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

LNCS 8118

Human-Computer Interaction – INTERACT 2013

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

2 Part II

ifip

Springer

Lecture Notes in Computer Science 8118

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

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 II

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 978-3-642-40479-5 e-ISBN 978-3-642-40480-1
DOI 10.1007/978-3-642-40480-1
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

Julio Abascal, Spain
Al Mahmud Abdullah, The Netherlands
Silvia Abrahão, Spain
Funmi Adebesin, South Africa
Sisira Adikari, Australia
Ana Paula Afonso, Portugal
Jose Luis Alba, Spain
Jason Alexander, UK
Dzmitry Aliakseyeu, The Netherlands
Majed Alshamari, Saudi Arabia
Margarita Anastassova, France
Craig Anslow, New Zealand
Jean-Yves Antoine, France
Mark Apperley, New Zealand
Caroline Appert, France
Nathalie Aquino, France
Liliana Ardissono, Italy
Carmelo Ardito, Italy
Pedro Arezes, Portugal
Myriam Arrue Recondo, Spain
Mark Ashdown, UK
Nikolaos Avouris, Greece
Chris Baber, UK
Cedric Bach, France
Jonathan Back, UK
Sebastiano Bagnara, Italy
Gilles Bailly, Germany
Martina Balestra, USA
Emilia Barakova, The Netherlands
Simone Barbosa, Brazil
Shaowen Bardzell, USA
Javier Bargas-Avila, Switzerland
Louise Barkhuus, Sweden
Barbara Rita Barricelli, Italy
Gil Barros, Brazil
Remi Bastide, France
Eric Baumer, USA
Gordon Baxter, UK
Russell Beale, UK
Elke Beck, Austria
Yacine Bellik, France
David Benyon, UK
Francois Berard, France
Regina Bernhaupt, France
Olav W. Bertelsen, Denmark
Nigel Bevan, UK
Ganesh Bhutkar, India
Nicola Bidwell, South Africa
Matthew Bietz, USA
Mark Billinghurst, UK
Dorrit Billman, USA
Thomas Binder, Denmark
Jeremy Birnholtz, USA
Fernando Birra, Portugal
Renaud Blanch, France
Marco Blumendorf, Germany
Mads Bødker Bødker, Denmark
Cristian Bogdan, Sweden
Davide Bolchini, USA
Birgit Bomsdorf, Germany
Rodrigo Bonacin, Brazil
Marcela Borge, USA
Sebastian Boring, Denmark
Aviaja Borup, Denmark
Paolo Bottoni, Italy
Matt-Mouley Bouamrane, UK
Don Bouwhuis, The Netherlands
Chris Bowers, UK
Doug Bowman, USA
Giorgio Brajnik, Italy
Willem-Paul Brinkman, The Netherlands
Gregor Broll, Germany
Judith Brown, Canada
Nick Bryan-Kinns, UK
Sabin-Corneliu Buraga, Romania
Marina Buzzi, France
Maria Claudia Buzzi, Italy
Daragh Byrne, USA

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

Carleen Mailtland, USA
Alessio Malizia, UK
Jani Mäntyjärvi, Finland
Javier Marco, Spain
Ana-Maria Marhan, Romania
Gary Marsden, South Africa
Mark Marshall, UK
Hannah Marston, Germany
Benoît Martin, France
Diego Martinez, Spain
Célia Martinie, France
Masood Masoodian, New Zealand
Sara Mastro, USA
Maristella Matera, Italy
Akhil Mathur, India
Max-Emanuel Maurer, Germany
Eva Mayr, Austria
Davide Mazza, Italy
Emanuela Mazzone, UK
Scott McCrickard, USA
Tom McEwan, UK
Kevin McGee, Singapore
Indrani Medhi, USA
Gerrit Meixner, Germany
Guy Melancon, France
Alex Mitchell, Singapore
Jose Pascual Molina Masso, Spain
Tonja Molin-Juustila, Finland
Francisco Montero Simarro, Spain
Pedro Moreira, Portugal
Ann Morrison, Denmark
Christiane Moser, Austria
Omar Mubin, Australia
Florian Mueller, USA
Maurice Mulvenna, UK
Dianne Murray, UK
Rod Murray-Smith, UK
Moyen Mustaquim, Sweden
Linda Napoletano, Italy
David Navarre, France
Luciana Nedel, Brazil
Vania Neris, Brazil
Jeff Nichols, USA
Colette Nicolle, UK
Laurence Nigay, France
Erik G. Nilsson, Norway
Valentina Nisi, Portugal
Leonel Nobrega, Portugal
Sylvie Noel, Italy
Manuel Noguera, Portugal
Monique Noirhomme, Belgium
Nuno Nunes, Portugal
Clemens Nylandsted Klokmose, Denmark
Michael O'Grady, Ireland
Kenton O'Hara, USA
Marianna Obrist, UK
Lars Oestreicher, Sweden
Amy Ogan, USA
Marta Oliva, Spain
Eva Oliveira Gonçalves, Portugal
Stefan Oppl, Austria
Lidia Oshlyansky, Sweden
Nuno Otero, Sweden
Benoit Otjacques, Luxembourg
Saila Ovaska, Finland
Ana Paiva, Portugal
Philippe Palanque, France
Jose Ignacio Panach Navarrete, Spain
Emanuele Panizzi, Italy
Nadia Pantidi, UK
Luca Paolino, Italy
Eleftherios Papachristos, Greece
Narcis Pares, Spain
Tessy Pargman, Sweden
Timo Partala, Finland
Avi Parush, Canada
Oscar Pastor, Spain
Fabio Paternò, Italy
Andrew Patrick, Canada
Celeste Lyn Paul, USA
Andriy Pavlovych, Canada
Annelise Mark Pejtersen, Denmark
Helen Louise Petrie, UK
Greg Phillips, Canada
Lara Piccolo, Brazil
Emmanuel Pietriga, Chile
Thomas Pietrzak, France
Marcelo Pimenta, Brazil
Ian Pitt, Ireland

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 II

Long and Short Papers (Continued)

E-Input/Output Devices (e-Readers, Whiteboards)

Facilitating Social Behaviour and Collaboration I

Facilitating Social Behaviour and Collaboration II

Gaze-Enabled Interaction Design

Gesture and Tactile User Interfaces

Gesture-Based User Interface Design and Interaction I

Gesture-Based User Interface Design and Interaction II

Health/Medical Devices

Humans and Robots

Human-Work Interaction Design

Interface Layout and Data Entry

Learning and Knowledge-Sharing

Learning Tools, Learning Contexts

Managing the UX

Mobile Interaction Design

Mobile Phone Applications

Leisure-Based Reading and the Place of E-Books in Everyday Life

Annika Hupfeld[1], Abigail Sellen[2], Kenton O'Hara[2], and Tom Rodden[1]

[1] Mixed Reality Lab, University of Nottingham, NG8 1BB, United Kingdom
[2] Microsoft Research, 21 Station Road, Cambridge, CB1 2FB, United Kingdom
psxah4@nottingham.ac.uk, tar@cs.nott.ac.uk,
{asellen,keohar}@microsoft.com

Abstract. With the proliferation of digital reading technologies and their underlying ecosystem, practices of reading are currently undergoing significant changes. Despite the currency of the topic, we find there is little empirical research on how people incorporate digital reading technologies into their existing leisure-based reading practices. In this paper, we explore the place of e-reading for pleasure in daily life, and how it is co-evolving with practices surrounding printed books. We present a diary study with 16 readers tracking their behaviors and motivations surrounding e-book use. Our findings are relevant to designers of digital reading technologies in highlighting the values guiding people's choices and behaviors concerning e-book use.

