. 7284, pp. 1–12. Springer, Heidelberg (2012)
19. Lampe, C., Wash, R., Velasquez, A., Ozkaya, E.: Motivations to Participate in Online Communities. In: Proc. CHI 2010, pp. 1927–1936. ACM Press (2010)

20. Li, I., Dey, A.K., Forlizzi, J.: Using Context to Reveal Factors that Affect Physical Activity. ACM Trans. Comput.-Hum. Interact. 19, 7:1–7:21 (2012)
21. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B.: Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 261–278. Springer, Heidelberg (2006)
22. Lloyd-Jones, D., Adams, R.J., Brown, T.M., Carnethon, M., Dai, S., Wylie-Rosett, J.: Heart Disease and Stroke Statistics-2010 Update. Circulation 121, e46–e215 (2010)
23. Madan, A., Moturu, S.: t., Lazer, D., and Pentland, A.S.: Social Sensing: Obesity, Unhealthy Eating and Exercise in Face-to-Face Networks. In: Proc. Wireless Health 2010, pp. 104–110 (2010)
24. Maitland, J., Chalmers, M.: Designing for Peer Involvement in Weight Management. In: Proc. CHI 2011, pp. 315–324. ACM Press (2011)
25. Maloney-Krichmar, D., Preece, J.: A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community. ACM Trans. on Comput.-Hum. Interaction 12, 201–232 (2005)
26. Mamykina, L., Miller, A.D., Mynatt, E.D., Greenblatt, D.: Constructing Identities Through Storytelling in Diabetes Management. In: Proc. CHI 2010, pp. 477–486. ACM Press (2010)
27. Mamykina, L., Mynatt, E.D., Davidson, P.R., Greenblatt, D.: MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In: Proc. CHI 2010, pp. 477–486. ACM Press (2008)
28. McQuail, D.: The Rise of Media of Mass Communication. In: McQuail, D. (ed.) Mass Communication Theory: An Introduction, pp. 1–29 (1994)
29. Newman, M.W., Lauterback, D., Munson, S.A., Resnick, P., Morris, M.E.: It's Not That I Don't Have Problems, I'm Just Not Putting Them On Facebook: Challenges and Opportunities in Using Online Social Networks for Health. In: Proc. CSCW 2011, pp. 341–350. ACM Press (2010)
30. Papacharissi, Z., Rubin, A.M.: Predictors of Internet Use. J. Broadcast. Electron. 44, 341–350 (2000)
31. Pate, R.R., Pratt, M., Blair, S.N., Haskell, W.L., Macera, C.A., Wilmore, J.H.: Physical Activity and Public Health: A Recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. J. Am. Med. Assoc. 273, 402–407 (1995)
32. Prochaska, J.O., Velicer, W.F.: The Transtheoretical Model of health Behavior Change. Am. J. Health Promot. 12, 38–48 (1997)
33. Purpura, S., Schwanda, V., Williams, K., Stubler, W., Sengers, P.: Fit4Life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight. In: Proc. CHI 2011, pp. 423–432. ACM Press (2011)
34. Richardson, C.R., Buis, L.R., Janney, A.W., Goodrich, D.E., Sen, A., Hess, M.L., Mehari, K.S., Fortlage, L.A., Resnick, P.J., Strecher, R.J.Z., Piette, J.D.: An Online Community Improves Adherence in an Internet-Mediated Walking Program. Part 1: Results of a Randomized Controll Trial. JMIR 12, e71 (2010)
35. Ruggiero, T.E.: Uses and Gratification Theory in the 21st Century. Mass Commun. Soc. 3, 3–31 (2000)
36. Sillence, E., Briggs, P., Harris, P.R., Fishwick, L.: How do Patients Evaluate and Make Use of Online Health Information? Soc. Sci. Med. 64, 1853–1862 (2007)
37. Toscos, T., Faber, A., An, S., Gandhi, M.P.: Chick Clique: Persuasive Technology to Motivate Teenage Girl to Exercise. In: Ext. Abstracts CHI 2006, pp. 1873–1878. ACM Press (2006)

Who Would Pay for Facebook? Self Esteem as a Predictor of User Behavior, Identity Construction and Valuation of Virtual Possessions

Jiaqi Nie[1,2] and S. Shyam Sundar[2,3]

[1] Keio-NUS CUTE Center, Interactive & Digital Media Institute,
National University of Singapore, 119613, Singapore
nicol018nie@gmail.com
[2] Department of Interaction Science, Sungkyunkwan University, Seoul, Korea
[3] Media Effects Research Laboratory, Penn State University, University Park, PA, USA
sss12@psu.edu

Abstract. Self-presentation is a major preoccupation in Facebook. Users carefully construct their online profile and assiduously edit postings on their wall in order to strategically shape their online persona. This study examines some psychological antecedents and consequences of such actions. In particular, we propose that users' self-esteem affects their sense of agency and self-monitoring tendencies, with the former leading to a fuller description of their profile and the latter contributing to more frequent customization of their wall. In turn, these behaviors are hypothesized to contribute to users' personal and social identity respectively, *en route* to affecting their valuation of Facebook as a virtual possession. Structural equation modeling analysis of survey data (N=221) largely supports this model and reveals that the personal identity reflected in one's Facebook account is a major predictor of the degree to which one values it as a possession. We discuss the implications of "I" vs. "Me" in self-esteem with regard to virtual possessions in social networking environments.

Keywords: self-esteem, Facebook activity, online identity construction, virtual possession, survey, quantitative methods, structural equation modeling (SEM).

1 Introduction

It is natural for humans to view the self as being situated in a social experience. Individuals construct their identity by carefully articulating their personality and paying close attention to how they present themselves to others. With the arrival of networked media, this articulation becomes even more pronounced. Internet users, particularly those who have social media accounts, must now consider their online personas in addition to—or as part of—their real-life identities. It has been well recognized by scholars that the Internet is challenging and shaping one's identity [1–3]. In fact, many studies have identified effects of Internet use on individual identity. Research has focused on the changes in self-concept ushered in by the Internet, giving

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 726–743, 2013.

rise to the notion of "second self" as an online extension of one's psychological and social self [4], leading to a conscious accumulation of "virtual possessions" [5], such as digital images, music files and even the number of contacts on one's social networking site.

All this suggests that while the psychology of one's offline self may predict our online activities related to shaping our online self, these activities in turn would likely dictate the degree to which we define ourselves based on our online self and the extent to which we value our virtual possessions. An understanding of the psychological mechanisms involved in the translation of one's self concept to one's online identity can be quite insightful in enhancing our theoretical understanding of user motivations for specific activities in social media sites. It also has practical implications for the valuation of such sites (e.g., how much users of social networking sites would be willing to pay for the service if it were not free) and the design of identity-enhancing features that would maximize the value of virtual possessions for users.

We begin with a review of the theoretical framework underlying self-concept by focusing on the two facets of "Me" and "I" based on Mead's theory, and propose a series of hypotheses involving psychological concepts as well as user behaviors and attitudes pertaining to Facebook, the popular social networking site [6]. The model emerging from these hypotheses is then tested using structural equation modeling analysis, followed by a discussion of the specific associations discovered in the study and their implications for theory and practice.

2 Literature Review

2.1 Mead's "I" & "Me" in Self-esteem

Several decades before the advent of social media, G.H. Mead [7] developed a systematic theory of the social nature of self by introducing the concepts of the "I" and "Me." According to this theory, individuals construct the "I" and the "Me" based on personal selfhood [6]. From a perspective of the world which can be reinterpreted by our actions and the feedback we receive, there are, on the one side, individuals who care about the "I" in selfhood, trying to know themselves by observing their own behaviors; on the other hand, the "Me" in one's self-concept is focused on what is learned in the interaction with others, and more generally with the environment.

Individuals' specific views about selfhood and the way they frame these views identify and contribute to their global self-esteem [8]. Individuals who see positive self-views from their own behavior are especially likely to be high in self-esteem. On the other hand, individuals who derive self-views from their interactions with others and the environment tend to be low in self-esteem.

Self-esteem lies at the core of one's self-concept and serves as the starting point for our examination of user identity in social media. Insofar as this study is concerned, 'self-esteem' refers to the evaluation of oneself, which is conceived both as a trait and a state of being [9]. We adopt the term 'self-esteem' to depict an individual's selfhood from a broad perspective, as proposed by Campbell et al. [10]. People with high self-esteem are consistent and clear about themselves, with the “I” dominating in one’s persona; whereas people with low self-esteem tend to be confused about their self concept and feel anxious about self-relevant images in public; their character is strongly influenced and controlled by “Me” rather than “I” [11–13].

More broadly, the twin aspects of “I” and “Me,” generated by high and low self-esteem respectively, are likely to lead to two different paths in influencing our online attitudes and behaviors, shaping our online identity, and eventually changing our valuation of our virtual possessions, as discussed in the sections that follow.

2.2 Psychological Factors and Behaviors Based on Self-esteem

There is consensus among scholars that self-esteem shapes the way we perceive ourselves and the strategies we use to present ourselves. On the one hand, self-esteem could be a proxy for the psychological sense of agency. Individuals who score high on self-esteem have a higher sense of self-control and enhanced initiative [11]. Higher self-esteem refers to a more favorable evaluation of oneself, which has profound consequences for one’s “sense of agency” [14]. The term “sense of agency” is used to describe the “sense that I am the one who is causing or generating an action” ([15], p.15). Therefore, we hypothesize:

H1: The higher the level of self-esteem, the higher the degree of sense of agency.

On the other hand, when it is low, self-esteem is related to self-monitoring in public presentations. The need for impression management and self-monitoring is critical when those with low self-esteem express themselves publicly in SNS [16]. Empirically, social media researchers adopt self-esteem, an overall affective psychosocial development measure, in evaluating one’s self or personality, to observe virtual selfhood in social networking sites like Facebook. For example, individuals with lower self-esteem have been shown to experience a higher level of bridging social capital by increasing their intensity of Facebook usage, measured in terms of the customization behaviors on one’s Facebook wall [17]. This demonstrates the “Me” aspect of self-concept, a tendency to monitor oneself constantly and rely on others for boosting one’s self-esteem. Thus, we propose that:

H2: The lower the level of self-esteem, the higher the level of self-monitoring.

According to Mead’s social behaviorism [6], “individuals become aware of and learn to guide their behavior” during their process in defining and studying the idea of the self (as cited in [18], p. 165). In other words, individuals adjust their behaviors and shape their images available to others based on their view of themselves [19]. Facebook enables its users to communicate with a diverse set of affordances (“action possibilities”), such as posting comments, sharing pictures, and liking or highlighting contents [20]. This is specifically true after Facebook released its timeline version, which offers users more self-monitoring options on their Facebook wall. Therefore, we can expect users with low levels of self-esteem to be overly conscious of their wall and proceed to use the available customization options, such as editing certain pictures, removing tags and deleting comments. On the other hand, the profile space in Facebook is a key location for self-presentation that can help boost one’s sense of agency. Online SNS profiles are the epitome of Web-enabled authorship wherein they can exercise enormous agency. The completion of the Facebook profile, insofar as defining one's presence on social networks, indicates how much individuals want to construct one's full identity on Facebook. Previous studies have found that individuals report higher self-esteem when they make changes to their online profiles, because of heightened self-awareness and self-presentation [21].

Based on the preceding review of theoretical and empirical work, we propose the following hypotheses for study in the SNS context:

H3: The higher the sense of agency for individuals, the higher the level of profile completion on Facebook.

H4: The higher the level of self-monitoring, the greater the frequency of wall customization in Facebook.

2.3 Facebook as Personal Identity vs. Social Identity

Facebook affords the ability to portray oneself in the best possible light. By either monitoring the Facebook wall or customizing the presentation on Facebook, users can carefully mold one's online identity. However, no matter how changeable an identity could be in Facebook, the original characteristic in oneself would still be the determining factor in one's online identity. In other words, although users of Facebook tend to create certain forms of online identity for themselves and are allowed more liberty to shape it than in the offline world, they still operate within the constraints of their original offline identity. The relationship between online identity and offline identity is bidirectional and mediated by a host of psychological factors and behaviors.

In general, identity, conceptualized by previous researchers as "an individual's perception of selfhood," contains two basic facets [22, 23]. One focuses on individuals' self-consciousness and self-reflection. The other facet, labeled "social identity," goes one step further by paying attention to personal behaviors within a group or public context [24]. Different from the personal identity, which is based on personal accomplishment and material attainment, social identity becomes a matter of impression management, directed at motivating others to achieve and maintain positive concepts of oneself in public [25, 26].

Empirical studies have confirmed the direct relationships between different types of Facebook user behaviors and their identity. DiMicco and Millen [27] conducted a survey based on the Facebook profiles visible when browsing IBM's network to examine the relationship between a Facebook users' profile behaviors and their online identities. They found that individuals use different profile pages in order to maintain multiple self-identities on Facebook. Ellison et al. [28] showed that people with low self-esteem gained more social capital, which can strengthen their social identity through Facebook if they update their wall more intensely. Basically, what individuals "invest" in "who they are" allows them to reconfirm the personal identity facet inside one's selfhood; while what they "invest" in "who they are communicating with" leads them to awaken the social identity component inside them.

Thus, under this line of reasoning, it is to be expected that identity, perceived by individuals within the context of Facebook, will also contain two dimensions determined by selfhood and self-esteem. Specifically, we propose that the "I"-dominated high self-esteem will connect to personal identity while the "Me"-influenced low self-esteem will relate to social identity. Based on this rationale, we hypothesize,

H5. The higher the level of profile completion on Facebook, the higher the degree of perceived personal identity from Facebook.

H6. The greater the frequency with which individuals customize their Facebook wall, the higher their degree of perceived social identity from Facebook.

2.4 Identity and Virtual Possession

If indeed our online activities can shape our identity, the products of those activities (the conversation threads archived, the digital images tagged, and so on) are likely to become treasured "virtual possessions". Given the increasing amount of time and resources spent on social networking and other online activities, we are entering the realm of "real virtuality," wherein the mediated world is a major aspect of our every-day "reality" [29]. The boundary between material possessions and virtual possessions is now blurring. Odom and his colleagues define virtual possessions as "artifacts that are increasingly becoming intangible (i.e. books, photos, music) and things that have never been traditionally considered as tangible goods, (i.e., SMS archives, social networking profiles)" ([30], p. 1491).

Facebook provides individuals with specific, virtual information wherein users benefit from a sense of oneself. In the present study, we suggest that Facebook acts as a virtual container of online social information (about oneself) that may be considered as "virtual possessions" by its users. Given this, it is necessary to analyze the factors that impact the role of one's own network on Facebook as a virtual possession and how we value this possession.

Any possession is based on attachment with certain specific meanings. As Odom et al. note, a possession should have its value in either fulfilling one's life or helping establish an idealized future goal [30–32]. Researchers argue that individuals form a meaningful attachment toward possessions through the process of self-extension [33]. Therein, lessons learned about possession-attachment develop and form a template (internal working model) on which relationships are built for life, which is relatively close to self-concept and self-identity. By mapping these mechanisms from material possession constructs, the following hypotheses can be posited about the relationship between online identity and virtual possessions.

H7: The higher the degree of personal identity perceived from Facebook by individuals, the higher the value they place on Facebook as a virtual possession.

H8: The higher the degree of social identity perceived from Facebook by individuals, the higher the value they place on Facebook as a virtual possession.

Taken together, a hypothetical structural model is proposed to predict how psychological differences of Facebook users affects their valuation of virtual possessions by examining different types of Facebook user behaviors, leading to different attitude outcomes pertaining to their personal and social identity (See Fig. 1). A person's self-esteem is expected to positively predict their sense of agency, which in turn is associated with more Facebook profile behavior and a higher degree of personal identity perceived in Facebook. On the other hand, self-esteem is expected to negatively predict the level of self-monitoring, which is associated with more Facebook wall customization and a higher degree of social identity perceived in Facebook. Both personal identity and social identity derived from Facebook are said to predict the value that users place on Facebook as a virtual possession.

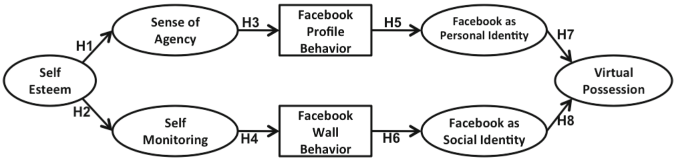


Fig. 1. Hypothesized Model

3 Method

3.1 Procedure

An online survey was conducted in spring, 2012 to test the hypotheses in this model. Participants from a Korean University were recruited for this study by posting the survey link on the school community website, with an offer of 2000 won as compensation for participation. A total of 225 participants using Facebook produced usable responses. Respondents were 51% female, 61.5% used the original Facebook interface (while the rest used the timeline-based interface, which was introduced shortly before our data collection began), and the average age was 22.2 years.

On the first page of the study, participants were informed that they were participating in an investigation of individuals' Facebook usage. All participants were offered an informed consent form and a pre-knowledge question prior to their participation. Respondents were given the sample picture of the two types of Facebook interface—original interface and timeline interface. They were asked to choose the Facebook interface that they are currently using. Upon consenting with a click, they were directed to the corresponding questionnaire.

Before launch of the study, we pretested the entire questionnaire with 24 students (14 females and 10 males) from the same Korean university, who were also paid 2000 won for their participation. Based on the insights gained from the pretest, we were able to prepare the final version of the questionnaire by modifying and adding some items, especially questions about Facebook use behaviors.

3.2 Measurement

According to Facebook user menu and pretest results, we operationalized user behaviors on Facebook profile as "the degree of completing the profile fields" and Facebook wall behaviors as "the frequency with which users customized their Facebook wall." Respondents were asked to log into their Facebook account and answer the questions based on what they saw on their Facebook profile and wall. In the Facebook profile section, 33 categories were listed for measuring their profile behavior, such as "work", "relationship", "family", etc. Participants were asked to answer whether they have filled in those fields or not by selecting "yes" or "no". The number of categories that were indicated as "yes" was summed up as a continuous measure for "Facebook Profile Behavior" (M = 12.72, SD = 6.58). For the Facebook wall section, due to different functions provided in the timeline interface and the original interface, we chose the common items across the two interfaces. Specifically the customizable items are "change friend list", "delete posts", "change date", and "add position". The frequency for the four items was measured by using 0-7 scale (0 = Never, 7 = daily). The mean value of the four items was calculated as the value for Facebook wall behavior (M = 2.69, SD = 1.31).

In order to measure individual's level of self-esteem, six items were adopted from the Rosenberg Self-Esteem Scale [34], including "I feel that I'm a person of worth, at least on an equal plane with others", "I have a bunch of friends that care about me", and so on (Cronbach's α = .81). Responses were reported on a 7-point Likert scale with higher scores indicating higher self-esteem.

Sense of agency was measured via nine items based on Kim & Sundar [35]. These items consisted of "I feel that I can control my destiny", "I feel like I can exercise my free will", "I feel like I can access information that is appropriate for me", and so on. These items were measured using a 1-7 scale, ranging from strongly disagree to strongly agree (Cronbach's α = .86).

The Self-Monitoring scale has proven useful in measuring individuals' sensitivity to social presentation [36]. However, given the present study's focus on presenting oneself in the most positive light to others and based on psychometric data about the scale from past studies, six questions were adapted from the Other-Directedness Factor of the Self-Monitoring Scale [37]. Example items include "In different situations and with different people, I often act like very different persons", "In order to get along and be liked, I tend to be what people expect me to be rather than anything else", and so on. These items were measured using a 1-7 scale, ranging from strongly disagree to strongly agree. The ratings across the six items were averaged to yield the "Self-monitoring" index (Cronbach' α = .75).

We selected four items capturing the degree of public attachment from Ball and Tasaki's nine-item scale and revised them to reflect the participant's relationship with Facebook (e.g., "If someone hacked my Facebook page, I would feel like I have been personally attacked," etc.) [38]. The questions loaded under the factor labeled "Facebook as Social identity" (Cronbach's α = .68).

As we suggested, the emotional significance of one's Facebook profile would strengthen the personal facet of attachment. Therefore, an amended version of the Possession-Self link Scale [39] was adopted to measure the personal identity derived from Facebook. The answers to these questions were recorded on a 7-point Likert scale with a higher score indicating greater degree of perceived personal identity from Facebook. Reliability among the five items (e.g., "Facebook has helped shape my identity", "Facebook has helped me narrow the gap between what I am and what I try to be", etc.) was high (Cronbach's α = .90).

To yield the Virtual Possession index, five questions were created to capture the degree to which participants value Facebook. The monetary value relating to a social networking site as virtual possession was measured by the following items: "I don't care if all my data on Facebook is transferred to another website (Reverse)", "I would like to pay to add music on my Facebook", "If Facebook starts selling Facebook Emoticons, I am willing to pay for it", "If Facebook starts selling Facebook Wall paper, I am willing to pay for it", and "If Facebook starts selling Facebook Wall themes I am willing to pay for it" (Cronbach's α = .71).

In order to control for potentially confounding factors, individual demographic questions were asked (e.g., age, gender, grade, etc.). The Facebook Use Intensity Scale [28] was also used to control individual differences in Facebook use behavior as well as virtual possession attachment (e.g., "Facebook is part of my everyday activity", "Facebook has become part of my daily routine"). The six questions were asked on a 1-7 scale, with 1 representing strongly disagree and 7 representing strongly agree (Cronbach's α = .80).

3.3 Model Specification

Structural equation modeling (SEM) using AMOS 18.0 was used to test the hypothesized model. The hypothesized model estimated the extent to which individual differences on self-esteem predict their values on Facebook as a virtual possession, by identifying two different paths based on the theoretical discussion of the psychological differences of self concept on "Me" versus "I"-focused psychological characteristics and behaviors. One path was concerned with the concepts of sense of agency and Facebook profile behavior while the other one pertained to self monitoring and Facebook wall customization activity (see Fig. 1). Self-esteem was entered as the lone exogenous variable, with the two psychological variables (sense of agency, self-monitoring), Facebook use behaviors (profile behavior, wall behavior) and their attitude outcomes (Facebook as personal identity, Facebook as social identity), as well as the valuation of Facebook as a virtual possession, serving as endogenous variables.

3.4 Model Fit Indices

This study followed the indicators provided in AMOS to assess the overall model fit, including chi-square ($\chi 2$), the root mean square error of approximation (RMSEA), the goodness of fit index (GFI), the standardized root mean square residual (SRMR) and comparative fit index (CFI). Considering the fact that the chi-square test is sensitive to sample size [40], the other fit indices were consulted for determining the degree to which the data fit the hypothesized model.

4 Results

4.1 Preliminary Analysis

The data were checked for normality and outliers before the analyses. Based on the Mahalanobis distance criterion, four cases were identified as multivariate outliers and therefore removed from the dataset. As a result, the final sample size was 221. Using Kline's guideline for univariate normality, where an absolute value of less than 3.0 for skewness and 10.0 for kurtosis are considered normal, all variables in the model were normal [41]. In addition, multivariate normality was checked. The criterion for multivariate normality is that the Mardia's coefficient should be lower than p (p+2), where p is the number of observed variables [42]. The Mardia's coefficient for the current study was 150.06. Given that we had 25 observed variables, we can conclude multivariate normality.

Table 1. Correlations and descriptive statistics for all measured variables

	1	2	3	4	5	6	7	8
1. Self-esteem	1							
2. Sense of Agency	.693**	1						
3. Self-monitoring	- .155*	- .078	1					
4. FB as Personal Identity	- .028	.178**	.010	1				
5. FB as Social Identity	.247**	.323**	.088	.431**	1			
6. FB as Virtual Possession	-.182**	.064	- .030	.524**	.153*	1		
7. FB Profile Behavior	.040	.081	.040	.245**	- .020	.183**	1	
8. FB Wall Behavior	- .116	- .076	.073	.220**	.055	.243**	.325**	1
M	5.44	4.88	4.19	3.29	4.60	2.69	12.72	2.61
SD	.99	.95	1.12	1.29	1.15	1.09	6.58	1.27

Note: **. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Before fitting the hypothesized model, a confirmatory factor analysis (CFA) was performed to verify the factor structure of the observed variables in the model. The model fit was poor according to the following indices in CFA: $\chi 2$ = 1782.405, df = 838, p < .001; RMSEA = .072 (CI = .067- .076); CFI = .837; GFI = .716. Two items from 'Self-esteem', five items from 'Sense of Agency' and two items from 'Self-monitoring' variable were iteratively removed from the dataset because they were not significantly correlated with and/or had low factor loadings on their respective latent constructs. After eliminating these items, fit indices improved: $\chi 2$ = 699.507, df = 384, p < .001; RMSEA = .061 (CI = .054 -.068); CFI = .918; GFI = .825.

Furthermore, modification indices suggested allowing the following three pairs of error terms from 'Facebook as Personal Identity' and a pair of error terms from 'Self-esteem' and 'sense of agency' to covary: (1) "Facebook has helped shape my identity" with "Facebook has helped me narrow the gap between what I am and what I try to be"; (2) "Facebook has helped me narrow the gap between what I am and what I try to be" with "My Facebook page is part of who I am"; (3) "My Facebook page is part of who I am" with "I derive some of my personal identity from my Facebook page"; (4) "I have a distinct identity" with "I feel that I can control my destiny"; (5) "I take a positive attitude toward myself" with "On the whole, I am satisfied with myself". Finally, the revised model had an acceptable fit: $\chi 2$ = 363.999, df = 255, $p < .001$; RMSEA = .044 (CI = .033 - .054); CFI = .965; GFI = .885.

Based on the result of the CFA, reliability for each construct was evaluated with the remaining items. In order to test the interrelationships among the constructs in the model, the reliability (Cronbach's alpha) and the zero-order correlations were computed (see Table 1). Most of these bivariate correlations were significant and in a direction that was consistent with the hypotheses.

4.2 Testing the Hypothesized Model

To test the hypothesized model, SEM was conducted with the maximum likelihood estimation method. The result showed that the overall model was a reasonable fit in terms of RMSEA and CFI, but not SRMR and GFI: $\chi 2$ = 591.125, df = 314, $p < .001$; RMSEA = .063 (90% CI: .055 - .071); CFI = .912; SRMR = .1166; GFI = .842. Therefore, model modifications were conducted based on theoretical rationale and modification indices.

First, a pathway from 'Facebook wall behavior' to 'Facebook as Social Identity (FASI)' (β = 0.085, p = .288) was detected as non-significant. Owing to the fact that the study failed to support this originally hypothesized path, and considering the fact that 'FASI' holds no theoretical value as an independent exogenous variable, this latent construct was removed from the model entirely. After the deletion of 'FASI', overall model fit improved: $\chi 2$ = 400.537, df =244, $p < .001$; RMSEA = .054 (90% CI: .044 - .063); CFI = .947; SRMR = .0852; GFI = .869. Yet, the fit of this model hovered around acceptable criteria.

After considering the modification indices and based on theoretical rationale, a path from 'Facebook wall behavior' to 'Facebook profile behavior' was added to the model. The change resulted in the following fit statistics: $\chi 2$ = 370.213, df = 243, $p <$ .001; RMSEA = .049 (90% CI: .039 - .059); CFI = .957; SRMR = .0782; GFI = .877. Since 'FASI' was removed, the direct path from 'Facebook wall behavior' to 'Facebook as Personal Identity (FAPI)' was suggested by the modification indices. The data suggest that the more Facebook users engage in customization activities on the wall, the greater the sense of personal identity perceived from Facebook, instead of a sense of social identity. From a theoretical perspective, customization of the wall is likely to be associated with personal identity according to the agency model of customization [2]. After adding the path from 'Facebook wall behavior' to 'FAPI', the overall model fit improved to a respectable level: $\chi 2$ = 361.897, df = 242, $p < .001$;

RMSEA = .047 (90% CI: .037 - .057); CFI = .959; SRMR = .0729; GFI = .880. Fig. 2 depicts the final model that was retained for interpretation.

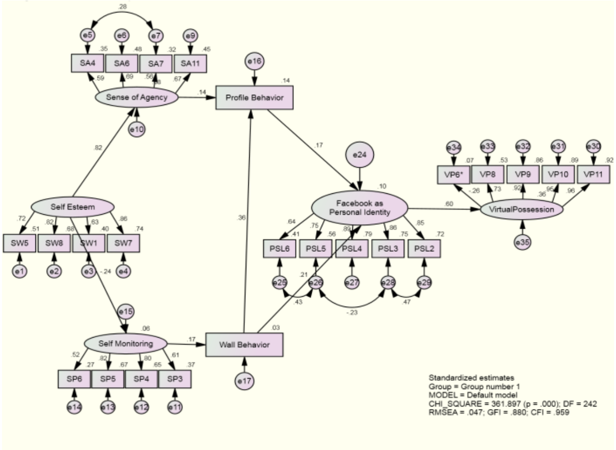


Fig. 2. Standardized path coefficients for the final model

4.3 Summary of Findings

The analysis showed that self-esteem has a significant positive effect on sense of agency (β = .82, $p < .001$; H1) and a significant negative effect on self-monitoring (β = - .24, $p < .005$; H2), meaning that individuals with a higher self-esteem have a higher sense of agency, while individuals who have a lower self-esteem tend to use self-monitoring strategy to control how they present themselves in the public (H1, H2 were supported). The results also suggest that self-esteem had stronger direct effects on sense of agency than on self-monitoring.

In addition, sense of agency is associated significantly with the information that was inputted in one's profile (β = .14, $p < .05$; H3); and the relationship between self-monitoring and Facebook wall behavior was also statistically significant (β = .17, $p < .05$; H4) (H3, H4 were supported). Although the variances explained by each psychological factor (sense of agency, self-monitoring) on the respective Facebook behaviors are not large, they are meaningful enough to reveal the relationships between self-esteem and the different kinds of Facebook activities by users. Furthermore, the path from Facebook wall behavior to Facebook profile behavior showed a significant relationship (β = .36, $p < .001$). This means that if individuals customize their Facebook wall more frequently to promote themselves in public, they are more likely to input more information in their profiles and thereby establish a more comprehensive online presence.

More interestingly, both the profile behavior and wall behavior are significantly associated with Facebook as Personal Identity (FAPI). Thus, both the amount of information in their profile (β = .17, $p < .05$) and the degree of wall customization (β = .21, $p < .05$) are positively related to the sense of personal identity derived from Facebook

(H5 was supported) rather than social identity (H6 was rejected). Ten percent of the variance for perceived personal identity from Facebook is accounted for by these two behaviors.

Also, results show that the higher the degree of personal identity perceived from Facebook, the higher the value of Facebook as a virtual possession for the users ($\beta = .60$, $p < .001$). Therefore, H7 was supported. The estimated variance of virtual possession accounted for by FAPI is 0.36.

Table 2. Estimated (Standardized & Unstandardized) S.E. and C.R. for the supported hypothesized paths

Paths	Estimates (Unstandardized)	Estimates (standardized)	Standard Error	Critical Ratio
Self-monitoring <--- Self-esteem	-0.28	-0.24	0.10	-2.90
FB WB<--- Self-monitoring	0.26	0.17	0.11	2.27
Sense of Agency <--- Self-esteem	0.84	0.83	0.13	6.64
FB PB <--- Sense of Agency	0.17	0.14	0.09	1.95
FB PB<--- FB WB	0.25	0.36	0.04	5.72
FAPI <--- FB PB	0.23	0.17	0.98	2.33
FAPI <--- FB WB	0.20	0.21	0.70	2.92
Virtual Possession <--- FAPI	0.68	0.60	0.74	9.21

Note: All of the paths were statistically significant at $p < .05$
FAPI: *Facebook* as Personal Identity; FB: *Facebook;*
WB: Wall Behavior; PB: Profile Behavior

In sum, the data show that self-esteem is positively related to sense of agency and negatively related to self-monitoring, with these psychological characteristics predicting user behavior on Facebook. We found that the more users engage in using Facebook, both for completing Facebook profile and customizing Facebook wall, the more they attached their personal identities to Facebook because both the profile and the wall end up reflecting the self. Table 2 provides an overview of the estimates, standard errors and critical ratios for the paths in the model.

5 Discussion

Overall, our analysis suggests, based on structural equation modeling, that the psychological factors underlying Facebook user behaviors influence the degree to which Facebook is seen as a reflection of one's personal identity and the extent to which this social-media site is valued as a virtual possession.

5.1 The "I" vs. "Me" Paradigm in Facebook

This study proposed two possible paradigms, the "I" and the "Me", in one's selfhood within the virtual reality of SNS. As we hypothesized, self-esteem is related to the two facets of selfhood, each associated with distinct psychological factors and

Facebook behaviors. The sense of "I"-ness focuses on a reflection of true inner self; the sense of "Me"-ness is externally influenced, wherein one's sense of self is derived from one's public presentation. For those individuals who have the "I" oriented personality, i.e., a higher level of self-esteem, a higher sense of agency was reported; while for those individuals with a "Me" orientation and a lower level of self-esteem, a higher need for self-monitoring was noted. These psychological states are able to predict distinct user behaviors on Facebook. Individuals who perceived a higher sense of agency tend to complete their profiles with more information, showing their true self and constructing a detailed online presence. On the other hand, individuals who show a higher level of self-monitoring tend to customize their online public presentations (by altering entries in their Facebook wall) more frequently. These behaviors have implications for the design of audience awareness mechanisms. On the one hand, interface features could be designed to inform users about the number of times their profile and wall postings were viewed by others. On the other hand, a system utility can be designed to keep track of the frequency of changes to the user's profile and modifications of wall postings.

As it turns out, these two activities are related. We found that individuals who customize their Facebook wall more frequently have a higher level of profile completion. It suggests that the degree to which individuals promote themselves in public has an influence on their desire to construct a formal and complete online presence. In turn, profile behavior emerged as an important construct connecting one's self-esteem to the tendency for projecting one's online personal identity through Facebook in the final model. All this implies that users value the profile space in Facebook as an important part of their online persona construction. Therefore, design changes to Facebook profile should be geared toward affording users the ability to customize in ways that can project their personal identity.

One of the notable findings of this study is that instead of increasing a sense of social identity as hypothesized, behavior on the Facebook wall (that is presumably meant for carefully articulating one's public self-presentation) was actually positively correlated with private identity-focused attachment. While some previous research has shown that the other-directedness in self-monitoring is negatively correlated with self-focused identity building, other research has shown that self-monitoring could also be seen as self-expression rather than self-promotion [43]. Instead of simply advocating oneself, self-monitors may be more likely to use this strategy to construct a self-concentrated identity [44]. Our data suggest that, over time, users accept Facebook as part of their routine and do not see it simply as a networking tool, but rather as a venue for expressing and shaping one's self identity, be it their true self or artificial self. This has implications for Web 2.0 application design, in that greater emphasis should be placed on tools that help users shape their personal identity. Based on their online activities, individual users could be offered a personalized palette of online personas to choose from so that they can visualize how others may view them. Various framing strategies can be offered to users with the help of tools that ask users whether they want to be seen by others in this light or that way, and then proceed to identify those profile elements and wall activities that are consistent and those that are inconsistent with their chosen identity, so that users can make judgments about their own online identity as well as decisions about projecting the type of identity that they wish to present to others.

5.2 Personal Identity and Virtual Possession

A statistically significant portion of the variance in perceived attachment to Facebook as personal identity can be explained by the two kinds of behaviors we examined in this study. The results indicate that Facebook successfully attracts users who create different strategies for representing themselves online and, more importantly, for expressing their true selves. As mentioned above, the present study only distinguishes two types of user behaviors on Facebook — information completion in the profile section and customization frequency in the wall section. While there are clearly many other behaviors in Facebook, we could only include those that lend themselves to counting and are explicitly recorded on Facebook, given the limitations of the self-report survey method that we employed. Hence, it is not surprising that the variance in self-focused attachment to Facebook that is explained by the two kinds of behaviors measured in the current study is only 10 percent, but it is nevertheless a significant indicator of the behavior-attachment connection.

Even though we focused on Facebook's potential to promote both social identity and personal identity as potential predictors of users' valuation of Facebook as a virtual possession, it was clear from the data that Facebook as Personal Identity played the more important role in the model. This demonstrates how closely linked Facebook is to one's representation of self-identity and how it affects the extent to which one values Facebook. This effect differs from the public-focused self-identity that users perceive from Facebook, which failed to show a significant association with Facebook user behaviors measured in this study. In fact, compared to the frequency of wall customizing, the extent of profile completion has a greater influence on the value of Facebook as a virtual possession, through personal identity. This has important implications, as discussed in the next section.

5.3 Value System of Virtual Possessions

Ultimately, it appears that self-esteem is a driving force behind our SNS use, and the extent to which the SNS reflects our personal identity is indeed predictive of the extent to which we value it as a possession. If it reflects our personal identity, we are willing to pay money for purchasing accessories on Facebook. By extension, users who see a strong personal identity reflected in their Facebook environment are more likely to pay for Facebook as a service if it were to charge usage fees. This insight can be quite useful for designers of social media sites and social apps for mobile devices. To the extent these sites and apps provide affordances for users to construct and shape their personal identity, they are likely to be valued highly by users.