Keywords: leisure e-reading, e-books, e-readers, tablets, smartphones, books, diary study.

1 Introduction

Nearly 15 years after the launch of the first commercial e-book reader, digital reading devices and associated e-books have finally reached a mass market. According to recent survey data [21], almost one third of the American population now owns an e-reading device or tablet, and that trend is set to rise sharply. At the end of 2011, e-books were outselling the top 50 bestselling printed books (USA Today). Last year, in the US, e-book sales rose across the board for adult books (49%), children and young adult books (475%), and religious publications (151%). What is clear from this is that the digital is both transforming and being transformed by everyday leisure reading.

It is interesting to contrast this with work-related reading. There has been a long history of HCI work focusing on why digital technologies often fail to support the deep kinds of reading we need to do for knowledge work. This kind of reading, often called "active reading" has been studied in work and academic settings [e.g. 1; 16; 27] with research suggesting that these kinds of activities require good support for annotation as well as more fluid and flexible ways to navigate and deal with multiple documents. Digital technologies often fall short of the affordances of paper in such tasks.

For leisure-based or ludic reading [17], however, the story is clearly quite different. It is curious then, that for this type of reading we know relatively little about why people

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 1–18, 2013.

choose e-reading over printed books, what the relative advantages and disadvantages are of digital versus physical books, and how e-book reading is becoming embedded into everyday practices. Discourse has largely been confined to debates on the impact of e-books on physical books, either mourning or celebrating the future demise of the latter [2, 8]. More than anything, it reflects the ambiguity we encounter during this 'late age of print' [3]. Alternatively, there are survey data (such as that offered by [21; 4]) that look broadly at trends, without delving deep into people's practices, choices and motivations. Rouncefield et al are the exception here in a chapter that beautifully highlights the situated and embodied nature of digital and physical reading in the home [23].

Our aim is to explore the mundane ways in which e-books are finding a place in everyday life. In particular we are concerned with the social and material circumstances that come to shape e-book practices. Rather than simply focusing on reading itself, we are concerned with the broader lifecyle of e-book acquisition, ownership and sharing. More specifically, we wish to ask: how is it that people make choices about what they read and how they read in the course of daily life?

In seeking to document the diversity and richness of the everyday use of e-readers, we shall see that these practices contrast with the presumptions of use embedded into many of the commercial ecosystems surrounding e-readers which tend to focus on supporting the consumption of content by individuals. We will argue that existing commercial infrastructures and software do not reflect the diversity of ways in which people actually acquire, manage and dispose of books. Finally, and most critically, we will also underscore that little recognition is currently given to the inherently social nature of leisure reading. Addressing these current disconnects opens up a design space for future developments that might drive such technologies forward.

In order to capture such detail over an extended period of time and in the context of everyday life, we used a diary study coupled with interviews aimed at unpacking the detail surrounding real instances of e-book use. In terms of the approach of our study, we chose to focus on the reading of e-books (the digital content) for leisure, regardless of the platform on which they are read. Given our interest in everyday practices we do not rule out exploring instances where e-book devices might be used for purposes other than leisure-based reading (although this is not the focus of the study), or where e-reading is done alongside or in conjunction with paper-based reading.

2 E-Reading Literature

Human-computer interaction and information and library sciences have had a long standing interest in e-reading and its potential in professional and academic environments including a number of empirical studies on the impact of technology on reading, Early studies were often focused on comparing digital with paper reading [e.g. 1; 6; 10; 18]. Later, research also looked at e-reading on single devices, including e-readers [27], tablets [19; 26], and smartphones [16]. Most recently, more attention has been paid to e-reading across devices [e.g. 12]. Unfortunately, the research here is focused on *active* reading practices in work or educational settings. There are also some critiques of the design of e-reading technologies [14; 20; 28]. The concern here is to improve on e-reader usability for learning and knowledge work.

The humanities also have a great deal to say about the nature of reading, but very little to say about the role of technology. In literary theory, reading, the reader, and the text are usually conceptualized as abstract entities removed from the historically variable social context from which they emerge. The disregard of the embodied, situated, and material nature of reading has been critiqued by Littau [13]. Mangen [15] elaborates on this point arguing that the immateriality of digital texts prevents us from getting immersed in the way a print book would. Her argument seems to hold for particular pieces of literary hypertext rather than properties inherent to digital technology. The reading of digital media alongside print has also been given attention in the digital humanities under the names of *close*, *hyper*, and *machine* reading, but more as tools for textual scholarship rather than as a leisure activity [9]. The approach is predominantly analytical, critical or speculative, as opposed to empirical.

Empirical approaches to leisure reading can be found in the social science literature. Reading for pleasure, or ludic reading, has been extensively studied by Nell [17]. He defines ludic reading as a play activity in that it 'absorbs the player completely, is unproductive, and takes place within circumscribed limits of place and time' and is 'usually paratelic, that is, pursued for its own sake'. Scales [24] found leisure reading habits to be determined by gender, race and education. While leisure reading is often associated with fiction reading, it has been shown to be motivated by a range of factors, including reading as part of the self, to improve reading efficacy, for social recognition, and to do well in other realms, such as work or education [25]. Closer to our concern of understanding the situated everyday practices of e-reading is Rouncefield and Tolmie's work on e-reading at home [23]. They take an in-depth look at the social nature of bedtime reading, as well as highlighting e-reading as a material, social, and situated practice. Related to this, Follmer et al [7] and Raffle et al [22] have deployed digital reading applications to support bedtime reading with children. These studies give us a good starting point but are limited in focus on the home as a setting and single device use.

In sum, we find that despite the rich picture we can gain of reading from these efforts, each discipline has tended to take a particular focus with respect to the *types* of reading and *contexts* in which reading takes place. Across all of these, there has been a lack of studies focusing on the ecosystem of reading technologies, namely the e-reader, tablet and smartphone, across the types of reading and contexts in which they are naturally used.

3 Method

We chose to use a diary study coupled with interviews as a way of collecting rich and situated data about people's on-going practices in real world contexts. Similar to a raft of previous studies [e.g.1], we used digital cameras to provide memory prompts to provoke discussion around specific episodes of e-reading . This allowed participants to give detailed accounts of everyday episodes that might have otherwise remained unnoticed or be considered too mundane to mention. The approach, then, was deliberately chosen to provide grounded detail about a range of specific instances of e-book use rather than to make general claims about what most people do, or how frequently they do it. The data are therefore not quantitative in nature, nor are they aimed at supporting or refuting any particular hypotheses.

3.1 Participants

Sixteen participants (9 male, 7 female) were recruited using a combination of mailing lists and snowball sampling to achieve a mix in age, gender, and device use (see Table 1). Here, due to the exploratory nature of the study, we were seeking diversity of the sample, rather than any particular balance or contrast within it. The only criterion was that they needed to have read e-books on one or more mobile devices. Participants had on average 1.8 years (or 22 months) of experience reading e-books, with a minimum of 6 months and a maximum of 7 years. All but 3 owned a specialized e-reader (Kindle).