5.4 Implications for Web 2.0 Applications Design and Testing

In addition to the aforementioned design implications, our study offers insight into measuring user experience. While traditional HCI literature informs us how to evaluate the quality and usability of websites and applications with usability testing (assessing such constructs as ease of use, usefulness, and navigability), our study identifies latent attributes that affect users' loyalty and perceived value of these sites. Our results and discussion clearly show that personal identity is a significant context of use for Web 2.0, meriting greater attention from designers. The strong connection between personal identity and value of Facebook as a virtual possession indicates that the realization of personal identity should be among the quality attributes used for testing. As Orehovački [45] suggested, the methodology for evaluating the quality of social networking applications like Facebook should be extended from usability to user experience. Based on the results of our study, we recommend that personal identity, as a subjective quality, be reflected in user-experience metrics when evaluating Web 2.0 applications for design improvements.

5.5 Limitations and Directions for Future Studies

This study is not without the usual limitations of exploratory research. First of all, because this study examines the value placed on virtual possessions, with some of them in monetary terms, a more heterogeneous sample — rather than the convenient college group used—is needed to predict users' valuation of SNS as a virtual possession. Second, this survey-based study also suffered from specific disadvantages due to the self-reported responses, especially for questions about specific behavior frequency and situations. Participants may have forgotten pertinent details. Furthermore, as mentioned above, this study did not capture all the categories of Facebook behaviors. It would certainly be interesting to study other categories of Facebook user behaviors such as frequency of profile photo updating and type of privacy setting. In addition, the current study only examined variables in the context of Facebook. For further studies, different SNS services could be used for testing the theoretical model.

Despite these limitations, the current study marks an important step in developing a psychological model about self-esteem and self-identity that involves both the antecedents and the consequences of Facebook user behaviors. Implications for SNS owners and advertisers may be derived from the results. For example, designers could consider all the ways in which they can enable users to build their true selves on SNS because a strong personal identity can result in greater attachment and valuation. Advertisers and social media marketers who can leverage the affordances of Facebook to offer users novel and creative opportunities to construct and exhibit their personal identity are likely to be valued more highly than their counterparts who simply focus on exposing Facebook users to their commercial appeals. Likewise, SNS groups that appeal to users' personal identities are likely to garner more support and following, especially when these groups appear as an integral part of one's profile on Facebook.

6 Conclusion

Social networking technology has opened up numerous possibilities for ideal self-presentation, which not only helps us realize our individual identity but also shapes our public persona. On Facebook, we customize our wall from time to time by deleting unwanted posts or choosing to experience different levels of interpersonal relationships with different individuals. As shown by our second (and final well-fitting) model, the lower one's self-esteem, the higher the possibility that one will adopt the self-monitoring strategy of camouflaging one's true self. However, our data also suggest that these strategies are not simply adopted to create a perfect social identity, but to articulate one's own ideal self-identity [46].

While some scholars (e.g., [29]) argue that technology is diminishing our true selves by encouraging us to constantly seek attention from others, our data suggest that College-age users view SNS technology as an extension of oneself, not a substitute. Individuals still possess a strong eagerness to define their individual identity and to control their own feelings in a social networking environment, especially when they have a high sense of self-esteem. The eagerness to have a full online presence seems to be stronger than the desire to monitor oneself in public social networks. Further, it is clear that self-esteem affects the value we place on Facebook as a virtual possession largely because we see Facebook as a vehicle for shaping and representing our personal identity. In conclusion, our valuation of Facebook as a virtual possession is largely a result of the "I" dominated selfhood rather than a "Me" dominated outlook. We value Facebook and spend so much time with it because it offers us numerous opportunities to construct, shape and reflect our personal identity, not because it allows us to project a certain social identity to members of our network. Therefore, despite all its socialness, Facebook is a deeply personal medium.

Acknowledgement. This research is supported by the Singapore National Research Foundation under its International Research Center Keio-NUS CUTE Center @ Singapore Funding Initiative and administered by the IDM Program Office.

References

1. Greenhow, C., Robelia, B.: Informal Learning and Identity Formation in Online Social Networks. Learning Media And Technology 34, 119–140 (2009)
2. Sundar, S.S.: Self as Source: Agency and Customization in Interactive Media. In: Konijn, E.A., Sonja, U., Martin, T. (eds.) Mediated Interpersonal Communication, pp. 58–74. Routledge, New York (2008)
3. Turkle, S.: Life on the Screen: Identity in the Age of the Internet. Simon & Schuster, New York (1995)
4. Turkle, S.: The Second Self: Computers and the Human Spirit. The MIT Press, Cambridge (1985)
5. Odom, W., Zimmerman, J., Forlizzi, J.: Virtual possessions. In: Proc. of the 8th ACM Conference on Designing Interactive Systems, pp. 368–371. ACM, New York (2010)
6. Mead, G.H.: Mind, self, and society: From the Standpoint of a Social Behaviorist. University of Chicago Press, Chicago (1967)
7. Kolb, W.L.: A Critical Evaluation of Mead's "I" and "Me" Concepts. Social Forces 22(3), 291–296 (1944)
8. Pelham, B.W., Swann, W.B.: From Self-conceptions to Self-worth: on the Sources and Structure of Global Self-esteem. Journal of Personality and Social Psychology 57, 672–680 (1989)

9. Crocker, J., Wolfe, C.T.: Contingencies of Self-worth. Psychological Review 108, 593–623 (2001)
10. Campbell, J.D., Trapnell, P.D., Heine, S.J., Katz, I.M., Lavallee, L.F., Lehman, D.R.: Self-concept Clarity: Measurement, Personality Correlates, and Cultural Boundaries. Journal of Personality and Social Psychology 70(1), 141–156 (1996)
11. Baumeister, R.F., Campbell, J.D., Krueger, J.I., Vohs, K.D.: Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? Psychological Science in the Public Interest 4, 1–44 (2003)
12. Bem, D.J.: Self-perception Theory. In: Berkowitz, L. (ed.) Advances in Experimental Social Psychology, pp. 1–62. Academic Press (1972)
13. Tice, D.M.: Self-concept Change and Self-presentation: the Looking Glass Self is also a Magnifying Glass. Journal of Personality and Social Psychology 63, 435–451 (1992)
14. Judge, T.A., Erez, A., Bono, J.E., Thoresen, C.J.: Are Measures of Self-esteem, Neuroticism, Locus of Control, and Generalized Self-efficacy Indicators of a Common Core Construct? Journal of Personality and Social Psychology 83, 693–710 (2002)
15. Gallagher, I.: Philosophical Conceptions of the Self: Implications for Cognitive Science. Trends in Cognitive Sciences 4, 14–21 (2000)
16. Zarghooni, S.: A Study of Self-Presentation in Light of Facebook. Institute of Psychology, University of Oslo (2007)
17. Steinfield, C., Ellison, N.B., Lampe, C.: Social Capital, Self-esteem, and Use of Online Social Network Sites: a Longitudinal Analysis. Journal of Applied Developmental Psychology 29, 434–445 (2008)
18. Manis, J.G., Meltzer, B.N.: Symbolic Interaction: a Reader in Social Psychology, 3rd edn. Allyn and Bacon, Inc., Boston (1978)
19. Goffman, E.: The Presentation of Self to Others. In: Manis, J.G., Meltzer, B.N. (eds.) Symbolic Interaction: A Reader in Social Psychology, pp. 171–178. Allyn and Bacon, Inc., Boston (1978)
20. Ellison, N.: Facebook Use on Campus: A Social Capital Perspective on Social Network Sites. In: ECAR Symposium, Boca Raton, FL (2007)
21. Gonzales, A.L., Hancock, J.T.: Mirror, Mirror on My Facebook Wall: Effects of Exposure to Facebook on Self-esteem. Cyberpsychology, Behavior and Social Networking 14, 79–83 (2011)
22. Bong, M., Clark, R.: Comparison between Self-concept and Self-efficacy in Academic Motivation Research. Educational Psychologist 34, 139–153 (1999)
23. Byrne, B.M.: The General/Academic Self-concept Nomological Network: A Review of Construct Validation Research. Review of Educational Research 54, 427–456 (1984)
24. Reynolds, K.J., Turner, J.C., Branscombe, N.R.: Interactionism in Personality and Social Psychology: An Integrated Approach to Understanding the Mind and Behaviour. European Journal of Personality 482, 458–482 (2010)
25. Gecas, V., Burke, P.J.: Self and Identity. Social Psychology: Sociological Perspectives on Social Psychology, pp. 41–67 (1995)
26. Hogg, M.A., Reid, S.A.: Social Identity, Self-Categorization, and the Communication of Group Norms. Communication Theory 16, 7–30 (2006)
27. DiMicco, J.M., Millen, D.R.: Identity Management: Multiple Presentations of Self in Facebook. In: The 2007 International ACM Conference on Supporting Group Work, pp. 383–386. ACM Press, New York (2007)
28. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook "friends": Social capital and college students' use of online social network sites. Journal of Computer Mediated Communication 12(4), 1143–1168 (2007)

29. Turkle, S.: Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books (AZ), New York (2012)
30. Odom, W., Zimmerman, J., Forlizzi, J.: Teenagers and their Virtual Possessions: Design Opportunities and Issues. In: Proc. of the 2011 Annual Conference on Human Factors in Computing Systems (CHI 2011), pp. 1491–1500. ACM, New York (2011)
31. Csikszentmihalyi, M., Halton, E.: The Meaning of Things: Domestic Symbols and the Self. Cambridge University Press, Cambridge (1981)
32. Kleine, S.S., Baker, S.M.: An Integrative Review of Material Possession Attachment. Academy of Marketing Science Review 1(1), 1–39 (2004)
33. Belk, R.W.: Possessions and the Extended Self. Journal of Consumer Research 15(2), 139–168 (1988)
34. Rosenberg, M.: Society and the Adolescent Self-image, revised edn. Wesleyan University Press, Middletown CT (1989)
35. Kim, Y., Sundar, S.S.: Can Your Avatar Improve Your Health? The Impact of Avatar Attractiveness and Avatar Creation. In: Proc. of the 61st Annual Conference of the International Communication Association, Boston, MA (2011)
36. Berinsky, A.J.: Can We Talk? Self-Presentation and the Survey Response. Political Psychology 25(4), 643–659 (2004)
37. Briggs, S.R., Cheek, J.M., Buss, A.H.: An Analysis of the Self-monitoring Scale. Journal of Personality and Social Psychology 38, 679–686 (1980)
38. Ball, A.D., Tasaki, L.H.: The Role and Measurement of Attachment in Consumer Behavior. Journal of Consumer Psychology 1, 155–172 (1992)
39. Ferraro, R., Escalas, J.E., Bettman, J.R.: Our Possessions, Our Selves: Domains of Self-worth and the Possession–self Link. Journal of Consumer Psychology 21, 169–177 (2011)
40. Bentler, P.M.: Comparative Fit Indexes in Structural Models. Psychological Bulletin 107, 238–246 (1990)
41. Kline, R.B.: Principles and Practice of Structural Equation Modeling. The Guilford Press (2010)
42. Bollen, K.A.: Structural Equation Models. In: Armitage, P., Colton, T. (eds.) Encyclopedia of Biostatistics, vol. 7, pp. 4363–4372. Wiley Online Library, New York (1998)
43. Lee, S.: Psychological Effects of Culture on Aesthetic Motivation for Cosmetic Customization of Mobile Phones. PhD dissertation, The Pennsylvania State University (2011)
44. Snyder, M., Copeland, J., Giacalone, R.A., Rosenfeld, P.: Self-monitoring Processes in Organizational Settings. In: Giacalone, R.A., Rosenfeld, P. (eds.) Impression Management in the Organization, pp. 7–19. Erlbaum (1989)
45. Orehovački, T.: Development of a Methodology for Evaluating the Quality in Use of Web 2.0 Applications. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 382–385. Springer, Heidelberg (2011)
46. Bargh, J.A., McKenna, K.Y.A.: The Internet and Social Life. Annual Review of Psychology 55, 573–590 (2004)

Authoring Support for Post-WIMP Applications

Katharina Gerken, Sven Frechenhäuser, Ralf Dörner, and Johannes Luderschmidt

RheinMain University of Applied Sciences, Wiesbaden, Germany
Katharina.B.Gerken@student.hs-rm.de,
{Sven.Frechenhaeuser,Ralf.Doerner,
Johannes.Luderschmidt}@hs-rm.de

Abstract. Employing post-WIMP interfaces, i.e. user interfaces going beyond the traditional WIMP (Windows, Icons, Menu, Pointer) paradigm, often implies a more complex authoring process for applications. We present a novel authoring method and a corresponding tool that aims to enable developers to cope with the added level of complexity. Regarding the development as a process conducted on different layers, we introduce a specific layer for post-WIMP in addition to layers addressing implementation or traditional GUI elements. We discuss the concept of cross layer authoring that supports different author groups in the collaborative creation of post-WIMP applications permitting them working independently on their respective layer and contributing their specific skills. The concept comprises interactive visualization techniques that highlight connections between code, GUI and post-WIMP functionality. It allows for graphical inspection while transitioning smoothly between layers. A cross layer authoring tool has been implemented and was well received by UI developers during evaluation.

Keywords: authoring processes, authoring tools, post-WIMP interfaces, cross layer authoring, collaborative user interface development, combined post-WIMP interactions, visual validation.

1 Introduction

Authoring processes and tools for creating classical GUIs (Graphical User Interfaces) that adhere to the WIMP (Windows, Icons, Menus, Pointer) paradigm have been explored and refined over the years and developers have been honing their skills in creating these types of user interfaces. With post-WIMP [14] becoming more commonplace [11], the authoring process for user interfaces is becoming more complex, as the diversity of interaction methods rises [1, 2]. Novel tasks need to be carried out during authoring and new skill sets are required, for example expertise in reality-based interaction [5], in order to develop post-WIMP user interfaces. As single authors seldom possess the whole range of skills required, collaboration of multiple author groups with specific expert knowledge is often needed [4, 12]. We identified another factor that distinguishes authoring of post-WIMP user interfaces: many post-WIMP interface elements (e.g. gestures, speech input) have no graphical

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 744–761, 2013.

representation in the user interface and therefore make it more difficult to graphically represent them in authoring tools. In addition, authors may not understand from visually inspecting the user interface which interactions are implemented and how they work together. As a consequence, existing tools for GUI development may not be well suited for post-WIMP. Dedicated authoring tools are necessary to create these types of user interfaces efficiently, to ensure high quality, to facilitate meaningful inspection and tests, to enable authors to contribute their expertise to the development and to reduce overhead due to several authors having to cooperate in the development.

This paper presents an authoring process and an authoring tool for post-WIMP user interfaces that take the peculiarities of post-WIMP methods into account and foster collaboration between different author groups. In our concept, we distinguish between four author roles: the programmer writing code, the GUI designer building the application screens, the post-WIMP designer who adds further input channels to the screens and the client who wishes to customize the application. In contrast to using several different tools for the respective development tasks, we expect the authors to better understand the connections between code, GUI and post-WIMP interaction when working with only one tool [17]. This may lead to a more comprehensible development process. It also supports the refactoring and iterative refinement of applications. We call this cross layer authoring, as we regard the development of a post-WIMP application as a process that is conducted on several layers denoting the contributions of the distinct authors. Each layer corresponds to a different author role with different task areas. The authors should be able to work on their respective layers independently, although certain dependencies do exist: The code layer can be regarded as the fundamental contribution to the application being developed. On top of that, the GUI designer can build the application's appearance, which can then be enhanced with interaction by the post-WIMP designer. Each layer offers abstract connection hooks for the other layers, allowing the authors to be able to work on their particular layer only. Thus we expect them to be able to focus and immerse themselves in their respective task better, which might lead to an increase in quality of the resulting applications. In order to inspect the connections between layers, specific interactive visualization techniques are introduced. Transitions between layers are designed to be seamless in order to reduce the loss of orientation when switching layers. In our approach, we aim at preserving traditional GUI authoring as many authors are familiar with this and have already acquired a profound experience. Thus, post-WIMP is added as additional layer and is separated from developing the WIMP aspects of a user interface. Our paper offers the following contributions:

- We introduce the concept of a single cross layer authoring dedicated for the development of post-WIMP user interfaces.
- We identify four layers as one authoring dimension and four editing modes as an orthogonal dimension resulting in a 4 x 4 state matrix of the authoring tool.
- We conceive visualizations for interactive transitions between layers and for the creation and inspection of post-WIMP interactions in a cross layer authoring tool.
- We present a prototype implementation demonstrating the feasibility of our concept.

The paper is organized as follows: First, related work is presented to illustrate the state of the art in authoring for post-WIMP applications. Then our cross layer authoring concept and CLAY (Cross Layer Authoring Tool) are being introduced and illustrated with a small use case. After explaining the basic navigation elements of the tool, we focus on the creation and configuration of post-WIMP functionality. Finally, we discuss our concept and the CLAY editor by evaluating user feedback to our prototype and stimulating possible future research topics.

2 Related Work

There exist several frameworks and libraries which enable programmers to add post-WIMP features to existing applications or to develop applications with post-WIMP user interfaces, e.g. PyMT [3] for touch or the Kinect SDK[1] for full body interaction. Those frameworks provide powerful tools for programmers, but are less fitted for interface designers [8, 9]. In the following, we focus on concepts assisting programmers on the one hand and supporting visual development (for the GUI designer, post-WIMP designer or client) on the other hand.

One approach to handle the complexity of the development of post-WIMP applications is to enhance a classical WIMP UI by merging post-WIMP capabilities on the same abstraction layer, as can be found in Apple's authoring tool named Xcode[2], for instance. With Xcode, the WIMP UI can be created graphically in a GUI editor. The UI elements are then bound to UI events by dragging graphical connection lines to methods or properties specifically marked in the source code. However, there is no visual support to add post-WIMP input (e.g. custom gestures) to the UI, so it can hardly be achieved without fundamental knowledge of the required programming techniques. This might lead to the programmer and designer having to spend more time negotiating specifications, which may add a significant overhead and prohibits focus on each author's individual task. Also, Xcode's concept does not include support for testing gestures directly while creating them. To inspect the created behavior, the use of a simulator or device is required. This makes the authoring process of post-WIMP capabilities inefficient, because the entire resulting application must be compiled and executed to test the behavior every time the logic or the parameters of the specific gesture change.

While Xcode's approach focuses on multi-touch interactions, there are other concepts of authoring tools that support the development of applications featuring all kinds of post-WIMP interactions. One of them is Squidy, a zoomable design environment for natural user interfaces [10]. It employs other concepts like the MaggLite authoring toolkit for post-WIMP interfaces [4] or the OpenInterface Framework [13], which offer a way to create post-WIMP user interfaces with the concept of visual dataflow programming. Squidy addresses the issue that the authors involved in the authoring process of post-WIMP UIs often use multiple toolkits or frameworks to create the desired UI and

[1] Microsoft Kinect for Windows: `http://www.microsoft.com/en-us/kinectforwindows/`

[2] Apple Developer: `https://developer.apple.com/`

its behavior. This is achieved by tying together relevant frameworks and toolkits in a common library, while a visual language is introduced to create natural user interfaces. To see the current development status as a whole, Squidy introduces the concept of semantic zooming, making it possible to control the level of complexity shown to the particular author developing the application. However, Squidy has a focus on providing a way to take input data of hardware devices (e.g. movement data recorded by a Kinect), process the data and send it to listening applications, e.g. via TUIO [6]. This focus on technical aspects is important to develop post-WIMP input alternatives actually available for a specific application. Still Squidy lacks the possibility to design the user interface and to add the created post-WIMP input methods to it. In addition, the capabilities of adding and validating the post-WIMP input on the actual application UI are missing. Thus, designers without any technical knowledge about the specific implementation of the post-WIMP input may have difficulties using the technical base and design as well as validating the UI as a whole.

Hartmann et al. present a rapid prototyping tool called Exemplar [15] that allows for authoring of sensor-based interactions by demonstration, thus enabling short iterative demonstrate-edit-review cycles. Yet, it only covers one step in the application development process, forcing the author to switch between different tools, whereas we seek to embed interactive editing (particularly of post-WIMP behavior) seamlessly into the overall authoring process.

Juxtapose [16], a prototyping code editor and runtime environment for UIs, allows for creating multiple alternatives of application logic and real-time tuning of application parameters. Juxtapose facilitates the work of UI developers, yet it only addresses authors with both UI design and programming knowledge, as the parameters are directly extracted from the code. The CLAY concept shares the goal to enable rapid parameter testing, however, it strives to provide it on a more abstract level for UI designers who do not necessarily know how each parameter is implemented in the program code. Generally, the Juxtapose concept lacks the distinction between different author roles.

Simpson and Terry [17] point out the problem of the use of heterogeneous toolsets for application development and present GUIIO, a design tool which uses ASCII text for rendering UI mock-ups. While it offers a fluid transition from program code to UI, it does not offer possibilities to inspect the UI in the way it would actually present itself to the user, nor does it provide means to graphically represent post-WIMP interaction. Instead of just visualizing the connection between code and UI from the perspective of the coder, we introduce a concept comprising the points of view of all the authors involved in the process, allowing them to work together with only one tool.

DENIM [18] is an informal website design tool presented by Newman et al. in 2003. It allows for informal sketch-based prototyping of websites and offers different representations of the sites through zooming. Similarly to CLAY, DENIM uses a slider with eleven levels to navigate from site map to storyboard and individual page view, thus enabling an almost smooth transition. Yet, the zooming in DENIM is only used to reveal more or less details of only one field of activity (the graphical user interface design), whereas CLAY uses seamless transitions to navigate between views of entirely different natures (user view, post-WIMP functionality, classic GUI and code). So instead of zooming "horizontally" on one aspect of an application, the CLAY concept allows for a "vertical" zoom through the different areas of expertise of the authors involved.

To sum up, authoring tools with a focus on a specific platform provide visual means to develop the WIMP part of the UI. Yet the development of post-WIMP capabilities has to be done entirely in the source code by programmers with specific technical knowledge. Besides, the resulting behavior is hard to inspect without compiling the application and testing on the target device [7]. More general approaches which support adding all kinds of post-WIMP capabilities to applications often have a focus on technical aspects. They lack the support for designing the application's UI, integrating post-WIMP capabilities and visually testing the added post-WIMP interactions. Different authoring roles are not differentiated enough to allow authors to concentrate on their specific contribution to the post-WIMP application.

3 Cross Layer Authoring

We distinguish between four different author roles in the process of a post-WIMP application development: the coder, the GUI designer, the post-WIMP designer and the user. We regard the authors' contributions to the application being developed as layers placed above one another, with the code layer being at the bottom. Above that, there are the GUI layer, the post-WIMP layer and the user layer. The code layer can be regarded as the fundament of the application, which all other layers depend upon. Likewise, the post-WIMP layer builds upon the interface layout and traditional elements (such as buttons, text fields) created in the GUI layer. Each layer is designed to suit the tasks and skills of the respective author only, e.g. source code can only be edited in the code layer, graphical interface design only takes place in the GUI layer.

Apart from distinguishing between author roles, we further split up the task areas of each of these roles into four editor modes, namely content, layout, parameters and behavior, which indicate what kind of changes can currently be made to the application. Dividing all the editable attributes into four different modes aims at keeping the screen organized and minimizing the risk for accidental changes (e.g. changing the position of a text label when intending to change its content). An overview of the resulting sixteen development modes and the respective functionality available in each mode can be found in Figure 1.

During the authoring process, cross layer connections between GUI components, post-WIMP input methods and the application logic on the code layer have to be created. By visualizing the connections between the layers when navigating between them, information about how the layers work together can be provided to each author. Transitions between the author layers are designed to suit the authoring task at hand and to happen seamlessly in order to avoid disorientation due to abrupt changes in the UI of the authoring tool.

In order to be able to evaluate this concept, we implemented the CLAY editor. CLAY is realized as a classic desktop application as from our observation many developers prefer a PC-based environment with multiple screens. Development especially on smartphones with small screens does not seem to be practical. Nevertheless, the concept could be adapted to work on specific post-WIMP devices as well.

	Content	***Layout***	***Parameters***	***Behavior***
User	Edit text fields, change images etc.	Arrange images, texts, buttons etc. on the screens	Adjust fonts, colors etc.	Inspect application behavior by moving the cursor over UI elements
Post-WIMP	Simulate/test post-WIMP functionality	Compose post-WIMP actions, connect them to the GUI	Adjust e.g. speed for a double tap, length of a pan gesture, voice input parameters	Connect code actions or screen transitions to post-WIMP input
GUI	Set label texts, images etc.	Arrange screens and GUI elements on the screens	Adjust size, color, appearance of GUI elements	Connect code actions or screen transitions to WIMP input
Code	Display and edit available code files	Create and edit UI components to be used by the GUI and post-WIMP layers	Define access rights for component properties for the GUI and post-WIMP layer	Create actions to be used by the GUI and post-WIMP layer

Fig. 1. Overview of the different development modes of CLAY and their respective functionality: We distinguish between four author roles (leftmost column) and four editor modes (top row)

3.1 The CLAY Prototype

In the CLAY user interface, we place graphical representations of the previously described layers and editor modes at the borders of the authoring tool (see Figure 2). They are arranged as two orthogonal axes forming a two-dimensional matrix where one dimension serves to distinguish between the authoring layers and the other dimension is used to switch the editor mode. With this, we provide a simple mental model how the authoring functionality in the tool is structured. There is constant visual feedback in which state the authoring tool is set.

The slider on the left-hand side allows for a smooth transition between author role layers. When the slider is positioned directly over a layer icon, only the respective layer is visible at 100 percent opacity. When the slider is moved towards a different layer, the current layer becomes more and more translucent to show information from the layer the slider is being moved to. This blending should eliminate the need for entirely switching between layers in many cases. If, for instance, a button is connected to a code action, selecting the button on the GUI layer and moving the slider towards the code layer will result in the connected method shining through the GUI layer. When approaching the code layer, the semantic level of detail of the code information increases until the slider has reached the code layer icon, where the whole code file is shown in a text editor (see Figure 3). In contrast to a navigation where layers are discretely switched, we expect the slider to provide a better orientation and understanding of connections between layers.

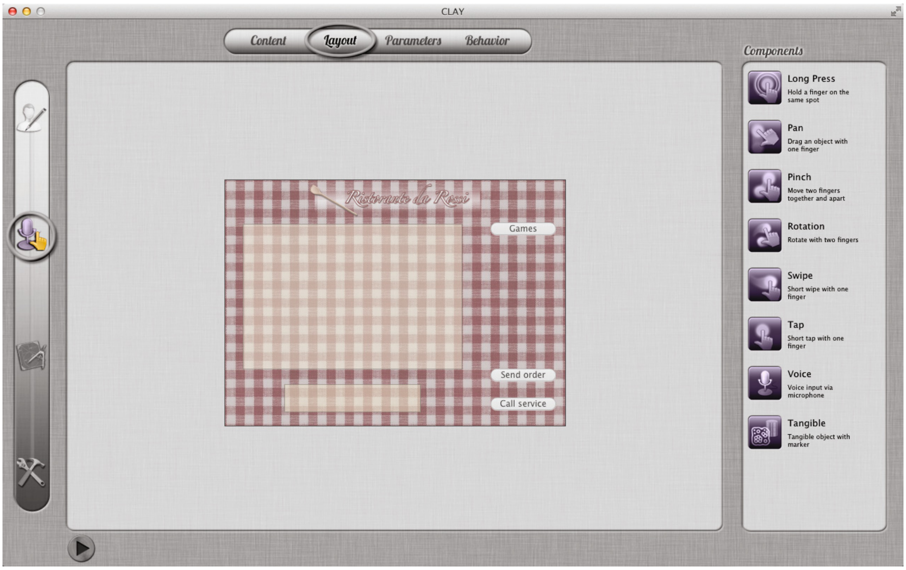


Fig. 2. The CLAY interface: A slider on the left allows for navigation through the author layers, a slider at the top can be used for switching editor modes.

Furthermore, the interface is divided up into two frames: a main frame in the center and a smaller frame on the right. The main frame serves to type code or put together GUI or post-WIMP elements, depending on the active layer. The right frame on the other hand is used to display the elements or options available in the current layer and editor mode. On the uppermost layer, the user layer, there is no right frame. Instead, the main frame is made wider to make more room for the screens of the application being inspected.

On the horizontal axis at the top, the developer can switch between the four editor modes content, layout, parameters and behavior. When switching between editor modes, the right frame's content adapts accordingly to provide a list of the specific components or parameters needed.

The play button in the bottom left can be accessed from all layers at any time. Pressing the button results in the application being compiled and run in a simulator in full screen mode. With the target device connected, this button also allows for a test run on the specific device. Though the resulting application can always be inspected in the user layer, it is important to provide a simulator where no other graphical elements of the authoring tool itself are visible in order to get a good understanding for the resulting application. It is even more important to provide this functionality as CLAY runs on desktop computers where post-WIMP input can often only be approximated by WIMP input.

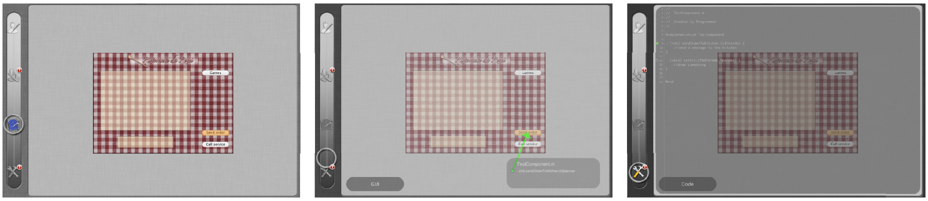


Fig. 3. Three phases of the smooth transition from the GUI to the code layer: As the slider on the left is moved from the GUI to the code layer, the interface builder becomes translucent and reveals more and more code layer details.

3.2 The CLAY Authoring Process

Throughout the next sections we illustrate the CLAY authoring process employing the following use case: Restaurant owner R has furnished his restaurant with multi-touch tables and needs a tabletop software which should allow his guests to skim through the food menu, place orders and play games while they are waiting for their meals.

There are three authors involved in the development process: Coder C is responsible for the actual functionality of the program developed. GUI designer G is responsible for the selection and layout of classical WIMP user interface elements and the transitions between screens. Post-WIMP designer P works on the screens provided by the GUI designer and adds post-WIMP functionality to the application. Restaurant owner R might be involved in the process at times, and may customize the product to a restricted extent.

Our tool provides views for each of these four roles. The view of the end user, a guest in the restaurant, is included in form of a simulator. In the following section, we elaborate in the challenges of each of the authors, introduce their layers and the associated editor modes in detail and motivate the need for smooth transitions between layers.

Code Layer. The programmers need a text editor with syntax highlighting and refactoring options in order to implement the custom GUI and post-WIMP component classes and the methods called by them. In addition to writing code, they have to provide hooks for the upper layers to enable them to connect to their methods and use their classes without intruding into the source code.

In our concept, these hooks are called components and action elements, which are being created by the coder as abstract representations for classes and methods. They create a level of abstraction for the upper layers and allow for exchangeability on the code layer without the GUI and post-WIMP layers having to perform any changes. These abstract elements also allow for simultaneous development across the layers, as the actual implementation of e.g. a specific method does not need to be completed in order for the upper layers to be able to connect to it.

In our example, coder C creates a new file in the content mode. After implementing a method, e.g. for sending an order to the kitchen, C switches to the behavior mode where she can create and edit action elements which are displayed in the right frame. She has to provide a name and description for the action (e.g. "Send order to kitchen”), which will be displayed in the GUI and post-WIMP layers. Additionally, she can restrict the action to be only available to certain components like a button or a tap. In doing so, C would prevent G or P from inappropriate connections, e.g. the event created by interaction with a tangible object to a specific action. Wiring up the action with the code can be done graphically by dragging a line from the newly created action element in the right frame to the small circle next to the code. Circles are being displayed next to the starting point of classes and methods. An unfilled circle indicates that the method is not yet connected to an action element. After connecting, the circle displays a small green filled circle inside (see Figure 4).

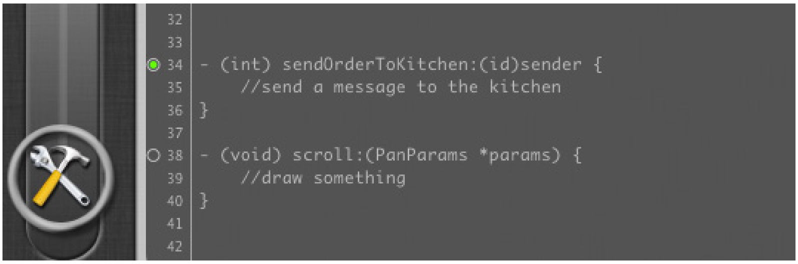


Fig. 4. Detail of the code layer: Code methods are tagged with a small circle. If connected to an action element, which makes them available to the GUI and post-WIMP layer, they also display a filled green circle within.

Creating GUI and post-WIMP components works similarly. In order to create e.g. a custom button, C switches to the layout mode where existing component elements are listed in the right frame. Along with giving a name and description, property access rights can be granted to the GUI and post-WIMP designer for every component and parameter individually in the parameters mode. For instance, the color and size of the button can be made editable, while the label can remain fixed. These access rights are not enforced in the tool, they serve rather as a reminder that an author did not deem it sensible to make certain alterations. Thus, by changing layers an author is able to override restrictions.

GUI Layer. Graphical interface designers do not always have broad programming skills and may wish to focus on the design and layout of the screens. It is their job to put together standard components (screens, buttons, labels etc.) or specific custom components created in the code layer. Furthermore, certain code actions or screen transitions can be connected to them. As the GUI designers are usually not aware of the source code, they need an abstraction of the methods available to connect to.

We conceived the GUI layer to resemble already existing authoring tools where the creation of the application screens is being done graphically via drag and drop. In order to connect certain components to code methods, the action components come into play. They can be connected to the WIMP components via drag and drop as well, eliminating the need for the GUI designer to modify any of the source code.

In our scenario, graphical interface designer G creates a screen with two boxes and several buttons by dragging these components from the right frame to the canvas in the layout mode. The components can be rearranged and edited according to the active editor mode selected on the top slider: To change text labels or images, the author has to switch to content mode. Changes in size, shape or color can be performed in the parameters mode, according to the access rights granted to the components in the code layer. Actions and transitions between screens can be attached to components like buttons or table cells in the behavior mode. To connect the “Send order” method to a button, G drags a line from the associated action element on the right. When hovering over the elements on the canvas, those components potentially able to connect to an action are highlighted by rendering them in green color. Alternatively, a component can be selected by clicking first, then the right frame will only display the methods available for the specific component.

The transition from GUI layer to code layer is designed to allow for inspecting the connections from the user interface to the code. The amount of the information shown from the code layer increases as the user moves towards that layer. We illustrate this transition with our use case in the following.

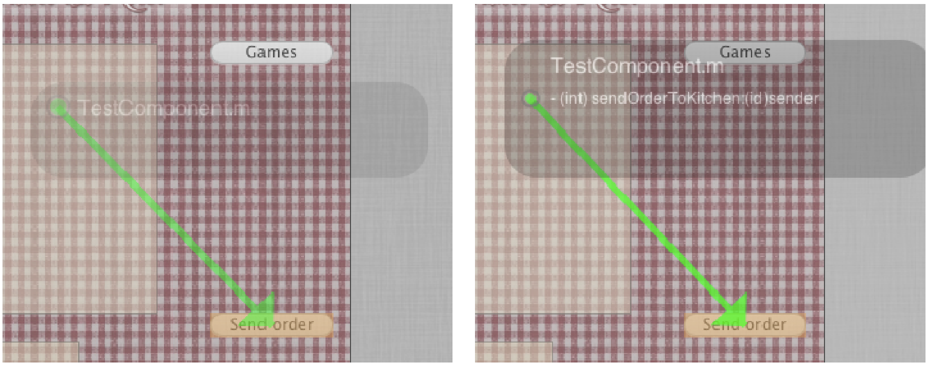


Fig. 5. Two phases of the transition from GUI layer to code layer: A small pop-up first displays the source file only (left image), then also the method the selected GUI button is connected to (right image)

As soon as G selects the mentioned button, a small number appears on the code icon on the slider to indicate how many connections to the code the button has, in this case a “1” for the connected “Send order” method. Thus G can instantly see whether or not there is code layer information attached for further inspection. When moving the slider towards the code layer, a small pop-up appears next to the button showing the name of the source file containing the connected method. When moving further to the code layer, the pop-up displays the method name (see Figure 5), and in a next step the method source code. When the slider has reached the code layer, the complete source file containing the “Send order” method is being displayed.

Post-WIMP Layer. Post-WIMP designers complement the screens designed by the GUI designers with post-WIMP functionality like touch gestures, voice input or tangible object support. Such designers might be experts in usability, but maybe not know much about coding. We provide a visual authoring tool similar to the tools already available for graphically developing WIMP interfaces, where designers simply drag and drop post-WIMP elements, combine them or adjust their parameters. However, while classic WIMP UIs need to cope with visible input and output channels only (buttons, text fields etc.), post-WIMP interactions do not always have an instant graphic representation. While a button is visible and known to be clickable, the area for e.g. a swipe gesture does not need to have a perceivable boundary. Without a graphical representation, it is hard to detect which post-WIMP interaction is offered by an application.