Table 1. Participant: *age, gender, and device type (e-Reader (e.g., Smartphone, Tablet, Laptop)*

Participant	Age	Gender	Device Used for Reading			
P1	50+	male	e-Reader	smartphone	tablet	
P2	50+	female	e-Reader			
P3	18-29	male	e-Reader		tablet	
P4	30-39	male	e-Reader			
P5	18-29	male			tablet	
P6	40-49	female	e-Reader			
P7	30-39	female	e-Reader			
P8	50+	male	e-Reader			laptop
P9	50+	female	e-Reader			
P10	30-39	male	e-Reader			
P11	30-39	female	e-Reader	smartphone		
P12	18-29	female	e-Reader			
P13	30-39	female	e-Reader	smartphone		laptop
P14	18-29	male	e-Reader	smartphone		
P15	18-29	male		smartphone		laptop
P16	18-29	male		smartphone	tablet	

3.2 Procedure

The study proceeded in three parts:

1. *Initial interview*: to position participants in terms of current and past patterns of e-book and reading device(s) use. These typically lasted between 25 and 90 minutes.
2. *Diary period*: to capture particular instances of use over a set period of time.
3. *Follow-up interview*: to gather additional information on each captured episode of use.

During the initial interview, participants provided background information on their device usage, type of content consumption and acquisition, reading history and patterns, and use of physical and digital media. Whenever possible, interviews took place in participants' homes, or else at their workplace. Where a face-to-face meeting was not possible, interviews were conducted over the phone.

The diary period lasted between 7 to 15 days depending on frequency of use, with an average of 10 days. Participants were asked to document all instances of book and e-book reading, as well as any behaviors relating to book and e-book use, such as acquisition, annotation, organization, sharing, etc. Participants were also asked to record any other types of reading on devices they also used for e-book reading, such as reading news, magazines, work documents, etc. While our focus was on e-book reading, capturing other types of content helped us contextualize their reading within the wider digital media ecosystem. Participants were asked to use a digital camera to capture reading episodes. Three participants created written diaries, two of them in addition to the photo diary and one in place of it.

The photographs were then used as memory prompts during the final interviews to elicit detailed accounts of the particular social and material circumstances and behaviors pertaining to book, e-book, and device use. Participants were encouraged to tell a story about each captured episode, telling us about where and when the episode took place, how long it lasted, what terminated it, whether it was interrupted, why they were doing what they were doing, who else was present and what were they doing. All final interviews were conducted face-to-face, either at the participant's home, workplace, or in coffee shops. Final interviews lasted between 20 and 65 minutes. On completion of the study, participants received a £50 Amazon voucher as a thank you.

The combination of participant diaries and in-depth interviews resulted in a rich data set encompassing roughly 20 hours of audio recordings, 147 photographs, and three written diaries. This amounts to an average of about 38 minutes of interview recordings and about 10 photographs per participant[1] among those keeping photo diaries. Audio recordings were partially transcribed and analyzed for emergent themes. Where relevant, photographs were used to complement the analysis of the interview data.

4 Findings

Our study covers a broad set of practices relating to e-reading and the social, material, and contextual concerns underpinning them. We begin with a look at the initial motivations and expectations surrounding e-reader ownership and how these relate to subsequent practices. Following this, we consider the broader ecosystem of content acquisition, ownership and storage before considering reading practices with the devices in the home, and out and about. Finally, we discuss sharing practices around e-books.

[1] This average appears low as some participants documented recurring events, such as a nightly bedtime reading, only once rather than each individual instance of it (as requested). The minimum number of photos taken per participant was three, with a maximum of 22.

4.1 Finding a Place for E-Reading in Daily Life

For most participants, purchasing an e-reading device was driven by some primary anticipated benefits. For participants who were new to digital leisure-based reading, motivations were in line with what one might expect: to save on storage space for books (p3), to avoid carrying around weighty books while travelling (p2), or for ecological reasons, such as reducing their use of paper (p1). Other reasons were economic, with the expectation that the cost of e-books would be cheaper than print books (p10). These are reasons that marketers of e-book appliances have long understood.

Some participants, however, came by their devices more reluctantly, sometimes being given them as presents or being given older devices that had been replaced by newer models. As p9, said: *I didn't ask for it, my husband got it for me as a present. I was resistant to it because I like books, the feel of them, the variety of them, and that hasn't stopped through having an e-book, it has just supplemented it.* In this instance, we can see some concern that something might be lost through the transition from physical to e-books. To p12, on the other hand, receiving a device was a revelation: *because I've not really read much for a long time because of the slight dyslexia [] I thought I can't imagine it's going to make that much of a difference but it really has; [] I'd say that I definitely be reading more, it just makes things more accessible for me really and I enjoy it.*

Fig. 1. Integrating the e-book into an established morning routine of studying the bible in English (left) and Greek (centre) with the help of a Greek dictionary (right)

As users explored the device and e-book ecosystem they came to understand its particular affordances for their everyday reading practices. This phase of use was often marked by a period of 'playing around' during which assumptions were tested, unexpected uses discovered and users came to an understanding of which kinds of reading the devices supported very well, and which kinds they did not. For example, p1, p4, p8 and p14 had been hoping to be able to perform non-linear reading tasks that involved flicking back and forth through pages on their e-readers, but soon abandoned these attempts. In p8's case, he had expectations of using his Kindle for bible reading in church, but found the navigation mechanisms for moving between passages too cumbersome: *I've tried and people do use the Kindle in church for bibles but to go jumping around it's really painful.* The fact that looking at more than one page or document simultaneously was not supported was also problematic. P1 struggled with integrating his e-reader into an established routine of reading the bible in Greek and English every morning alongside an English-Greek dictionary and came up with his own way of doing so (see Figure 1). P1 and p15 also bemoaned the fact that visual information was lost on the black and white display, particularly if color was an important aspect of the content or character of the book.

The issue here is not so much to point to problems with e-reading devices as it is to show how, through exploration, people begin to make particular choices for particular reading contexts. These choices in turn begin to carve out new ways in which these practices fit with other aspects of their everyday lives.

4.2 Managing the E-Reading Process

In this section we consider how people managed the process of acquiring, using, organizing, deleting and keeping digital content for e-reading, sometimes in contrast to these same, more ingrained practices with print books.

Finding e-Books. Participants discussed how the move to e-books changed the way they became aware of and found new books to read. Some reported on the difficulty of browsing for leisure reading as it was hard to get a sense of the content of a book from the image and blurb alone. As a result, participants relied on bestseller lists, special deals, and recommendations based on their purchase history. Downloading e-book samples was also a common practice to help decide whether to buy a book or not. P7's comment makes this apparent: *I think it's because I just read a trilogy and wasn't too impressed with the third one and then this is a similar story because it was a recommendation on the Kindle so even though the sample was good I didn't want to get into another book that was very similar to what I've just read.*

Participants also relied heavily on recommendations from friends and colleagues, but this too appeared to be shifting online. P6 and p11 said that, whereas previously they would regularly meet a friend at the pub to discuss and exchange books, now they exchanged emails to recommend books to buy.