Our goal was to create a post-WIMP layer that closely follows the concept of the GUI layer, i.e. it enables the author to graphically create input alternatives and connect them to code actions without the need to change the source code. In order to achieve this, we conceived possibilities to graphically assign post-WIMP functionality to application screen elements. The process of creating and inspecting post-WIMP elements will be further elaborated in the section "Authoring of Post-WIMP Interactions".

User Layer. Developers may want to take a look at the application screens they are currently working on without distracting connection symbols or tooltips and without having to compile the application first. Methods for user-centered design call for customers to be able to influence the development process from time to time without having to work with complex tools and risking to damage anything. That is why we included the user layer where screens are displayed the way they would look in the application when run on the target device. Here, editing is restricted to changes of the appearance or the content like exchanging texts or images, changing of colors and fonts, or rearranging of the components present. Thus, this layer may not only help developers during the authoring process, but also serves as a presentation mode for meetings with the clients.

In CLAY, the user layer consists of a big canvas showing the application's screens. When in behavior mode, the application can be inspected by moving the mouse over the screens and their elements. Whenever actions are connected to a component, they will be represented graphically. In our example, restaurant owner R may, for instance, adjust the colors to match his restaurant's interior in the parameters mode or change the texts of the food menu in the content mode. The goal is to include the customers in an iterative development process, enabling them to contribute their feedback more directly and allowing them to make changes to the application without the need to have to rely on the other authors.

3.3 Authoring of Post-WIMP Interactions

Standard authoring tools today are not designed for graphical authoring of post-WIMP interaction, but instead require programming knowledge to enhance their applications with post-WIMP features (see section 2). Our concept of cross layer authoring strives toward enabling the author of each layer to work independently, e.g. a post-WIMP designer does not necessarily need to perform changes on the code layer. In this section, we illustrate our approach for graphically creating, combining and validating post-WIMP elements in this section by exemplarily describing some use cases of our previously introduced scenario.

Creating Post-WIMP Interactions. After the graphical UI of the application has been designed and developed, it is the task of post-WIMP designer P to enhance the interface with post-WIMP functionality. The application screen shows a box where a list of dishes is being displayed and a smaller box underneath. P wants to enable the user to scroll through the food menu by panning over the smaller box from left to right. In the layout mode, he drags a pan gesture onto the canvas resulting in a small translucent gray box containing the respective element symbol. As further configuration of the element is necessary, it is marked by a small yellow warning sign.

In order to connect a post-WIMP interaction to an application screen, the author has to define where the interaction can be conducted and which area, if any, is being affected by this action. For this, we conceived a visualization of the activity and output areas that can be defined for every post-WIMP interaction. Activity area in this context denotes the region where e.g. a touch gesture is being executed, while the optional output area denotes the region that responds to that gesture. After defining these areas, they are graphically highlighted to make it easier for the author to inspect the post-WIMP elements they created. Post-WIMP interactions do not necessarily affect elements of the user interface, they might also affect the application logic. In this case, a dedicated area is provided in the authoring tool that serves as a representation of the output area.

In our example, post-WIMP designer P wants the user to pan inside the box at the bottom, so P assigns the pan's activity area to that box by dragging from the pan symbol to the box while holding down the left mouse key. While dragging, a red arrow is being displayed pointing away from the post-WIMP element towards the cursor. After connecting, the bottom box shows a red overlay with a preview of the pan gesture, i.e. a small white circle, indicating a finger, moving up (see Figure 6).

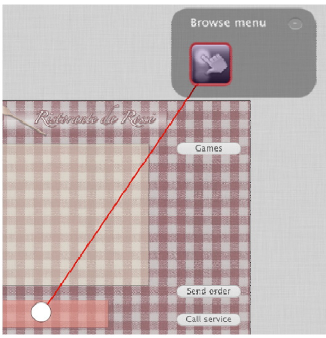


Fig. 6. Detail of the post-WIMP layer: A newly created pan gesture is being displayed in a small box. After connecting the pan to the area where it can be executed ("activity area"), the respective region shows a red overlay. A red frame around the pan symbol indicates that it has been connected to an activity area.

The pan gesture created by P provides default values containing the direction, the length and the duration of the pan, which were defined in the code layer. As coder C has made these values changeable, P can adjust them to his needs in the parameters mode, e.g. change the angle to make it a horizontal pan instead of a vertical one. The changes are instantly being reflected in the preview on the activity area.

Next, P defines the output area for the pan. In this example, he wants the box at the top to be scrolled whenever the user carries out a pan in the lower box. Assigning the output area works similarly to defining the activity area, except for now, P has to press the right mouse key and the displayed arrow and the accentuation of the components are painted blue. After connecting, the upper box has a blue overlay. A bicolored frame around the post-WIMP symbol indicates that it is connected to both an activity and an output area (see Figure 7).

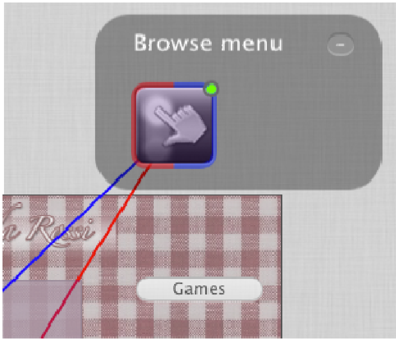


Fig. 7. Detail of the post-WIMP layer: The bicolored frame around the pan gesture symbol indicates that its activity and output areas are assigned. The green filled circle shows it is connected to a code method.

Connecting post-WIMP elements to code actions can be done in the same manner as in the GUI layer. When dragging the scrolling action to the newly created post-WIMP element, it is highlighted green. After successfully connecting the action, a small green filled circle is displayed in the upper right corner of the pan symbol (see Figure 7). As scrolling is a standard action in the tool, it also has a graphical representation, which is immediately visible on the output area overlay. Parameters like scrolling direction or speed can be set in the parameters mode along the lines of the configuration of the pan gesture.

Combined and Alternative Post-WIMP Elements. In addition to adding single post-WIMP elements to application screens, it is also possible to graphically combine several elements to one larger element group or to offer alternative input methods. This is a novel approach to deal with complex post-WIMP input, as it can be connected to a code action without the code layer having to know what exactly triggered the action.

In our next use case, P wants to add two alternative possibilities to order a meal on the food menu screen: the first is to perform a long press on the desired dish and say

"Order" at the same time, the second is to perform a long press on the designated food menu item and then swipe to the right. To achieve this, P drags a long press gesture to the canvas, and then adds a voice input element by dropping it onto the long press box. Both elements are now arranged next to one another, indicating that they have to be performed consecutively. After defining the activity areas, configuring the elements in the parameters mode (e.g. setting the voice input to the word "Order") and wiring up the post-WIMP group with the corresponding code action, the setup of the first input alternative is completed.

To implement the second input method, P adds a swipe gesture to the group, which should be performed alternatively to the voice input. To achieve this, P drops the swipe element directly onto the voice input symbol, which results in the swipe being displayed above the voice input. This indicates that either both actions have to be performed at the same time or alternatively. The behavior for these kinds of stacked elements can be switched by clicking on the label appearing above the elements, reading either "Together" or "Choice" (see Figure 8). Further configuration concerning e.g. the time interval between the long press and the swipe or voice input can be configured in the parameters mode. No changes have to be made in the behavior mode as the post-WIMP group has already been connected to a code action that will be triggered whenever one of the alternative chains of post-WIMP interactions has been performed.

Fig. 8. Detail of the post-WIMP layer: More complex post-WIMP interactions can be created by combining single gestures or events. In this case, a long press combined with either a swipe or voice input form a combined post-WIMP element, which can be connected to an action.

3.4 Visual Validation

In order to be able to inspect the post-WIMP interactions, which have been added to an application, we conceived what we call visual validation. It is meant to visualize the sequence of a specific post-WIMP interaction (or group) with their respective activities and output areas. This enables the author to test the course of the post-WIMP interactions without having to compile the application first, thus targeting on a reduction of the time needed for testing and adjusting the interaction parameters [7].

In our CLAY prototype, visual validation can be performed in the content mode. When switching to that mode, all post-WIMP boxes turn into little players with a timeline and a play button (see Figure 9). Pressing play starts a real-time animation of the post-WIMP elements on the screen they are connected to. Touch gestures are once more previewed by a graphical representation of the user's finger moving over the activity area, standard output elements are being represented by e.g. an arrow moving along the scroll direction. This is especially useful when several elements have been combined to one more complex element group, as their interplay can be inspected chronologically. In case a post-WIMP element directly triggers a code action, this is shown by an arrow directed towards an application logic icon, which then appears in the bottom right corner of the canvas. Having these players in the content mode aims at making it easy to inspect the temporal sequence of the post-WIMP elements and check the result at a glance. Additionally, the players can be organized in different groups, which should lead to more clarity on the screen. These groups can be added, shown and hidden in the right frame when in content mode.

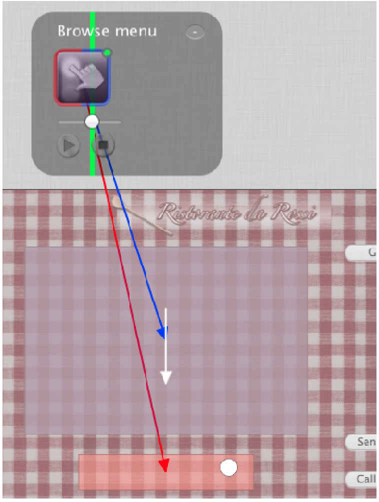


Fig. 9. On the post-WIMP layer, visual validation of e.g. a pan gesture can be performed in the content mode. Pressing the play button in the "Browse menu" box starts a real-time animation of the post-WIMP elements on the screen they are connected to.

4 User Feedback and Discussion

We evaluated our concept by presenting the prototype to three computer scientists and two media designer to gather qualitative feedback. All participants were between 20 and 40 years of age and had at least fundamental knowledge and experience regarding the design and development of post-WIMP applications.

The organization of the tool in four author layers was generally appreciated and found to be valuable, especially for large projects, though there was disagreement concerning the need for a user layer. While one subject found this layer particularly useful to include customers into the authoring process, another participant stated that customers might have no interest in actively participating in the development.

The division into four editor modes was regarded as useful. The fact that, for instance, the content mode offers rather different functionality on each layer was not regarded as confusing.

The concept of fading between layers using a slider instead of switching discretely was generally appreciated and regarded as enjoyable. Still, some participants pointed out the importance of the possibility to switch between layers faster, especially when using the tool for a longer time.

Creation of component and action elements in the code layer and usage of those elements in the upper layers was considered easy and intuitive. However, all subjects missed a possibility for the interface and post-WIMP designers to predefine custom components and actions they need the programmer to implement. In the CLAY concept, applications are built from the code layer to the upper layers, according to specifications negotiated between the authors. Future work might engage in a concept to allow the definition of custom components or actions not only in the code layer, but also in the GUI and post-WIMP layers.

It was generally considered problematic to create large applications with the tool as it was expected to become gradually overloaded. Nevertheless, our concept offers several approaches to handle complex applications. For instance, the author has the possibility to display connections between layers for every component individually and on demand. Post-WIMP element groups can be minimized to take up less screen space and organized hierarchically in groups, which might lead to a less overloaded screen. When inspecting several connections to the code layer at once, the author can at some point choose which of the boxes to expand in order to see the full source code of only the specific source file of interest.

The visual creation and validation of post-WIMP interactions was considered innovative and helpful, though some explanations had to be given concerning the setup of the interactions, alluding to the fact that it is not an intuitive concept which would benefit from explanatory tool tips.

In conclusion, all participants stated that they generally found the CLAY concept to be an improvement of the authoring process of post-WIMP applications and would consider working with a tool based on that concept in the future.

5 Conclusion

We presented a concept called cross layer authoring for a tool enabling the collaborative development of post-WIMP applications for several different author groups. Layers in this context denote the contribution of the particular author groups to the developed application. We identify four different layers, namely the code, GUI, post-WIMP and user layer, each corresponding to an author role. Furthermore, we split up the tasks for each of these author roles into the four editor modes content, layout, parameters and behavior. We suggest organizing cross layer authoring tools in form of a matrix, with one dimension serving to distinguish the author layers, and the other dimension used to switch between editor modes, which received consistently positive feedback in our user interviews. Our goal is to provide a tool enabling authors to work on their respective tasks independently. In order to allow UI designers to work autonomously, we include possibilities to graphically create the GUI and post-WIMP functionality. A concept for visually enhancing user interfaces with post-WIMP input methods has been introduced. Furthermore, the graphical creation and inspection of combined post-WIMP interactions has been presented and found to be innovative and useful by the users. In order to enable connections between layers, we introduce the concept of abstract connection hooks. These connections can be graphically inspected by traversing the layers. We propose a seamless transition between layers, which we realized in our prototype CLAY with a slider as a tool for navigation which users found easy and enjoyable to use.

Future work might engage in a concept to soften the hierarchy of the layers, allowing GUI and post-WIMP designers to create custom component or action elements they need the programmer to implement. This might lead to a more bidirectional communication between layers and thus encourage the concurrency of the development process, which might result in a reduction of the development time.

Acknowledgements. This research work has been financially supported by the German Federal Ministry of Education and Research (BMBF-FHProfUnt grant no. 17043X10).

References

1. Bastide, R., Navarre, D., Palanque, P.A.: Model-Based Tool for Interactive Prototyping of Highly Interactive Applications. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2002, pp. 516–517. ACM, New York (2002)
2. Beaudouin-Lafon, M.: Instrumental Interaction: an Interaction Model for Designing Post-WIMP User Interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 446–453. ACM, New York (2000)
3. Hansen, T.E., Hourcade, J.P., Virbel, M., Patali, S., Serra, T.: PyMT: a Post-WIMP Multi-Touch User Interface Toolkit. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS 2009, pp. 17–24. ACM, New York (2009)
4. Huot, S., Dumas, C., Dragicevic, P., Fekete, J.-D., Hégron, G.: The MaggLite Post-WIMP Toolkit: Draw it, Connect it and Run it. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, UIST 2004, pp. 257–266. ACM, New York (2004)
5. Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-Based Interaction: a Framework for Post-WIMP Interfaces. In: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 201–210. ACM, New York (2008)
6. Kaltenbrunner, M.: reacTIVision and TUIO: a Tangible Tabletop Toolkit. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS 2009, pp. 9–16. ACM, New York (2009)

7. Khandkar, S.H., Sohan, S.M., Sillito, J., Maurer, F.: Tool Support for Testing Complex Multi-Touch Gestures. In: ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010, pp. 59–68. ACM, New York (2010)
8. Kin, K., Hartmann, B., DeRose, T., Agrawala, M.: Proton: Multitouch Gestures as Regular Expressions. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI 2012, pp. 2885–2894. ACM, New York (2012)
9. Klemmer, S.R., Li, J., Lin, J., Landay, J.A.: Papier-Mache: Toolkit Support for Tangible Input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 399–406. ACM, New York (2004)
10. König, W.A., Rädle, R., Reiterer, H.: Squidy: a Zoomable Design Environment for Natural User Interfaces. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA 2009, pp. 4561–4566. ACM, New York (2009)
11. Lawson, J.-Y.L., Coterot, M., Carincotte, C., Macq, B.: Component-Based High Fidelity Interactive Prototyping of Post-WIMP Interactions. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI 2010, pp. 47:1–47:4. ACM, New York (2010)
12. MacIntyre, B., Gandy, M., Dow, S., Bolter, J.D.: DART: a Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, UIST 2004, pp. 197–206. ACM, New York (2004)
13. Serrano, M., Nigay, L., Lawson, J.-Y.L., Ramsay, A., Murray-Smith, R., Denef, S.: The Openinterface Framework: a Tool for Multimodal Interaction. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2008, pp. 3501–3506. ACM, New York (2008)
14. van Dam, A.: Post-WIMP User Interfaces. Commun. ACM 40(2), 63–67 (1997)
15. Hartmann, B., Abdulla, L., Mittal, M., Klemmer, S.R.: Authoring Sensor-based Interactions by Demonstration with Direct Manipulation and Pattern Recognition. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 145–154. ACM, New York (2007)
16. Hartmann, B., Yu, L., Allison, A., Yang, Y., Klemmer, S.R.: Design as Exploration: Creating Interface Alternatives Through Parallel Authoring and Runtime Tuning. In: Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST 2008, pp. 91–100. ACM, New York (2008)
17. Simpson, J., Terry, M.: Embedding Interface Sketches in Code. In: Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology, UIST 2011 Adjunct, pp. 91–92. ACM, New York (2011)
18. Newman, M.W., Lin, J., Hong, J.I., Landay, J.A.: DENIM: An Informal Web Site Design Tool Inspired by Observations of Practice. Hum. Comput. Interact. 18(3), 259–324 (2003)

Existing but Not Explicit - The User Perspective in Scrum Projects in Practice

Åsa Cajander[1], Marta Larusdottir[2], and Jan Gulliksen[3]

[1] Uppsala University, Box 337, SE-751 05 Uppsala
[2] Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland
[3] KTH – Royal Institute of Technology, Lindstedsvägen 3, 10044 Stockholm
asa.cajander@it.uu.se, marta@ru.is, gulliksen@kth.se

Abstract. Agile software development processes are becoming more common, but this does not mean that the user perspective in the development is catered for. It has its challenges to integrate the users' aspects in Scrum projects in practice. In order to better understand these challenges we have interviewed IT professionals using Scrum focusing on four different areas: responsibility for the user perspective, emphasis on usability and user experience through documentation, usability activities with users and the organisational and contextual settings for emphasizing the user perspective. Results show that the responsibility for the user perspective is unclear in Scrum projects, and that often the user perspective is neither discussed nor described in the projects. However, the user perspective is often present through informal feedback used to understand the context of use and inform design for example. Finally the paper presents implications for working with the user perspective in Scrum projects.

Keywords: Usability, user experience, user perspective, responsibility, agile software development, Scrum.

1 Introduction

Systems development processes and methods used in industry vary over time, and historically different value sets have guided the development of processes for IT projects. The purpose of a process is often to provide structure, predictiveness and quality to software development practices and to be a container of the collective knowledge and experiences of software development work within a particular context [1]. Since the 1990s, agile processes, and especially Scrum, have become more popular [2; 3] as a reaction to the size and complexity of the presiding processes, such as RUP or the waterfall model. Scrum as a development methodology has received increasing attention over the last years, mainly due to its agile nature and openness for flexibility, and for the skills and driving forces of the team members. One of the basic values of Scrum is speed and communication [4]. However, the requirements on quality, usability and other aspects of great importance for the users are not explicit in the process.

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 762–779, 2013.