Acquiring e-Books. Price was often mentioned as an important factor in guiding participants' purchase decisions. Low-priced e-books (under three pounds) would be bought without much deliberation by most participants (9 out of 16). For more expensive e-books, participants would often obtain a sample before making the decision to buy. Alternatively, participants would opt for the print version if the price of the e-book was comparable, as articulated by p10: *Mostly prices are very similar to normal books, maybe a few pounds cheaper for the newest stuff and I wouldn't bother buying that. Most of the newer books I get from the library in paper copy, because I only read them once and never again, so I don't really want to keep them.* The flipside of this was that very low cost e-books were cited by five of the participants (p2, p6, p7, p8, p11) as changing their reading habits, leading them to buy books they wouldn't normally have considered. P7 in particular celebrated this fact, saying that whereas previously she would have chosen from a limited selection of cheap or free library and charity shop books, she now had access to a greater variety of content online.

Nevertheless, the fact that not all books were available in e-book format meant that most participants would fall back on buying print instead P1, on the other hand, said he would deliberately choose to buy print when he wanted to read the books again, share them, give them as gifts, or read them in the bath. These considerations were also linked to concerns participants had about their e-book collection being tied to a particular platform, and thus being vulnerable to potential incompatibility issues should the technology change in the future. Needless to say, some of the restrictions associated with e-books, such as the inability to share content, were circumvented by illegally obtaining content. For one of our participants, owning both a legal and illegal copy helped him achieve two ends: to pay content providers, such as authors and publishers for work he liked, and to have the freedom to read e-books on a range of platforms.

The ability to instantly download content online seemed to enable participants to more fluidly move from one book to the next. Typically, new e-books were bought shortly before finishing a book. P7 and p16 also spoke of looking for their next e-book when they got bored with their current reading. Content might also be almost instantly acquired in the context of learning about a book, such as reading a review in a newspaper (p7), hearing about a book at a conference (p6), or on a TV or radio program (p15, p2). P1 was able to download a copy of the Lord of the Rings during his vacation on a campsite, the book his daughter was reading and had asked him questions about. Acquiring a copy there and then let him read the book in parallel and discuss it with her. Conversely, the reliance on being able to buy books anytime anywhere caused problems to one couple who assumed they'd be able to buy new books during a vacation but found they didn't have internet access for the duration of their stay (p2).

Many of the participants felt that the instantaneity with which content could be obtained caused them to read more. At the same time, the ease with which content could be obtained required participants to carefully manage their consumption. As p6 said: *It's dangerous, I often, if I'm into it I just keep buying books, most of the books that I've been buying have been 2.99, 3.99 because they've been quite old books, but obviously these have been new and out [] and they're £10 and things and it's lethal because you're not having to literally go and find £10 in your purse, you just press the button, and I have to check myself sometimes, because I think, blimey, I spent thirty pounds or forty pounds on books whereas I wouldn't do that in a shop, it's very easy to do on the Kindle, so I have to check myself sometimes.*

Organizing and Archiving. According to participants, the primary purpose of organizing books was to help them manage the reading process by classifying their books as 'unread', 'to read on vacation', 'currently reading' or 'read'. Most participants let their e-books accumulate on their device before they felt a need or pressure to manage them so as to make books they wanted to read easier to find. Some had an established practice of moving read books into collections or archiving them for that reason, i.e. removing them from the device, but not their account. A common frustration here was that moving e-books into collections did nothing to reduce the length of the books list due to the lack of a folder structure with the result that list navigation remained cumbersome. For some participants, the process of organization spanned both physical and digital books. In order to keep track of both his unread books and e-books p1 kept his Kindle with his books on the shelf as a visible reminder. None of the participants saw a particular need to delete books until they ran out of storage space, but considered deleting samples, free content, or books they didn't enjoy or want to re-read.

The visibility of a book collection also seemed to be an important factor in how participants related to their books and e-books. Reading both print and e-books, p1 faced a dilemma after buying an e-book that was part of a collection of physical books he already owned. Finishing the e-book left him with a desire to put it on the shelf and to wish he had bought the physical copy instead. P9 regarded her e-book collection to be short term, to be something she was not emotionally invested in: *I like having physical books and seeing them and think I must read that, I haven't thought about it; an e-reader collection is very different for me because it tends to be things that I'll read and then discard although I haven't got round to it yet and it's still a bit of a pain but it's not a library of things I am treasuring in the way I do with books; and I know some people are really keen on creating collections and putting them under different groupings, but with me it's slightly different; it feels like it's something more short term and more functional than, I certainly don't have any emotional attachment to my e-reader collection, whereas I do have a very strong emotional attachment to some of my books; books I've read when I was seven; I can't imagine having my e-reader collection when I was seven and keeping them until I was 55, it's just not the same'.*

4.3 Everyday E-Reading Practices

After an initial "settling in" phase, e-book reading found its place both within the home as well as on the move outside of the home. In this section we show how the particular affordances of e-book reading supported both routine and opportunistic reading alongside and sometimes in combination with more traditional reading, and discuss how and why participants made choices about e-reading versus paper based reading.

E-Reading in the Home. Within the home, e-reading took place in a variety of locations through the house: in the kitchen, dining area, lounge, bedroom, toilet, and garden (Figure 2). People were as inclined to curl up with an e-reader as they would have been with a traditional paper book. Much of the leisure reading done by participants in their home had a routine character, often being bound up in the routines and practices of other household members, such as choosing to read on a smartphone rather than tablet or e-reader at night to avoid disturbing their partners' sleep. Of interest then is how e-reading practices pertained to these shared routines within the household and how they enabled reading to be fitted in.

Routines emerged both as a consequence of the social situation of the household and the enabling properties of a particular device. In one example, p2 had begun to read regularly on her Kindle at the kitchen table during breakfast with her husband: *I read at the breakfast table while he is doing the crossword and when he's done, he'll ask me to finish ones he didn't get. I'll read because otherwise I'll talk to him and annoy him.*

She found reading on the Kindle particularly conducive to these circumstances in part because she could leave the device resting on the table without having to hold it open, thereby keeping her hands free for drinking and eating her breakfast. Indeed, this particular affordance of the device led to other circumstances of adopted use in the home. P7 found that the Kindle enabled her to enjoy reading outside in the garden chair: *It's so much easier; when I read outside before I had the pages rafting and it's just that you can sit outside and have a drink in one hand and all it is, is pressing it to get to the next page, there is no sun reflected off it all either*. Here then we see how the e-reader, with its one handed use and rigidity, was able to find its place within the social and material context of the home.

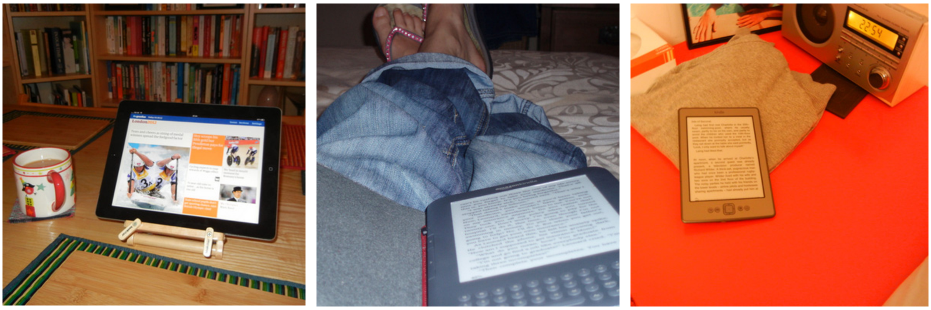


Fig. 2. Routine e-reading at home: (1) during breakfast, (2) after work, (3) before bed

But there were other circumstances where the e-book was not conducive. Notable here was reading in the bath where concerns about damage bring to the fore the expense of the device. As p1 articulated, in such circumstances, he would shift to reading a traditional paperback: *I'd only switch back to the paperback if I was reading in the bath, so if I had a paperback I'd read the paperback in the bath, just because I don't want to risk damaging the Kindle*.

Routine reading also involved making judgments about the text to be read and how to fit coherent chunks or sections of content into a reader's routine. Current e-readers did not readily support these judgments. With the Kindle for example, one of the problems is assessing the length of particular content and how long it might take to read. While there are certain visual indicators of structure, participants found these difficult to interpret in determining their decisions to read certain contents in particular circumstances: *Because it's on the Kindle you don't know what size it is and you only know what percentage through the book you are, and yes it does tell you how many pages, [] but you don't really see that and think yeah I'll give that a go because it's there and you don't know how much you'll be reading*. (p12).

If one of the key benefits of e-book readers is access to collections of e-books, another is the "at-hand" nature of these portable devices. Together, these affordances make for more opportunistic reading in the domestic space. A good example of this is the case of p11 whose days were largely driven by the needs of her newborn daughter and four year old son. For her, reading was something that had to be fitted in, such as when waiting for her daughter to go to sleep or while her son was watching television while having his breakfast or playing in the garden: *I don't really have time to get ready for bed, I don't really have a night time ritual anymore, because I follow [daughter]'s ritual, so basically I put her to bed and then I sit there for awhile waiting for her to fall asleep, and read.*

As another example, p16 started reading e-books on his iPhone when his daughter was born: *I only really started reading a couple of years ago on my iPhone and the reason for that predominantly is my daughter being born; having the books right there available on my iPhone means it's incredibly convenient so when I do have a moment to quickly read I can do so; it's not like having to make sure that you've got a paperback book with you all the time. Plus it means that with the Kindle application I've got multiple books with me and depending on what my mood is at a given point. If I wasn't reading an e-reader I wouldn't be reading anywhere near as much as I used to.* This shows that not only do e-books allow in the moment access to whole collections of books, they also allow for choice, allowing readers to choose the kind of content that will best suit their situation.

E-Reading Out and about. Again, as with traditional books, e-reading found its way into all sorts of locations outside the home including the workplace, pubs, cafes, parks and various forms of transport such as cars, buses, trains, and planes. Of course, as a high level concern, this is all relatively unsurprising, but what is significant are the particular ways that e-reading finds its place in these spaces. Again, particular features of the devices enabled both routine and opportunistic practices of reading to emerge.

Perhaps the most significant factor in the organization of these practices concerns the weight and form factor of the devices. With the Kindle, for example, its compactness and weight meant that it was easily carried in circumstances where traditional paperbacks would be left behind (p11). This in turn meant that people developed new routine opportunities for reading. Most notable here was during the daily work routine, where some participants were able to incorporate the reading of books into their commuting practices in ways that they had previously not done, often displacing other activities or media such as books, newspapers and magazines. Other participants, as a consequence of being able to carry the device, were able to have access to their books at work. E-reading, for some, was something that they would do during their lunch breaks (p2, p5, p7, p8, p13, p14). Again, the issue here was not simply one of enabling reading where it might not otherwise have been performed. Rather it was also about how these devices allowed fiction reading in more kinds of different locations.

As well as the routine practices that participants had constructed around particular devices, as in the home, we saw that e-reading played a significant role in more opportunistic reading when away from the home. Again, what was key here were the unconscious ways in which some of these devices were carried. This was particularly salient in the case of reading from smartphones, which were carried by some participants at all times. This is illustrated by the practices of p14 who found himself in a pub waiting for his cousin to arrive. Waiting in public places can be socially awkward and so he was able to busy himself and pass the time reading a novel on his smartphone. In another example, we saw how p14 used his phone to take advantage of an unanticipated period of sunshine on the way home from shopping. Noticing it was sunny, he stopped off to read and relax in the sun – again not something that he could have planned for with a traditional paperback.

Fig. 3. Opportunistic e-reading at home (top) and out and about (bottom): while (1) watching the children, (2) eating alone, (3) dinner is cooking, (4) the car is being washed, (5) having lunch at work in the car, (6) the sun is out at the pub

When reading in public places, participants oriented to a number of other concerns in the shaping of their practices with particular devices and content. For p3, there was an issue concerning the relative value and visibility of his different devices that led to him make certain decisions when reading on public transport. *It's [iPad] fantastic if you want to do a bit of work or something or get to your emails or write something but for reading on public transport it's not so good but maybe that's just a kind of self-consciousness thing, it just feels a bit uncomfortable taking an iPad out whereas it doesn't feel the same taking the Kindle; I think part of it is because it's expensive I think because it's a bit larger and more noticeable I think.* Here we can see how the iPad for this participant was somewhat conspicuous in this public setting. Not only was he self-conscious, he was also worried about its value to others around him. While such perceptions may be unfounded, they are nevertheless concerns to which he was orienting.

Related to this was the relative invisibility of reading content on these devices and what this meant for reading practices in public places. P12 for example claimed that she only started to read in public places on account of the fact that it was more difficult for others to see what one was reading on a Kindle, allowing her to keep her reading more private when in public. In one episode, when a stranger walked up to her and asked what she was reading she was able to avoid revealing it to him for reasons she explains: *I never lie to be quite honest but there was a guy and he asked me what I was reading the other week [] so I told him about a thing I had on there that I haven't even started reading yet, and I thought I just don't want to tell him because I was reading like this really interesting feminist book and I thought I just felt like this would elicit a strange reaction [] I'm never reading anything bad or anything like that, but I like that privacy about it, that you don't have to advertise to people in public space what you're reading*.

4.4 Social Practices of E-Reading

While individual reading is central to understanding e-book practices, social aspects of use (some of which we have already touched upon) are also important to consider. In this section we discuss how practices of sharing were played out in the context of the e-book ecosystem. In particular, we consider the sharing of content, accounts and devices. Important to note here is that in the United Kingdom the sharing of e-books is currently prevented by digital rights management software on most devices. Amazon allows up to five Kindles to be associated with one account through which books can be shared. What is of significance are the ways that our participants found various workarounds in these constraints.

Sharing Content. If we consider talking about books to be a form of sharing, digital reading devices showed good support in enabling participants to share particular passages during conversations with friends and colleagues. In particular, p12 used bookmarks on the Kindle to quickly revisit passages she wanted to talk to her friends about. Similarly, p15 created bookmarks to discuss certain passages with his colleagues at work: *I think when you bookmark stuff you can put a note alongside the bookmark and I occasionally do that, just if it's something that I want to show to someone at the office if it's something particularly interesting*.

P6 also downloaded an e-book for her daughter to read on her device. Her daughter wasn't interested in the book and she ended up reading it herself. Similarly, both p10 and his wife had Kindles, but on separate accounts. When his wife lost interest in an e-book she was reading on her device they swapped Kindles, so he could read it instead: *I think I can gift her a book or whatever and she can lend it to me but that gets a bit complicated so if I want to read her book or she wants to read my book we just swap Kindles*. Sometimes they would also read the same e-book on the same device which caused problems when the other person revisited the book: *If we're at home and I fancy reading a book that she's reading then I just take her Kindle, but it's a bit annoying because it only lets you maintain the reading position for one book so if she's at a different point in it she gets annoyed when she switches it on and it's on a different page*.