Many software development practitioners in industry regard Scrum as a user-centric process, for example by introducing user involvement through user stories, and by its iterative and communicative nature [5]. This aspect of Scrum would then concur with the values of many user centred design approaches, and this may then explain why Scrum has become so successful. However, the user perspective is not a mandatory part of the process and not something that can be taken for granted while applying Scrum in software projects. Even though many development organisations are describing great success of using agile development processes, none of these processes explicitly describe that usability activities should be included in the process [6] and Scrum has particularly been criticized for not sufficiently including the user perspective in the process [7]. One of the challenges mentioned by IT professional who are including the user perspective in software development is that it is hard to find time for usability activities such as user centred evaluation [8]. Additionally, it is challenging for the IT professionals to maintain an overview of the total user experience of the product in Scrum projects. It has been suggested that sharing documents, artefacts and particularly knowledge between the development team and the usability specialists is one way of maintaining the overview or the big picture of the user perspective [9]. Moreover, it has been suggested that having a usability knowledgeable person in the teams or at least that more face-to-face communication happened between the team members and the usability specialists [10].

Previously we have studied the application of usability evaluation methods in Scrum projects in practice [11]. From this study we concluded that many informants gather informal feedback on their design using qualitative evaluation and only very few conduct quantitative evaluation by asking more than 10 users to participate in thorough user evaluation. Typically thorough usability evaluation is conducted as seldom as twice a year, often by contracting an external usability expert [12]. The main reason for usability evaluation not being conducted more frequently is time constraints. However, some usability activities, such as workshops, are frequently used in Scrum projects [13]. These activities are often informal, ie do not include measurements or rigorous usability evaluations, and therefore these fit better to the fundamental principles of Scrum, which are speed and communication. Additionally producing incremental deliverables in short project periods is another popular and important Scrum feature. Another challenge experienced by the interviewed usability specialist in the Scrum projects was maintaining the overall vision of the user perspective, despite of the Scrum tradition of slicing projects in smaller parts.

In this paper we will further explore how the user perspective is affected by the fact that our informants are using the Scrum process in practice. We have chosen to address the occurrence of a user perspective in four important areas: responsibility for the user perspective in Scrum projects, activities including the user perspective in the development, emphasis on usability and the user experience through documentation, and the organisational and contextual settings for emphasizing the user perspective. Through such an analysis we aim to understand the need for improving the Scrum process to facilitate the inclusion of a user perspective more explicitly in the daily work of the IT professionals.

2 Background

One important task in planning your software development project is defining the responsibility for particular activities. Responsibility here may refer either to the state of having a duty to deal with something, or the state of being accountable or to blame for something [14]. This can be seen as either a rule based view of responsibility, or a consequence based view, as in [15]. The notion of responsibility for usability is closely related to discussions of responsibility generally in social science in relation to groups. Here phenomena such as "the diffusion of responsibility" and the notion of "somebody else's problem" are interesting to investigate. Diffusion of responsibility is a social phenomenon, which might occur, in larger groups, where no one in the group takes responsibility for phenomena. When a task is placed before a group of people, there is a tendency for each individual to assume someone else will take responsibility for it—so no one does. This is a negative outcome that may occur in groups where the responsibility is not clearly assigned. Previous research in the area has indicated that the diffusion of responsibility may have negative effects in systems development in practice [15].

Stating goals for usability or user experience is one way of describing the emphasis on the user perspective explicitly and giving the IT professionals some motivation for including the user perspective in their work. Stating usability goals has been highly rated by IT professionals as a way to integrate usability in software development work [16]. Identifying the usability requirements of a system as accurately as possible has been pointed out being a major challenge in agile projects, since it involves the customers rather than the actual users [6]. The results from a recent study show that stating usability goals in Scrum projects was highly rated by the participants, and still less than half of them did explicitly state usability goals [13]. Researchers have suggested that usability goals or issues could be combined with other existing documentation in Scrum, such as the description of the goals as acceptance criteria for the user stories [7; 17] and capturing usability goals in user stories or in the product backlog [18].

Recent results show that many IT professionals frequently conduct some kind of usability activities in Scrum projects, and that they generally rate the activities as being useful [13]. The highest rated usability activities according to this study were workshops, informal usability evaluation with users and meetings with users. It is noticeable that all of these activities are rather informal. Similar results have been presented in a study showing that all kinds of prototypes are used more frequently in Scrum projects than when using other software development processes [2]. A recent extensive literature study shows that the most common usability activities in agile development are low fidelity prototypes, user testing aimed at refining the prototypes in the next iteration and inspections [17].

Organisational and contextual settings for integrating user centred design into agile software development successfully have been explored in several research studies. Close collaboration between the development team and the usability specialist has been considered as one of the biggest success factors for integrating the user perspective in Scrum projects [10; 17]. The usability specialists' understanding of their job role and the need to establish, protect and communicate an overall team vision was pointed out as the two major themes highly important for the success of integrating user activities in agile development [19]. Often user experience issues are considered important both on strategic and operational level, but the current work processes and management styles can limit the impact of the usability specialists [20].

3 Method

The 21 informants were found through personal contacts, a presentation at an HCI interest group, social media and suggestions from informants already interviewed. The informants who participated in the study all worked with Scrum and were interested in integrating usability activities into the Scrum process.

In total there were 7 females and 14 males, their age reached from 29 to 55, and they had been working from 1 to 15 years in the industry. Their education varied from having no university education to having a PhD degree. When our informants were asked about their roles we got various answers such as: usability designer, user experience manager, usability specialist, application designer, interaction designer, senior system developer, user interface developer, usability specialist, business architect, sales person, developer, tester, Scrum Master and Product Owner. Some of the informants stated that they had two or three roles. The informants worked at 14 companies in various organisational contexts. The main types of organizations were product development and consulting companies. Some of the companies were international, having employees worldwide. The number of employees reached from 8 to 12.500. Generally, the Scrum process used had been adapted to the context of the organisation. Some companies had combined Scrum with features from other systems development processes such as Kanban and Extreme Programming. We asked the informants about their work in one specific project since we wanted to understand the context of their experiences, and we did not want them to describe several anecdotal things from a number of project. We also asked them to tell about a project that they had been working on lately so that they would have it fresh in memory.

The data gathering method used for our study was semi-structured interviews. An interview template was developed and the questions were adapted in accordance with the organizational role of each informant, their background, and their experiences working with Scrum and usability. The interview template included the following topics: Background including role in company, responsibilities and education; experience from using Scrum in one particular project; their activities for involving users in that project; their view on responsibility for usability and other remarks. The interviews lasted for about one hour. The interviews were recorded and detailed notes on paper were taken. Most interviews were conducted by two researchers interviewing one person, one acting as a conductor and the other as a note taker. All recorded interviews were transcribed verbatim, most of them by an external consultant. The quotations provided in the text are not always verbatim, but sometimes slightly rephrased in order to be more readable and representative. We refer to our informants as males despite their real gender when describing their comments.

Data from different interviews were compiled and iteratively analysed by two researchers. During a thematic analysis [21], data was reviewed, organised and read through to identify themes. At this phase, mind maps were used. Data was then reviewed again to iteratively develop these themes and to categorize the statements of the respondents according to the themes in a software program for data analysis called Atlas.ti[1]. Some themes represent a set of inductive constructs while others are rather a set of predefined ideas. At this stage all the authors made interpretations and discussions included search for alternative understandings and interpretations. Finally, the different themes were exported from the software tool in order to have an overview. This overview constituted the basis for the written text presented in this article.

In the analysis of the interviews we chose to group the roles of the informants in four categories:

1. **Scrum manager role –** Three informants had the *Scrum Masters role*, which are responsible for the project management of the Scrum team or the *Product Owner role*, which is responsible of specifying what is needed in the software from the customers' viewpoint and prioritize these needs as described by Schwaber and Beedle [22].
2. **Team member role –** Nine informants were working in one team and being responsible for delivering the outcome of the sprint as described by Schwaber and Beedle [22]. Their main activities were to design, develop and test the user interface.
3. **Usability specialist role** – Five informants had extensive knowledge of HCI. Typically, they were not members of one particular development team, but supported many teams. Their main job tasks were requirements analysis, interaction design and evaluation. They contribute to the project both during pre-studies and support the team members during the actual development. This role description is based on the Usability Engineer role as described by Mayhew [23].
4. **Business specialist role –** the main job task of four informants was to analyse the requirements of the software during pre-studies. Typically, they were not members of one particular development team, because their main job task was done in the process of forming the teams. Typically these informants did not have as extensive knowledge of HCI as the usability specialists.

4 Results

Here the results will be presented according to the four focus areas in the paper.

4.1 Responsibility for the User Perspective

During the interview sessions all informants were asked about their opinion on who is responsible for the usability of the system or product. All informants gave one single

[1] For further information see `http://www.atlasti.com/`

answer to the question. Some of the informants see responsibility as pointing out the person to blame when things go wrong, whereas others answer from the perspective of who has the duty to deal with usability activities. One of the usability specialist is worried that he might be seen as the one who is to blame since he perceives that the usability of the product they are selling is not really good. He discusses if the others will put the blame on him: *"I do not think that anyone will come and chop my head off - yes actually it worries me a little bit because from a usability perspective it is my product - I am the only usability expert here, so if it lacks the usability it is my fault."*

It is noticeable that the answers to the questions were quite diverse and distributed between four different ways of looking at the responsibility: Six informants state that the usability specialist was responsible, some state that no one (three informants), the whole team (six informants) and the Scrum managers (six informants), are responsible. One of the three interviewees that described that no one was responsible describes that dismissing usability work and the responsibility for the product to be usable was a deliberate choice in the Scrum project due to time pressure: *"Actually no one is responsible. I tried to take that role and tried to integrate it in those parts that I was responsible for. But it was the first thing that disappeared when time became short. You won't have the time to do this. You have to program!"*

Some results indicate that the responsibility for usability is not well defined in Scrum in practice, and that "diffusion of responsibility" and "someone else's problem" has occurred in some of the teams. One informant sums up the situation like this after having thought about the responsibility: "*I don't know who was responsible, that's the answer"*. Another informant describes the lack of responsibility as a problem: *"There is no one responsible for the actual full user experience. That's the problem."*

Table 1. Overview of answers to the question: Who is responsible for the product to be usable??

Answer / Answers from:	The Scrum managers are responsible	The whole team is responsible	The usability specialist is responsible	No one is responsible
Scrum managers	3 Scrum managers			
Team members	2 team members	3 team members	1 team member	3 team members
Usability spec*		1 usability spec*	4 usability spec*	
Business spec*	1 business spec*	2 business spec*	1 business spec*	
Total number	6 informants	6 informants	6 informants	3 informants

* spec = specialist.

One interesting finding is that four out of five of the usability specialists perceived that the responsibility for usability was theirs in their capacity as usability specialists, and at the same time all the Scrum managers perceived that the usability of the product was the Scrum manager's responsibility. This variety of answers might come as a result from the two different views of responsibility as either something you have the duty to deal with (as a usability specialist you are to deal with usability), and the overall responsibility for the quality of the product from a usability perspective (some might say that the quality of the final product is the manager's responsibility). One of the Scrum managers integrates both views of responsibility in his answer, and he describes that he is responsible or to blame for the usability of the final product, but that it is every ones responsibility to work with usability in the project: "*I of course know that I have the implicit responsibility due to the title I have and the things that I do in the project, people in the project certainly rely on me on making this product highly usable, but I can certainly see that everyone is working towards that goal*". These two views of responsibility are also present in the answers from two of the usability specialists who have said that they are responsible themselves, but also that the Product Owner is responsible in some way.

Furthermore, six of the informants perceive that the responsibility was theirs as a member of a team. This sums up to be 13 out of 21 informants who feel that they are responsible for usability in their role. This result is really interesting as the informants who took part in the interview study were chosen based on the fact that they worked with usability in Scrum, and even so 8 out of 21 answered that they were not responsible for usability from any perspective. Another interesting finding is that the team members all give different answers to the question of who is responsible for usability, and that their answers are distributed in different categories in Table 1. The answers of the business specialists vary in a similar way but none of them thinks that no one is responsible.

4.2 The User Perspective in Documentation

The informants described several different approaches on how to include the user perspective in written text in the Scrum projects. There are mainly four categories of answers in the interviews: 1) there is no text on the user perspective; 2) a broad and vague description of the user perspective is used; 3) a detailed text on the user perspective is written; and 4) the user perspective is defined by a standard.

Some of the informants described that they did not describe the user perspective at all in their software development (category 1), or any kind of usability specifications. The reason given for this is that the user perspective and usability are difficult to state in words. One of the informants expressed that usability is more a general quality and a feeling not possible to define, as in this quotation: *"sometimes it's like this, you have to feel it to know"*. One informant mentions that the usability specialist in their company must check all the functionality before it can be delivered, that is their way of including the user perspective in their development process, but they do not state any requirements for the usability of the product. Hence, no text or documentation is used to ensure the user perspective. Some of the informants explain the lack of the user perspective in Scrum documentation by saying that it would be of no use in the Scrum context. Documentation used in Scrum generally describes some small aspects of the system, and these generally do not describe the overall picture which is necessary for user perspective. There are generally no such overarching descriptions used in Scrum, according to the informant. In some companies the introduction of Scrum has changed the process of handling the user perspective and documentation generally, which he sees as very positive, and *"now it feels like we don't seem to do any specifications any more. It is so much more dynamic since they started with Scrum."*

About one third of the informants mention that their companies have written text on the user perspective in the text on non-functional requirements, and generally these requirements are not very strictly defined (category 2). The functional requirements, on the other hand, are strictly defined with levels of importance for the implementation. Some of the levels described were: *"must, if there is time, and if we have time, maybe we will do it"* but these are not applied for defining usability requirements. One informant describes how the user perspective often becomes very vague by explaining how it is included in workshops with users: *"It is very much up to the team in the workshops trying to define what is the soft parts of the products and it mostly comes like, this old system works like this and we don't like that"*. Some companies describe the user perspective in a document depicting a vision for the software system. However, it should be noted that having the user perspective in the vision document means that it is disconnected from the testing and evaluation of the system, as the vision document describes an idea or hope of how the system should work and be in the future. At the same time, including the usability goals in the vision documents helps when viewing the system as a whole which is good as the 'eagle view' of the system is described to be missing in many Scrum projects.

In a few cases there is precise description of the user perspective in documentation used in the projects (category 3). One informant explains: *"You're doing a login page then that login page should be easy and secure, and you have the corresponding measurement for that, what is the users satisfaction as opposed to a scale from one to seven, [...] so it's measured in that way"*. An interesting observation is that even though these informants did state a precise description of the user perspective they did not evaluate against these usability goal in a quantitative way later in in the process.

In two companies a standard is used to define the usability of the product (category 4). Both these products are safety critical. One of the informants explains: *"So even in the standards they promote usability to avoid hazards"*. This informant explains further that this was one of the reasons that the company wanted to hire a usability specialist. Another informant says that at his company, general usability goals are stated with references to standard. He explains: "*the general requirements can be the same for different projects and will be referred to in a general document covering all aspects of it, for example a style guide that contains references to standards"*. It is clear that there is not one way of stating the usability goals in practice, and that the challenges for the usability professionals are numerous.

4.3 Activities with Users

It is described in the interviews that the fundamental value in Scrum is speed, and this affects the way that the IT professionals choose techniques to incorporate the user perspective in the development. Our results show that the Scrum team communicates with a few users in an informal way. Often some prototypes are showed and discussed, or features are presented at a workshop. Prototypes are often showed both to the other team members and to the users. These meetings can be planned in advances, but often they occur when there is a need. Often they are qualitative and the focus is on feedback rather than on quantitative measures.

From a usability perspective it is interesting to look at the user stories since they could be closely related to usability activities and the user perspective in Scrum. A user story is a short text describing the user's interaction with the computer system and it is often used as a way to understand functionality. However, the user stories used by the informants in the study are seldom connected to other usability activities, such as for example interviews with users or field studies and they are often quite technical. One informant describes this as worrisome since the user stories do not help the developers understand the needs of the users. This informant is convinced that there is a need to make other kinds of user stories in the Scrum projects that he works in, and he suggests user scenarios as an avenue forward. Another problematic thing with the user stories mentioned is that the main objective with them is that they define what the different team members are supposed to do during the next day or days. Hence, they are used more for task lists, or to-do description in the projects and they are not really used as a way to understand the users', the users' needs or the user's perspectives. The user stories are more used as a description of who is doing what in the project, and informants use them as a way of remembering and talking about what they are doing in the project. One informant describes how user stories are used in the work of the IT professionals in this way: "they know which story is on-going, and who is working on it, and they have a day-to-day conversation." So the user stories are used to structure the tasks of each team member and for communication of how much is done in the development, and not for describing the user perspective of using the software.

Several of the senior usability specialists describe that they ask the developers what they need when they design the system. Here, the communication of the user perspective is in focus. Then the senior usability specialist gives the developers what they describe they need, and this can be Power Point presentations, documentations, UML diagrams or paper prototypes. One informant describes that: *"I usually try to have a discussion with the developers and ask them what they need to be able to do their job. If they need written requirements, I write it, if they need power point slides with descriptions I do that"*. Hence the communication as such is most important for extending the teams' understanding and motivation for incorporating the user perspective. This way of working also describes that when describing and presenting the users' needs in Scrum, the thoroughness of the information is not as important as providing artefacts to foster communication. When choosing the activities the informants do not talk about which activity would give them the best understanding, or the most correct representation of the user perspective. Instead communication is the core value that informs their choice. The informants also describe that paper prototypes are used in communication with users, and they are chosen as they are quick and easy. Often the feedback given is short and not documented.

One emerging activity with the purpose of including the user perspective in the Scrum projects is the use of user forums or blogs. One of the companies has a blog where they present and discuss their product. This blog has attracted around 500 users of the system who discuss, complain and describe areas of improvement of the product. The usability specialist working in the company describes these forums that emerge in connection to their blogs as a new usability activity that he has chosen to include in his projects as it is a quick and easy way to discuss with users and to understand the users' perspective.

4.4 Organisational and Contextual Settings for the Usability Work

When asked about the organisational and contextual setting for the usability work, and the foundations for the level of emphasis on the user perspective in their projects, we got interesting answers from our informants describing what is important to incorporate a usability perspective in the organisation. Some mention that the development company has to be interested in including the user perspective in the development work, other mention that the managers and the Product Owners' view on the user perspective is crucial and some mention that both the development organisation and the client need to be interested in these aspects. These various views on what is important in the organisational and contextual setting are elaborated below.

Some informants stress the importance of having a usability knowledgeable person in each team who can teach the other team members how to include the user perspective in the development work. One informant describes: *"I think that by having me on the team everyone was constantly reminded of the need for things to be easier and usability. I think that's one of the big advantages of working in a team. You educate the people you work with and you change their attitudes"*. It is also a general pattern in the interviews that informants who work with usability as team members are more satisfied with their work situation. They have the possibility to affect the work in the teams, and the others often accept and appreciate their contribution. The informants who are outside the teams, in a specialist role focused on usability, are more dissatisfied. They are working as consultants for several teams, and they do often not have the status or power to make a true contribution in the development. One of the informants describes his situation: *"I am like an add on"*. It is noticeable that these informants do not have a formal role according to Scrum.