Sharing Accounts. P2 shared her account with her husband and both of her daughters. Rather than being set up deliberately, the sharing of the account evolved out of the way she and her family shared print books before adopting e-reading. P2 was the first to own a digital reading device in her family. She then gave her daughter access to her account when she got curious to try reading e-books, first on her iPhone and later on a Kindle. She also shared the account with her other daughter who read e-books through the Kindle app on her iPad. More recently, she gave her husband a Kindle who then starting reading on the same account. P2 said she would occasionally look through the e-books her daughters bought and download the ones she was interested in onto her device. She also would send e-books she read and liked to her daughter's Kindle, followed by a text message to tell her about it.

P4 gave his father access to his account when passing on his old Kindle to him. Similarly, p7's Kindle was linked to her husband's account who bought it for her. Before buying a book she would usually seek his consent. P14 temporarily gave his mother access to his account when he lent her his Kindle to convince her to get her own. Before doing so, he bought her a number of books he thought she'd be interested in: *Well I'm just trying to sell my mother on the idea of e-books so I lent her for one week and she quite liked it. I lent it to my mother before. I bought a few books for her and put them in a separate folder and stuff. [She] wants to borrow it again and she tends to buy random stuff I'm not going to read so I mean I delete if off the kindle but it's still in my account, you can download it whenever you want.*

Sharing Devices. There was a general reluctance to sharing one's reading devices. Privacy was one concern, lack of access another. Unsurprisingly, unwillingness to share devices increased with how personal the content on the device was perceived to be. While none of the participants minded sharing their print books, only about half would share their Kindle or tablet, and none of them was willing to share their smartphone. Lack of access to the device (and its contents) tended to be participants' main concern when considering sharing their Kindles. As p9 summarizes: *If somebody said would you lend me a book I'd say yes not a problem, whereas the e-reader is a bit different; now why is that? I suppose a) it's more expensive than most books, b) it's a bit like giving someone access to your whole library, which is a bit personal really, because there'd also be (documents on it) as well; and also when I give away a book that I'm not reading, I'm not putting myself to any inconvenience but if I'm giving someone my e-reader then I am in an inconvenience because I won't be able to download anything new.*

P7 occasionally shared her Kindle with her two children, but since they preferred their print books, access to the device was never an issue. In p14's case, sharing a Kindle with his girlfriend for a period of six months before buying his own led to competition over access to it: *Fights, well, we'd basically take it book by book, so I'd have it for a book and then she'd have it for her book but when she wasn't using it I could just like borrow it and it kind of did get a bit, both competing for it quite a lot which did get a bit annoying that's why we decided to get another one. It was just an experiment and they came down in price as well.*

5 Discussion and Implications

The findings show that the experience of reading e-books is shaped not only by the affordances of e-readers, applications, and content providers, but by dint of a larger ecosystem - one that includes paper books and their attendant practices. On first encountering e-books, people explore their potential and their constraints. In turn, they find a place for e-books, in a way that alters, complements, and augments their existing reading practices. The result is that print books complement rather than replace e-books, or put differently, they fill the gap where e-books fall short - and vice versa. In making choices about what platform to read on, people make use of the affordances of current e-reading technology designs, like instantaneity and portability, while giving up some of the values associated with print books, such as visibility and shareability. What we see in people's practices and orientations surrounding e-books is a shift in emphasis from the book as artifact to a set of activities associated with reading. In the following, we discuss these tensions in detail and propose key themes for guiding potential future avenues for the design of reading technologies.

From Commodity to Service. One of the themes that echoed through the data was that, for books that matter, people choose printed books. Digital books were seen as more transitory; they were less about keeping and more about using. The findings point to a number of reasons for this. One had to do with long term access. As an ecosystem of reading devices, applications, content, and service providers, e-reading relies on the availability of each of these components, and more importantly, their mutual compatibility, both now and in the future. Concerns about potential incompatibilities between current and future reading technologies were raised, and may be one reason that people see e-books as not something to invest in for the long term. There were also issues of what it means to "own" a book. Physical artifacts can be displayed and collected. Participants spoke of emotional attachment to print books, not apparent with e-books. Finally, the ability to share might be another reason printed books are more valued. After finishing an e-book, the inability to pass it on to others leaves it as 'dead' content in people's accounts. Taken together these concerns might explain why our participants generally valued e-books less than printed books. This in turn was reflected in their reluctance to pay full "book" prices for e-books. It is clear that people think of the value of e-books differently, and more in terms of the activities that e-books allow them to perform rather than the artifact itself. This suggests that designers might also think differently about their value, moving away from conceiving of e-books as commodities toward thinking about services that enable experiences around reading. This could be manifest in different pricing models, such as allowing people to pay for a subscription service for e-books on a per use basis, or small fees for sharing, or by providing e-books that come with discounts for buying the print version, recognizing that people might want to own the books they want to treasure.

Discovering E-books. E-books are currently delivered through a small set of providers, like Apple and Amazon. This 'walled garden' model, however, does not recognize the many ways in which people find and acquire books. Discovering books to read is a diverse and open set of activities drawing upon a variety of sources, from chatting with friends and acquaintances to idly searching book stores and libraries. The experience online is much more restrictive as people have to rely on information presented to them by the main content providers. We suggest there is a potential for offering more serendipitous and social mechanisms for finding books online, such as allowing people to share recommendations device to device, or posting recommendations in more flexible, ad hoc ways. For example, people might post recommendations linked to a location that reminds them of a good book, or where they spent time reading. Later, others can “come across” these tags and instantly download the books they refer to. The point is that new digital mechanisms for discovery and awareness of books could be much broader and more flexible than it is currently.

Keeping the Reading in Order. Digital reading devices are able to carry large amounts of content, and much of their allure lies in the fact that people have an entire library available at their finger tips, anytime, anywhere. As we have seen, however, large numbers of e-books impose the need to manage the information. Offering some of the information management features that are now available on phone and PC operating systems, such as folders - or in our case ‘shelves’ - would help to alleviate the problem. We have also seen how the divide between digital and physical book collections creates problems in terms of managing people’s reading process, and more broadly, the ability to relate to their books as a single collection. We encourage designers to explore ways of bridging this divide by providing cross-platform visibility and management support. For instance, social media or book sharing sites, such as goodreads.com, could be linked with a user’s online purchase history, wish list, etc. to help them keep track of their reading and book collection. Alternatively, users could be enabled to print a physical prop for each purchased e-book to join their physical book collection.

Moving beyond Fiction. E-reading devices today are optimized for linear reading, which is the reading of a text from beginning to end. Either as a consequence or cause of this, the majority of reading done on e-readers is fiction reading. As we have seen, other types of non-linear or richly visual leisure-based reading, such as the use of cookery books or travel guides, are not well supported by current e-readers yet people (in our study at least) expressed a desire to use e-readers for these kinds of books too. We suggest exploring the design space of e-books and e-readers to support a wider range of reading activities, including ways to support parallel reading (two or more texts next to each other) and more flexible navigation and place-holding mechanisms for active, non-linear reading. This would see e-reading start to reach new markets, and enable new kinds of reading experiences for consumers.

Exploiting the At-Hand Nature of E-Books. The findings confirm that one of the real strengths of e-books is the way that they can support both ad hoc and routine reading practices in new ways. In particular, the lightweight form factor of e-readers, always present nature of smartphones, and ability to pick and choose from either a library of content, or to download new content means that e-reading can be tailored to suit many new situations. Recognizing that this is the case could be exploited more fully in the design of e-reader software and by content providers. For example, rather than classifying by genre, age group and so on, content could also be suggested for “a quick read”, or for longer, more engrossing sessions, or based on location, and so on. In other words the system could be geared toward recommendations which take into account not just a reader's taste, but their context. The interface could give better support for this too, such as giving more rich visual cues as to the length of an e-book, or allowing more flexible browsing of an e-book when time is short.

Sharing. Current business models are dominated by a consideration of the individual as the point of consumption and marketing. The sharing of e-books between users, applications, and devices is constrained by corporate digital rights management (DRM) software. We are aware that some of these constraints are specific to the UK market, but the workarounds people have developed to manage these constraints are indicative of the underlying values that drive people's behaviors and choices regarding e-book use and consumption. Whilst these constraints are in place, people are forced into sharing their devices and accounts with family, partners, and friends or bypassing DRM altogether. To address this issue, we suggest thinking of ways to facilitate the sharing of e-books. For instance, we can think of reading applications that allow people to share their library with friends and family without giving them access to their account details. Alternatively, we encourage designers to enable device sharing that is sensitive to the reader's privacy and personal use preferences, such as place in a book, bookmarks, annotations, font size, categorization, etc.

6 Conclusion

In this paper, we contributed to a rich space of research on reading by exploring the ways in which e-reading technologies have found their place in everyday life. Rather than replacing print books, we found e-books occupy a niche among people's paper based leisure reading practices. In making choices about what technology (including books) to read on, people create a particular experience of reading by drawing on the affordances provided by the particular ecosystem of content, application, device and infrastructure. Based on our findings we suggest to move away from conceiving of e-books as artifacts toward the *activities* and *experience* of reading, including acquisition, organization, and sharing. While running the risk of chiming in with the voices predicting the future demise of the book, it seems to us that books are at an early stage of the transformations software, music and film underwent in going digital. If their histories can serve as an example, and the industry moves from ownership to use based models, understanding the practices of reading will be of paramount importance in helping to design that future.

References

1. Adler, A., Gujar, A., Harrison, B., O'Hara, K., Sellen, A.: A Diary Study of Work-Related Reading: Design Implications for Digital Reading Devices. In: Proc. CHI 1998, pp. 241–248. ACM Press (1998)
2. Birkerts, S.: The Gutenberg Elegies: The Fate of Reading in an Electronic Age. Faber and Faber (1994)
3. Bolter, J.D.: Writing Space: Computers, Hypertext, and Remediation of Print. Lawrence Erlbaum Associates (2001)

4. Book Industry Study Group (2012), http://www.bisg.org/publications/
5. Darnton, R.: The Case for Books. Public Affairs (2009)
6. Dillon, A.: Reading from paper versus screens: A critical review of the empirical literature. Ergonomics 35(10), 1297–1326 (1992)
7. Follmer, S., Ballagas, R., Raffle, H., Spasojevic, M., Ishii, H.: People in Books: Using a FlashCam to Become Part of an Interactive Book for Connected Reading. In: CSCW 2012, pp. 685–694. ACM Press (2012)
8. Gomez, J.: Print is Dead: Books in our Digital Age. Macmillan (2008)
9. Hayles, K.: How We Think: Digital Media and Contemporary Technogenesis. University of Chicago Press (2012)
10. Hillesund, T.: Digital Reading Spaces: How Expert Readers Handle Books, the Web and Electronic Paper. First Monday 15, 4–5 (2010)
11. Kratky, A.: Re-thinking Reading in the Context of a New Wave of Electronic Reading Devices. In: Cipolla Ficarra, F.V., de Castro Lozano, C., Nicol, E., Kratky, A., Cipolla-Ficarra, M. (eds.) HCITOCH 2010. LNCS, vol. 6529, pp. 1–11. Springer, Heidelberg (2011)
12. Lindberg, S., O'Brien, H.: The Use of eReading Devices in Academic Environments. In: iConference 2012, pp. 473–474. ACM Press (2012)
13. Littau, K.: Theories of Reading. Polity (2006)
14. MacFayden, H.: The Reader's Devices: The Affordances of eBook Readers. Dalhousie Journal of Interdisciplinary Management 7 (2011)
15. Mangen, A.: Hypertext Fiction Reading: Haptics and Immersion. Journal of Research in Reading 31(4), 404–419 (2008)
16. Marshall, C.C., Ruotolo, C.: Reading-in-the-small: a study of reading on small form factor devices. In: Proceedings of DL 2002, pp. 56–64. ACM (2002)
17. Nell, V.: The Psychology of Reading for Pleasure: Needs and Gratifications. Reading Research Quarterly 23(1), 6–50 (1988)
18. O'Hara, K., Sellen, A.: A comparison of reading paper and on-line documents. In: Proceedings of CHI 1997, pp. 335–342. ACM (1997)
19. Pearson, J., Buchanan, G.: Real-time Document Collaboration Using iPads. In: BooksOnline 2010, pp. 9–13. ACM Press (2010)
20. Pearson, J., Buchanan, G., Thimbleby, H.H.: Design Principles for eReaders. In: BoosOnline 2010. ACM Press (2010)
21. Pew Internet Center (2012), http://libraries.pewinternet.org/subjects/e-readers/
22. Raffle, H., Revelle, G., Mori, K., Ballagas, R., Buza, K., Horii, H., Kaye, J., Cook, K., Freed, N., Go, J., Spasojevic, M.: Hello is Grandma there? Let's Read! Story Visit: Family Video Chat and Connected e-Books. In: CHI 2011, pp. 1195–1204. ACM Press (2011)
23. Rouncefield, M., Tolmie, P.: Digital Words: Reading and the 21st Century Home. In: Harper, R. (ed.) The Connected Home: The Future of Domestic Life. Springer (2011)
24. Scales, A., Rhee, O.: Adult Reading Habits and Patterns. Reading Psychology 22(3), 175–203
25. Schutte, N., Malouff, J.: Dimensions of Reading Motivation: Development of an Adult Reading Motivation Scale. Reading Psychology 28(5), 469–489
26. Sheppard, D.: Reading with iPads - the Difference Makes a Difference. Education Today, 12–15 (2011)
27. Thayer, A., Lee, C.P., Hwang, L.H., Sales, H., Sen, P., Dalal, N.: The imposition and superimposition of digital reading technology: the academic potential of e-readers. In: Proceedings of CHI 2011, pp. 2917–2926. ACM (2011)
28. Warren, J.: Innovation and the future of e-books. The International Journal of the Book 6(1) (2009)

Reading together as a Leisure Activity: Implications for E-reading

Michael Massimi[1], Rachelle Campigotto[2], Abbas Attarwala[2], and Ronald M. Baecker[2]

[1] Microsoft Research, Cambridge, UK
[2] Technologies for Aging Gracefully Lab (TAGlab), University of Toronto, Canada
mmassimi@microsoft.com, {rachelle,abbas,ron}@taglab.ca

Abstract. Reading from devices such as Kindles, Nooks, and tablets ("e-readers") is an increasingly common practice. A primary reason users purchase e-readers is to read for pleasure, as opposed to reading for work or school purposes. With paper, people sometimes read together from a single book (e.g., reading a bedtime story with a child) – a practice we call partnered reading. This practice, and the goals of e-reading for pleasure more generally, remain underexplored in the HCI literature. This paper contributes findings from a deployment study wherein participants used an e-reader application to read with a partner. These findings (a) provide descriptive accounts of how people use e-readers to read together, and (b) identify opportunities to improve the design of e-readers to support partnered e-reading for pleasure.