Others explain that the managers in the development company affect the emphasis on the users perspective vastly, their background and education plays an important role for the focus on usability. One informant says: *"There's a huge focus on usability in the entire organization, thanks to things that have happened. One is that we changed our management to someone who's not an economical person but someone who has an industrial designer background. He's kind of changing the focus to take care of the brand and rather than looking at the economic figures each month they look at the long term prospects for the products that we create"*. Another informant describes that the managers in the company emphasise the user perspective because they presume that they will earn money from it. He describes: *"The company has noticed that it makes money. I think that's the main force. When I started three years ago I tried to push for it as well. But they felt that it was a nice thing to have, but it wasn't crucial for sales. They see it as, in our field it has been low prioritized and* *they saw an opportunity there to enhance this focus. We can be market leading within our field. So I think they understand that they can make money of it"*. Even though a few informants describe the management's view of usability as important the majority of the informants did not mention the management when asked about the organisational and contextual setting of usability. This might be due to the fact that there had been a change in management in the organisations who mention this. It is interesting however that the informants who have a manager who is really interested in usability perceive that this strongly affects their work and the usability perspective in the organisation.

Others describe that the Product Owner plays an important role when deciding how much the user perspective should be emphasised. Formally the Product Owner should be responsible for the contact with the customer, but the customer does not need to be the user. One informant describes the different stakeholders' aspects by saying: "*I think that it's really important that the Product Owner has the usability perspective, because no one else at that kind of level has those glasses on. The sales director doesn't, and not the marketing director and not the management either really. So the Product Owner is really concerned about the product and how that will be perceived by the customer and the Product Owner is the one that has the most interaction and contact with the team"*. This informant describes that the communication between the management level and the team regarding usability is important, and that it is the responsibility of the Product Owner. It is however noticeable that this informant presumes that the CEO and the managers generally do not focus on usability, and therefore the Product Owner should have this focus. Another informant describes his experience of co-operating with a Product Owner that did not have a usability focus, and how this affects the usability work in a negative way. This informant describes the qualities of a good Product Owner focusing on usability like this: *"I would prefer another Product Owner, who understands the organization, that understands the task, understands really what the users like and what they need. This person was very good at making decisions, but maybe not from the user's point-of-view."*

Some mentioned that both the client and the development team needed to be interested in the user perspective. In some cases the client has asked for courses to extend their knowledge about usability activities. One informant describes that the client probably learned about the importance of usability the hard way. This informant explains: *"I was asked yesterday to have a lecture to the client on how we do work with usability, and how that can affect their role in ordering IT-systems on a regular basis, so they do have an interest in it and understanding for its importance, and I think they've learned this the hard way as many others have that they've ordered quite a few systems, and they're not satisfied with it"*. Another informant explains the importance of the co-operation between the client and the development team when keeping the user perspective in the development. He describes: *"It's probably because the clients have been interested in usability in combination with that we have been able to convince them that this is a good way of working. I guess it's a combination of both. I mean obviously as a usability consultant you know how you think it should be done. Luckily we have been given quite a lot of freedom in how to do it."* This view of the user perspective as a part of both the client and the development organisation's priorities is a central part of user centred design, and it is interesting to note that very few informants describe this idea in their interviews. It is also interesting to note that the informants with a long education in usability are the ones that mention this perspective in their interviews.

5 Discussion

In this section the different results are interpreted in order to understand the four areas of the article better.

5.1 Responsibility for the User's Perspective

From our previous studies we have seen very different results on who actually shoulder the responsibility, and also different opinions on who should have the responsibility for usability [24]. This study indicates that responsibility for usability is a problematic area also in Scrum projects in practice. The concept of responsibility is not discussed or expressed in the actual projects. However, Scrum projects are probably not worse than other projects regarding the responsibility for usability. Scrum emphasizes productivity and speed in the development projects above other quality features of the system such as usability. Furthermore, there is often no formal role that is given the responsibility for usability in Scrum projects, and this may be a problem as it has a tendency to become "no one's responsibility". However, one may wonder to what extent formal usability responsibility really would affect people in their daily development work and hence influence the quality of the final product.

How can we influence people in Scrum projects to accept responsibility for usability and to work with usability issues? Is it simply to express this need as an activity in the process and then allocate a certain resource to be in charge? Other researchers have answered 'No', since the problem then emerges that responsibility for usability is assigned to someone who does not have the proper knowledge to be able to do a good job [25]. We believe that people must have the right attitude towards usability, a motivation that they really would like to work with usability and an understanding of the role that the user perspective can play in the project if considered properly. Extensive knowledge about, and experience of usability work is also crucial for the responsibility, which is obvious from the study since the vast majority of those who have formal and thorough education believe that they are responsible for usability. People also need to have the skill to be able to shoulder the responsibility for usability, and to work with usability in complex settings. They need to be able to argue and to have the ability to persuade others that usability is important. Hence, we need to prepare them for this during their university education [26].

In this study some of the informants are prepared to shoulder the responsibility for usability, but in the end they are not given the space, mandate or resources to seriously deliver a usable product. This often means that they also need to abandon the responsibility for the user perspective. When you have the mandate and resources to be in charge of usability the possibility to deliver this quality is much more evident and subsequently the likelihood of delivering a successful project from a usability perspective is there.

5.2 Emphasis on the User Perspective in Documentation

The vast majority of respondents do not use precise or written descriptions of the user perspective in their projects. This is interesting since one way to look at maturity in the development process is to consider the teams' use of quantitative goals of usability [27]. These goals should then be used as a target for usability in all usability activities in development and measured continuously to check if these goals have been reached. However, from this study we can see that in the few cases where quantitative usability requirements or goals were used these were not measured in a quantitative way. The goals often got lost in the process, even if they were written at the start of the project.

Scrum is often presented as a dynamic development process in literature [5; 28]. It focuses on oral communication and on collaboration and it does not stress the need for documentation of any kind [29]. Instead it promotes the delivery of functioning programs and program code. In this context it should be noted that addressing the user perspective according to the principles of user centred design and the ISO standard [30] generally means gathering extensive data and documenting your studies to support design decisions in the project. Hence, this fundamental idea of documenting user data to support design is conflicting with the general idea of Scrum, which might give an answer to why the written documentation is often lacking. Moreover, our study shows that the documented goals for usability in Scrum, are often very general, short and mostly based on an oral common understanding. This becomes problematic as it is generally hard to evaluate those goals, and moreover it is not the same person that establishes the usability goals and subsequently evaluates them.

Another interesting finding is that the few organisations that write detailed usability goals do not use those goals when evaluating the product that they deliver. Even though these goals are precise, measurable and a part of the requirements they are not a natural part of the evaluation phase. These detailed usability goals are sometimes used in the design process to inform design. The activity of writing the detailed goals is hence used to inform design, but not to form the evaluation. Evaluations in Scrum projects are made to get informal feedback, to gather the users' opinion about the software, to find bugs and to check functionality [11]. Often the goal with the evaluation and testing is to check if the user stories are sufficiently developed, and not to check if the usability goals have been fulfilled. One of our informants expresses his view on usability goals by saying: "*As soon as something qualitative is written in the usability requirements, they get lost*". So he thinks that the goals are forgotten, after they have been written, and not used while designing and evaluating the user stories.

Another issue that affects the difficulty encountered when integrating usability in Scrum is the focus on things to be done, i.e. delivery of functionality, and an approach where the system to be built is documented in short descriptions, often in user stories. These small descriptions are not really well suited for usability work generally, as it is really difficult to describe general usability aspects of a system in such a way. It is very difficult to describe when something is usable, and especially from the perspective of a small piece of functionality. Usability needs to be addressed on a higher level than possible due to the granularity of the user stories in Scrum, and it includes many aspects that are simply impossible to address when the system is presented as small pieces of a puzzle constituting the system to be built.

5.3 The User's Perspective through Activities with Users

It is clear from the interviews that the values underpinning the choice of usability activities are speed and appropriateness for communication. Perhaps this focus on communication really affects the work with usability, and that talking about the system, its users and different problems relating to the system is really one way to incorporate the user perspective that is not a formal activity. The focus on communication makes it easier to work with more fuzzy quality aspects that are not easily defined in documentation. Oral communication often gives a richer picture and supports the user perspective in that sense. Another aspect when choosing usability activities is that Scrum puts a lot of emphasis on producing new parts of the software through using user stories, and not on redesigning older ones. The value in Scrum seems to be a very action oriented culture with a focus on producing new functionality and not so much on evaluating existing software and redesigning to make the software more usable. Two of the informants describe that they have worked around this in Scrum through rewriting the usability problems into new user stories.

Many researchers seem to envision mature system development projects with documentation regarding the user's perspective, an explicit focus on usability, well defined roles and responsibility with a usability focus [31]. According to this way of thinking the system development projects with oral communication, none existing formal roles and a focus on functionality instead of usability are less mature. This would imply that Scrum would be a less mature process from a usability perspective than its predecessors. However, we would argue that the words mature or immature are quite inappropriate to use, as there is no one best way of organising work in software development projects. Instead the activities need to be carefully chosen and managed given a specific context, as is described by for example Contingency theory and in the work by Morgan [32]. Researchers in organisational theory have discussed the appropriateness of organisational structures in a given context, and we concur with them that the choice of activities is dependent upon the internal and external situation.

Workshops are very commonly used in the Scrum projects [13], and it is especially common in the pre-study phase or in the very beginning of the project. Mostly they are used in order to understand the context of use and to understand the users' needs. However, the workshops are not explicitly connected to usability work in the projects, but are more seen as a way to "capture requirements". This is interesting, as it is obvious that these workshops are one way to incorporate the user perspective or users in Scrum. Hence the workshop as an activity can be seen as a Trojan horse hiding the usability activities in a more popular or attractive form, which could be one avenue forward to include the user perspective through activities with users.

5.4 Organisational and Contextual Settings for Usability Work

This study indicates that experienced and knowledgeable IT professionals may be frustrated about the lack of usability considerations in the agile processes. These professionals constantly strive to work with usability according to books and articles published in the subject [33], however they often feel that they fail miserably. The context of their project often makes it hard to successfully integrate the values and principles of user centred design, as discussed by for example [34]. Moreover, they often feel that the guidelines and methods presented by the research community are not supporting them in their work [35]. Hence, we need to reconsider how to present usability activities so that they are perceived as useful by practitioners.

Experienced usability professionals in the study sometimes find ways around these problem and they dress activities that relate to the user perspective in different clothing. Subsequently usability activities enter the Scrum projects as 'Trojan Horses'. In the interviews it is for example apparent that usability activities are integrated into the Scrum projects in the shape of workshops or informal feedback. This way of integrating usability is also described in other research, by for example [36].

In theory, Scrum contains possibilities for an enhanced user perspective through workshops, meetings with users and user stories. However, a clear usability perspective is needed from the project management as well as the organisational context for successful integration of the user perspective in Scrum.

6 Conclusions

To conclude, and as a take-away message, we would like to present a few recommendations for working with user-centred design in Scrum based projects. These are based on the results of our interviews, and also in relation to our previous studies, in line with the discussion above.

1. **There is no clear picture of the responsibility for usability** – Clearly the emphasis on the user perspective should be strengthened if the responsibility for working with usability is clarified and communicated explicitly. This includes both the aspect of who will work with usability, and who is responsible for the quality of the final product. The question is how it could be done in the context of Scrum, where there are no formal responsibilities for any quality aspects, such as security, privacy and performance. Stating that the responsibility for usability should be shouldered by the entire team does not clearly enough distinguish how the user perspective should be integrated to promote good usability. Perhaps this responsibility could be made more explicit using ideas from for example the concept "privacy by design" where it is still a shared responsibility but more explicitly addressed [38]. Moreover, usability specialists shouldering the responsibility for usability needs contextual and organisational support to be able to make a difference regarding usability. Some examples of organisational support needed are: sufficient mandate, support from management, organisational competence,

self-esteem and experience as well as a good position in the team to be able to contribute to better usability.

2. **Usability goals are unclear** – The question is to what extent clear, measurable quantitative goals are needed to focus development efforts and to drive the development of the product quality. Here we believe that more investigations are needed to find examples where high quality products have been delivered without explicit targets being set. Overall quality goals often serve the purpose better than usability goals, as they help deliver the overall design direction.
3. **The ad hoc nature of user involvement and design feedback –** Agile processes do maybe not support, but also not prevent user involvement. Rather the processes involve users informally. Often this is done by gathering feedback on design in an informal way. Most often this is made in an ad hoc manner and based on the team members' own initiative and knowledge about the user perspective rather than being systematically planned in the Scrum process. We believe that it would be beneficial if the process could be systematised, showing user involvement and design feedback as general activities in the development process.
4. **New methods arise –** The agile requirements on speed and efficiency, focusing on deliverables over extensive documentation and the existing criticism based on lack of user involvement gives incentives to develop new usability methods. The development of user forums to discuss and provide immediate feedback to the developing organization happens because of the goal of agile methods of delivering functioning code as fast as possible. Perhaps incorporating the user perspective more deliberately in the sprint review meetings could be one way, or a new activity checking how the users would use the software in reality, a sort of a "reality check" could be another option.

The user perspective may be considered in Scrum projects, if the team wishes, but there is no explicit support or demands for this in the process, unless the project owner or the development manager has a special interest in usability. Since there is no explicit support for the user perspective in the process itself, much of these aspects often exist through informal communication. A user-centred Scrum project needs to find ways of working with the user perspective within the overall agile nature of the project. Activities addressing the user perspective need to be efficient, with a low degree of overhead and formatively contributing to the product towards the final delivery. To sum up: There are changes that should be made in Scrum projects to better integrate the user perspective, since the user perspective in Scrum is existing but not explicit.

Acknowledgments. We would like to thank COST Action IC0904 called Twintide for the financial support that they have provided. Also, we would like to thank all informants in the study that took their time to participate in the study.

References

1. Boehm, B.: A View of 20th and 21st Century Software Engineering. In: Proceedings of the 28th International Conference on Software Engineering, pp. 12–29. ACM, Shanghai (2006)
2. Larusdottir, M.K., Haraldsdottir, O., Mikkelsen, B.: User involvement in Icelandic Software Industry. In: Proceedings for the 2nd International Workshop on the Interplay Between Usability Evaluation and Software Development - I-Used at the INTERACT Conference 2009, Uppsala, Sweden, pp. 51–52 (2009)
3. Hussain, Z., Slany, W., Holzinger, A.: Current state of agile user-centered design: A survey. In: Holzinger, A., Miesenberger, K. (eds.) USAB 2009. LNCS, vol. 5889, pp. 416–427. Springer, Heidelberg (2009)
4. Beck, K.: Extreme programming explained: embrace change. Addison-Wesley (2001)
5. Schwaber, K.: Scrum development process. In: Proceedings of the OOPSLA 1995 Conference, Languages and Applications - Workshop on Business Object Design and Implementation (1995)
6. Sohaib, O., Khan, K.: Integrating Usability Engineering and Agile Software Development: A Literature Review. In: Proceedings of the ICCDA 2010 Conference, pp. 32–38 (2010)
7. Singh, M.: U-SCRUM: An agile methodology for promoting usability. In: Proceedings of the AGILE 2008 Conference, pp. 555–560. IEEE (2008)
8. Larusdottir, M.K.: Usability evaluation in software development practice. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 430–433. Springer, Heidelberg (2011)
9. Beyer, H.: User-Centered Agile Methods. In: Carrol, J.M. (ed.). Morgan & Claypool Publishers (2010)
10. Kuusinen, K., Mikkonen, T., Pakarinen, S.: Agile User Experience Development in Large Software Organization: Good Expertise but Limited Impact. In: Human-Centred Software Engineering, Toulouse (2012)
11. Larusdottir, M.K., Cajander, A., Gulliksen, J.: Informal Feedback Rather Than Performance Measurements - User Centred Evaluation in Scrum Projects. Behavior of Information Technology (2013)
12. Larusdottir, M.K., Bjarnadottir, E.R., Gulliksen, J.: The focus of usability testing in the software industry. In: World Computer Congress, Brisbane, Australia (2010)
13. Jia, Y., Larusdottir, M.K., Cajander, Å.: The Usage of Usability Techniques in Scrum Projects. In: Winckler, M., Forbrig, P., Bernhaupt, R. (eds.) HCSE 2012. LNCS, vol. 7623, pp. 331–341. Springer, Heidelberg (2012)
14. Oxford Online Dictionary, `http://oxforddictionaries.com/`
15. Gotterbarn, D.: Informatics and professional responsibility. Science and Engineering Ethics 7, 221–230 (2001)
16. Gulliksen, J., et al.: Making a difference: a survey of the usability profession in Sweden. In: Proceedings of the NordiCHI 2004 Conference, pp. 207–215. ACM Press, Tampere (2004)
17. Silva da Silva, T., et al.: User-Centered Design and Agile Methods: A Systematic Review. In: Agile Conference (AGILE), 2011, pp. 77–86 (2011)
18. Düchting, M., Zimmermann, D., Nebe, K.: Incorporating User Centered Requirement Engineering into Agile Software Development. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4550, pp. 58–67. Springer, Heidelberg (2007)
19. Kollmann, J., Sharp, H., Blandford, A.: The importance of Identity and Vision to user experience designers on agile projects. In: Proceedings of the Agile 2009 Conference, pp. 11–18 (2009)

20. Kuusinen, K., Väänänen-Vainio-Mattila, K.: How to make agile UX work more efficient: management and sales perspectives. In: Proceedings of NordiCHI 2012 Coference, pp. 139–148. ACMPress, Copenhagen (2012)
21. Ezzy, D.: Qualitative analysis: Practice and innovation. Psychology Press (2002)
22. Schwaber, K., Beedle, M.: Agile software development with Scrum. Prentice Hall, Upper Saddle River (2002)
23. Mayhew, D.J.: The usability engineering lifecycle. Morgan Kaufmann Publishers, San Francisco (1999)
24. Cajander, Å., Gulliksen, J., Boivie, I.: Management perspectives on usability in a public authority: a case study. In: Proceedings of the NordiCHI 2006 Conference, pp. 38–47. ACM Press, Oslo (2006)
25. Karat, J., Dayton, T.: Practical education for improving software usability. In: CHI Conference, ACM Press/Addison-Wesley Publishing (1995)
26. Cajander, Å., Daniels, M.: Development of professional competencies in engineering education. In: Frontiers in Education Conference (FIE). IEEE (2011)
27. ISO 9241-11: Ergonomic requirements for office work with visual display terminals, International Organisation for Standardization, Geneva, Switzerland (1998)
28. Cohn, M.: Succeeding with Agile. Addison Wesley, USA (2010)
29. Robinson, H., Sharp, H.: Collaboration, Communication and Co-ordination in Agile Software Development Practice. In: Mistrík, I., Grundy, J., Hoek, A., Whitehead, J. (eds.) Collaborative Software Engineering, pp. 93–108. Springer, Heidelberg (2010)
30. ISO, ISO 9241-210: Ergonomics of human-system interaction - Part 210: Human-centred design process for interactive systems, International Organisation for Standardization, Geneva, Switzerland (2010)
31. Stone, D., et al.: User Interface Design and Evaluation. Morgan Kaufman (2005)
32. Morgan, G.: Images of Organization London. Sage Publication Ltd. (1996)
33. Gulliksen, J., et al.: Key principles for user-centred systems design. Behaviour & Information Technology 22, 397–409 (2003)
34. Cajander, Å.: Usability-Who Cares?: The Introduction of User-Centred Systems Design in Organisations (2010)
35. Wixon, D.: Evaluating usability methods: why the current literature fails the practitioner. Interactions 10, 28–34 (2003)
36. Kujala, S., Kauppincn, M.: Field studies in practice: Making it happen. In: Proceedings of the NordiCHI 2004 Conference, pp. 297–303 (2004)
37. Keil, M., et al.: A Framework for Identifying software project risks. Communication of the ACM 41, 76–83 (1998)
38. Cavoukian, A.: Privacy by Design... Take the Challenge, Information and Privacy Commissioner of Ontario, Canada (2009)

Fast Train to DT: A Practical Guide to Coach Design Thinking in Software Industry

Muktha Hiremath and Visvapriya Sathiyam

SAP Labs India, User Experience, ITPL Main Road, 138,
560066 Bangalore, India
{muktha.hiremath,visvapriya.sathiyam}@sap.com

Abstract. In this paper, we explain best practices of bringing together diverse teams from business, technology and user experience in a large-scale software development setup and coaching them to use design thinking as a methodology to product definition and innovation, in less than 5 weeks. This paper can serve as a primer for those who are new to design thinking and coaching. It contributes to a better understanding of the importance of a coach in nurturing a design thinking mindset.

Keywords: Design Thinking, Innovation, Human Centered Design, Coaching.

1 Introduction

Design Thinking as a human centered approach to innovation has gained immense popularity amongst multinational organizations[1] in recent years. In our organization, we define Design Thinking as

"an effective approach to developing the right products, with the right people, in the right environment".

Design Thinking (DT) is also increasingly used in software product development companies as a tool for innovation (examples: Nokia, SAP, General Electric). It follows a human centered triad approach to thinking giving equal importance to human needs, business requirements and technical feasibility [3].

In the past, we used User Centered Design (UCD) methods in our software product development process that did not succeed due to three critical flaws 1) UCD was part of an assembly-line approach to software development. Business experts stood first in line and defined the solution. Designers visualized and passed the designs to technical teams to implement it 2) We could not mandate end user research due to stringent timelines for delivery 3) It was too late to change the product road map even if end user research was conducted and usability testing only fixed interaction-level issues of the developed product. Design Thinking could fix all these flaws because 1) It is a

[1] Ideo popularized Design Thinking and has applied Design Thinking to a whole host of consumer companies (P&G; Nike; ConAgra etc); apart from using it to redesign Services (hospital emergency rooms; restaurants; NGOs etc).

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 780–787, 2013.

radical collaborative approach to software product definition where the business, technical and user experience experts together define the solution along with their customers 2) End user focus formed the core of human centered design 3) The process by itself is agile and therefore complements an agile development project. We practice design thinking in just 4-5 weeks (equivalent to 1 takt or sprint cycle in an agile development setup).

We coach design-thinking projects in our organization and this paper is a culmination of our best practices, especially in a software development setup. The projects we coached focused on designing different parts of our business software for small, medium and large enterprises. Our teams consisted of people with diverse backgrounds (business experts, technical architects, user researchers, interaction designers, developers and quality experts) and we typically run them for five weeks long. Here below we have summarized 1) overall best practices to coaching, 2) specific techniques and methods that can be used in each of the design thinking phases to nurture a design thinking mindset. The paper aims to provide a checklist to keep in mind for coaches and therefore does not go deep into project specific use cases and examples or comparisons empirical in nature. It serves as a "primer" to kick-start and successfully coach teams in a small-scale agile setup.

2 Design Thinking Coaching

Our role as coaches for design thinking in our organization was two-fold:

1. To coach the team to use the methodology for re-defining their software within 5 weeks, in a way that the design can be productized immediately.
2. To nurture and develop a design thinking mindset within the organization.

Our coaching practice began with a thorough understanding of how design thinking is taught in Stanford D-School [6] and by IDEO's online initiative that brings all design thinking educators on the same platform [3]. We fine-tuned the process to fit our software development model, the essentials of which we share below.

2.1 Challenges

As design thinking was still nascent in our organization, we had several challenges of which the critical ones were:

1. **Stakeholder Buy-in and Mindset change:** Our first hurdle was to convince management to experiment a collaborative definition process that is NOT driven only by solution experts.
2. **Employee participation:** As the process demands 4 weeks of full-time commitment, we had to plan several months in advance for participant availability.
3. **End User Reach:** In a business setup, getting direct access to our end users posed several legal and organizational constraints. We had to plan for end user research several months ahead and even use our personal networks to tap users

from a diverse set of background (potential customers, customer's social network, and next generation users).

Overtime we learnt that the key to coaching and running a successful design-thinking project is to **plan well ahead** and to ensure we **live by its truest spirit**. There were five aspects we paid close attention, to make coaching effective.

1. **A good preparation and presentation:** We introduced a scoping phase for team members to plan ahead, involve stakeholders in the process and get acclimatized to the design thinking mindset [6]. Similarly we wrapped up with a presentation phase to prepare and showcase the results of the design-thinking project.
2. **Design Thinking Space:** We used a dedicated space for the duration of the project to ensure that an immersive environment can be created with sufficient wall space, air, water, sunlight, music, projectors, sketching and modeling tools.
3. **Teaming strategy:** We split teams into two groups (4-6 participants each) to create a sense of competition, agility and bring in diverse perspectives. We also balanced the group based on gender, age, experience and expertise. Our ideal team setup had an interaction designer, user researcher, product owner, technical architect and a developer or a quality expert. We also analyzed individual personalities before forming the groups to ensure teams can self-regulate conflicts.
4. **Integrated coaching:** We had two coaches full-time with the team for the 5 weeks of the design-thinking project. Anything lesser (one coach, part-time coaching, consulting hourly) would be a compromise and would affect team's agility and quality of deliverables.
5. **Obsessive Time boxing and Concrete Deliverables:** Every phase began with a coach led practice session, followed by tightly time-boxed execution that concluded with concrete deliverables ready for the next phase.

3 Best Practices: Design Thinking Process

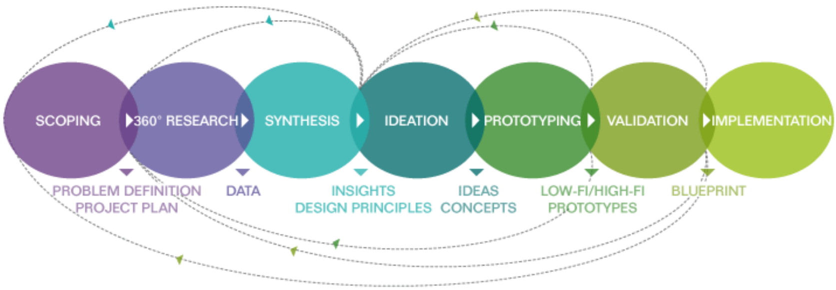


Fig. 1. Design Thinking process as followed in our organization

In our practice, we adapted the 5-step approach to design thinking (DT) by adding a 'scoping' phase at the beginning and culminating in a 'presentation' phase. In this section, we highlight the most important techniques and methods for each of these phases.

3.1 Scoping

A scoping phase is especially important in large-scale design thinking projects to a) plan customer and stakeholder interactions across the globe b) plan project member's availability from different teams for 5 consecutive weeks c) plan the 5 week design thinking project efficiently. Our best practices for this phase are the following.

a. **Start early:** Team members meet for 1-2 hours every day in the 1st week to get acclimatized into the mindset and the process.
b. **Co-define the problem space.** The coach lead the teams to articulate the problem space and in putting down their hypothesis, assumptions and constraints. Though the problem space would be revisited during the synthesis phase after meeting users, we found that a free-flow dialogue with the team brings everyone on the same page and helps team put down "what is known" versus "what is unknown".
c. **Share knowledge "stories".** Teams share their current understanding using stories. We avoid using any documents or presentations.
d. **Plan ahead for research.** Teams identify end user profiles to reach out to them and arrange for all local or remote meetings, ahead of research.
e. **Game every meeting:** Gaming brings agility and creativity and breaks any biases towards action. We even create games together that become an addiction [7].
f. **Plan for trials** for every phase. We quickly try out a technique without any bias towards action. In our practice, we have experimented by running a one-day crash-DT within the scoping phase of a longer DT of 5 weeks to get the team oriented to the spirit of the process.
g. **Kick-off as a team** officially with stakeholders at the end of this phase when the team has a detailed plan of the weeks ahead and in particular demonstrates a collaborative mindset already. This boosts stakeholder and team confidence and ensure higher rate of success and buy-in from everyone. A project that starts with a grand kick-off early without a team or a concrete plan often dies after a few weeks.

3.2 Research

In the research phase, design thinkers learn first-hand about their problem space by observing and interviewing end users and finding their needs and motivations. As coaches, we found that learning to be empathetic cannot be taught in classroom only and devised measures to help teams practice it in the field with heightened awareness.

1. **Use an empathy stick** as a self-regulatory mechanism to observe the basic rules: maintaining eye-contact, talking less, listening more, keeping questions open

ended and deferring judgment. Our teams have innovated using hand signs, picture cards and gestures to warn their group members.

2. **Keep the roles clear and simple:** Teams split themselves into groups of three with each one playing the role of a) initiator who watches what users say b) observer who watches what users do c) facilitator who watched how user feels.
3. **Prepare to ask empathetic questions.** We spent significant time framing the 'right' questions to ask users along with the teams.
4. **Game all interactions with end users.** Gaming keeps users engaged in an activity with 'all' his senses and elicits his 'unsaid' needs. As coaches we prescribed suitable games and also let our teams create new games.
5. **Take 30 minutes at end of day together as a team.** Often preached, but never practiced, we use the time to share our learning and to plan our next day.
6. **Plan for adjacent and analogous user analysis.** The 'wild' ideas during ideation stemmed up from making a connection to insights from these alternate situations. (studying how ants collect food as a colony to understand 'management')
7. **Take extreme users seriously.** We observed and interviewed users three levels down the value chain: our direct customers, our customer's customers and customers customer's end users to develop a holistic understanding of the problem.

3.3 Synthesis

Synthesis is most grueling of all the phases, as the teams work towards seeing connections over hundreds of facts obtained from research and making intuitive leaps about their understanding of the problem.

1. **Make one participant the persona or a user** to steer the discussion from the user's point of view and to make quick decisions where conflicts arise.
2. **Target minimum 3 conceptual models**. We have used several frameworks to analyze research data such as customer journey maps, semantic and temporal zoom and conceptual mind maps [5].
3. **Visualize and headline insightful moments** before they die. We kept a space on the design wall to record insights visually along with a user statement or story.
4. **Moderate to think as a group.** To think as a group requires one to listen to one's own thoughts and to listen to others, build on each other, and articulate it appropriately. We regulated the individual thinking time and the group sharing time until the team was able to group-think on its own.
5. **Check the team mood periodically.** We used two techniques – a fish bowl after every phase and the 'I like, I wish' [2]at the end of every day

---

[2] A fishbowl is a form of dialog that can be used in large participatory setup of more than 10 members with the coach moderating the conversation.

The 'I like I wish' is a method adapted from Stanford d-school's method cards and used to get quick and frequent feedback from colleagues and end users.

3.4 Ideation

Design thinkers use the 'brainstorming' technique popularized by Alex Faickney Osborn [10] for group ideation. The coaches play the role of a guide taking the participants through the steps of ideating, sharing, voting, selecting and generating actionable ideas. The coach also moderates the sessions ensuring the fundamental brainstorming rules [6] are not violated and in time keeping.

1. **Invite experts, customers and stakeholders to brainstorm.** Stakeholders' buy-in to build or sell the product increases when they contribute to the creative process. We have run brainstorming sessions with 15 participants in one session.
2. **Brainstorm in rapid-fire rounds.** We aim for 100+ ideas. Our brainstorming session has a 10-20-20-10 minute split with 10 minutes individual think time, 20 minutes sharing and building on other's ideas, 20 minute consolidating and agreeing on the most promising ideas, followed by a 10 minute break.
3. **Try a fantasy question as a teaser** to push imagination and think out of the box. This could be a teaser exercise when the creative spirit is dropping.
4. **Use a gong to enforce rules.** We hit the gong when a member does not follow any brainstorming rule. The gong has been more effective in regulating the group than a shout, whistle or any other prop.

3.5 Prototyping

In this phase, teams come together and use rudimentary materials and techniques to give their ideas a physical form and shape. The most commonly cited problem during this phase is when the team is building "one" prototype that is a mash of "all" ideas from the ideation phase. The primary aim of a coach would be to push the teams to deliver several different prototypes they can take to their customers to find out which one works best. A coach also can demonstrate methods of quick prototyping using paper, magnets and Lego blocks and ensure all members are creating parts of the prototypes atleast.

1. **Plan ahead for validation:** Planning for the validation phase brings better clarity on what to prototype.
   a. Decide the storyboards or user scenarios to prototype.
   b. Agree on the number of prototypes (alternatives, variations) to create.
   c. Fix the validation schedule with end users.
2. **Prototype alternatives and variations**, and seek for end user feedback. We split teams into sub-groups of 2 and prototype different parts or variations of the prototype to speed up the process. The aim is to create a bunch of prototypes.
3. **Time-box the phase** so that teams don't spend time re-thinking about the ideas, and build them rapidly. We split the phase into smaller sessions.
4. **Make experimental prototypes only**. The prototypes should also reflect the nature of the thought process – experimental. We use paper, magnets and stickies.

3.6 Validate

Validation is a second chance to meet the end users, but this time with physical prototypes to trigger deeper discussions with the end user. The most common pitfall in this phase is when the group sells the idea to the users instead of receiving feedback and learning from them. To avoid this, the coach guides the team into every detail of the validation process including planning of activities with the end user, creating the validation scripts, identifying the prototypes to be tested and techniques of deriving quantitative and qualitative feedback from the validation sessions. Here below are some good practices for productive validation sessions with end users.

1. **Plan for an action packed validation.** Our validation sessions are short and action-driven. Users interact with the prototype quickly and go by their intuition instead of spending time thinking and interpreting the prototypes. The validation sessions are 30 minutes and are repeated successively with 8-10 users a day.
2. **Meet and refine end of day religiously** adding, removing or modify prototypes based on user feedback.
3. **Use the empathy stick** within the group using gestures, postcards and other self-regulatory mechanisms to follow the rules of empathetic listening.
4. **Test with extreme users.** We often fail to involve customer's customer's users in validation process due time constraints, though we might have met them during the research phase. **Never fall into the rushing-to-finish syndrome.**

3.7 Presentation

We mark the closure of the project when we have a design blueprint well received by our stakeholders and ready for production. Coaching in the phase could include helping the team define a communication strategy to their stakeholders, creating a business case if needed and in creating a design proposal.

Make a persuasive presentation. The teams showcase their work in the form of an interactive demonstration, role-play or as a narrative. They also make a video demonstration that can be distributed to other stakeholders in other geographical locations. Though teams choose their medium of communication, we coach them in persuasive communication techniques and help them build their presentation content [9].
Create a blueprint that combines the best of all the prototypes so that production team can use it directly. This is done in collaboration with the production teams and other stakeholders and the design thinking project comes to a closure [8].

4 Conclusion

In the past we had experimented design coaching through other mechanisms such as consulting over a period of 3-6 months hourly or weekly and building a one-year classroom program. But with no real project or timelines, or dedicated coaching efforts these attempts were worthless. In contrast, the integrated approach to coaching with a five-week live project has just the right momentum, motivation and magic to create perfectly useful solutions and in the process creates passionate design thinkers. The best practices described in the paper helped achieve our goals as a coach, of:

1. **Agility:** we create a design blue print within 5 weeks that is ready for production
2. **Skill building:** our hands-on techniques help teams understand and cultivate a design thinking mindset within 5 weeks with heightened awareness and long-term retention.

And once the first project became a success we saw the organization embracing the approach more willingly. We also witnessed our trained design thinkers spreading awareness and becoming junior design thinking coaches and advocates over time.

4.1 Limitations

Our practices reflect the human desirability perspective in depth, but not business viability or technical feasibility. In particular we had not invested in creating business cases and models, as business viability was assumed to be a given. We also did not focus on technical feasibility as our architects in the projects evaluated feasibility after the design-thinking phase. As a word of caution, this process is best applicable for a software product definition scenario and the same has not been tested for designing customer or business services. It best fits a project setup that aims at creating a realistic design vision within a short time of 4-5 weeks without getting into the details of project and cost estimation, planning and delivery. It guarantees problem clarity, design roadmap, stakeholder and customer buy-in within a very short timeframe.

References

1. Thoring, K., Müller, R.M.: Understanding the creative mechanisms of design thinking: an evolutionary approach. In: DESIRE 2011: Procedings of the Second Conference on Creativity and Innovation in Design (October 2011)
2. Brown, T.: Design thinking. Harvard Business Review 86(5), 84–92 (2008)
3. HCD toolkit, `https://hcd-connect-production.s3.amazonaws.com`
4. Ideo's Design Thinking for Educators, `http://www.designthinkingforeducators.com/`
5. Kolko, J.: Exposing the Magic of Design: A Practitioner's Guide to the Methods and Theory of Synthesis. Oxford University Press (2011)
6. Stanford d.school's Bootcamp Bootleg, `http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf`
7. Gray, D., Brown, S., Macanufo, J.: Gamestorming: A Playbook for Innovators, Rulebreakers, and Changemakers. O'Reilly Media (July 2010)
8. Osterwalder, A., Pigneur, Y.: Business Model Generation. Wiley (2010)
9. Reynolds, G.: Presentation Zen –Simple ideas on presentation design and delivery. New Riders (2008)
10. Brainstorming, `http://en.wikipedia.org/wiki/Brainstorming`

Author Index