Keywords: E-reading, partnered reading, collaborative reading, pleasure, entertainment, leisure, iPad, ALLT.

1 Introduction

A recent survey by the Pew Research Center found that "one-fifth of American adults (21%) report that they have read an e-book in the past year, and this number increased following a gift-giving season that saw a spike in the ownership of both tablet computers and e-book reading devices such as the original Kindles and Nooks" [1]. Arguably, the growth of e-reading has largely been a byproduct of consumers purchasing e-readers for purposes of reading for pleasure [2].

According to the Pew Research Center, "80% of Americans age 16 and older say they read at least occasionally for pleasure. Some 36% read for pleasure every day or almost every day" [1]. Reading for pleasure outranked other reasons for reading, including current events, doing research, and reading for work or school purposes. The widespread appeal of reading for pleasure suggests that e-reading technologies may benefit from design decisions that support this specific type of practice.

While we often think of reading as a solitary activity, reading is always performed within a specific psychological and social context. We have motivations for reading, and direct our reading processes based on the social context that we act within [3, 4]. Collaborative reading in the workplace is often characterized by formal, professional

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 19–36, 2013.

relationships among readers, and with well-articulated goals such as synthesizing material to produce a report [5-7]. In educational settings, readers often try to achieve a deep understanding of the content and engage the material through annotation or note-taking in an active reading process [8, 9]. Learning to read is similarly a social process, and reading together between an adult and a child is a common way to impart literacy skills [3, 10].

Reading for pleasure is no exception, and is often performed in the home with family and friends. E-readers are increasingly available in this social context, but may lack characteristics that make them appropriate for co-located social reading. For example, one study found that 81% of respondents preferred printed books over e-books for reading with a child [1]. Of course, children are not the only types of people with whom one might sit down and read; reading together can be a pleasurable activity for people of any age. Other examples include reading with senior citizens who might have visual or motor impairments [11], or simply sharing a book with a friend sitting next to you. E-reading in the home for purposes of pleasure is still under-explored, despite the success of e-readers for individuals reading on their own.

Part of the pleasure derived from co-located reading is social [3, 4, 10]. The activity of reading provides a shared form of entertainment that can form the basis of conversation, and perhaps co-located reading can contribute to the strength of the relationship between readers. Reading aloud to another person also offers an opportunity to entertain and delight others. Even reading performed separately but in the presence of others can be a relaxing way to spend time together.

Despite the richness of co-located reading for pleasure, it remains a relatively underexplored area in the HCI literature. This paper contributes to what is known about digital reading by describing this practice in more detail, and in particular, by identifying design successes and challenges associated with using an e-reader for co-located pleasure reading activities. We find that e-reading for pleasure requires coordination between reading partners, and there are issues with respect to pacing and visibility of reading activities. E-reading for pleasure constitutes its own form of entertainment, but competes against a growing set of leisure activities available on tablet computers. These findings underpin implications that may improve the design of e-reader hardware, software, and associated services.

2 Background

Several studies concern the design and evaluation of reading devices, especially in comparison to reading from paper [5, 6, 12]. We first turn to the bulk of work in HCI, which has focused on how to improve productivity measures such as comprehension or reading speed. These fall under a process called active reading [8, 9, 13, 14]. Active reading refers to a set of strategies for engaging with written material and is "the combination of reading with critical thinking and learning, and involves not just reading per se, but also underlining, highlighting, and commenting" [15].

Designing digital systems that can support active reading techniques has been a fruitful research area [9, 16], and many commercial e-readers now permit at least some of these active reading techniques to occur [17-19]. Ongoing work has identified ways in which commercial e-readers present usability problems [20], and how users overcome these limitations in pursuit of academic research [21]. Other work explored how the availability of multiple reading surfaces – paper, computer monitors, tablets, slates, and so on – can be understood in work practice in order to optimize the reading and/or writing experience [12, 14]. Document manipulation techniques like navigation and page turning also become important when working across multiple devices [22].

While active reading can be performed with texts chosen for pleasure (e.g., a novel) it is often performed with a particular goal in mind. Adler et al. identify four kinds of reading goals they observed in the workplace: extracting information, integrating information, consistency checking, and critiquing or making comment [5]. For example, a researcher might read hundreds of papers and synthesize them into a literature review. Achieving these goals often involves working with others.

2.1 Social Reading

Social reading research has focused on productivity goals, often in the workplace, university, or other educational settings. Pearson notes that collaborative (or social) reading is not a single activity, but comprises a set of different types of reading tasks [23]. Most devices have been deployed to support a specific kind of reading environment. For example, the XLibris e-reading device allows users to share annotations like highlighting and e-ink [15]. This technology was deployed in an academic reading group where members would meet weekly to discuss conference or journal articles [24]. Pearson et al. created a laboratory scenario where participants were asked to use their BuddyBooks prototype in order to complete reading comprehension tasks, and with particular emphasis on supporting mutual navigation of documents [23].

At a Distance. Other work has focused on how to support social reading despite geographical and temporal distance. Kaplan and Chisik developed a desktop application called the Sociable Digital Library Book that allows groups of young adults between 10 and 14 to read independently but share notes and markings through the internet with remote others [25]. In a field deployment they found that their application encouraged conversation and interaction among readers.

Other work has focused on reading at a distance with children. The “Storytime with Elmo” system allows a distant relative to read a story to a child via the internet [10]. Children read using a physical copy of the book and a video of the remote relative is displayed in a case above the book. Relatives can monitor the child’s reading in their interface, and are shown potential discussion points to raise with the child.

Commercial systems have also taken advantage of social media to share information about readers’ habits and activities. For example, Kobo’s Reading Life platform allows users to publish information about what they’re reading to Facebook

[26]. They can additionally track their reading statistics (e.g., number of pages turned, time spent reading) and participate in online discussions.

2.2 Reading for Pleasure

In the present study we focus on reading for pleasure. By this we mean the pursuit of reading as a way to spend time and for entertainment. Reading for pleasure, described by Ross as "nongoal oriented transactions with texts" [27] is undertaken for many reasons. In a study consisting of 194 interviews, Ross found that those who read for pleasure do so to make sense of their own experiences in a variety of ways (e.g., identifying role models, finding new perspectives on the world).

Reading for pleasure is quite varied: it can mean reading a romance novel on a beach, studying the memoir of a prominent politician, skimming a magazine at the doctor's office, or any of a wide range of potential settings and materials. Nell uses the term "ludic reading" to refer to "an enormously complex cognitive act that draws on an array of skills and processes in many different domains – attention, comprehension, absorption, and entrancement; reading skill and reading-rate variability; readability and reader preference; and reading physiology" [28]. Indeed, while one may read in order to relax, it is not always a soothing process. As O'Hara notes, "[s]ometimes, when reading texts such as thrillers or mysteries, reading for enjoyment is characterized by concentration and high emotional involvement with the text. Such reading may involve trying to anticipate what is ahead in the text and finding relationships among specific ideas and events. This kind of reading will be in a linear fashion and require a high investment in time" [2].

Reading for pleasure can be a complex process to characterize. Relatively little work in HCI has focused on reading for pleasure in a social context however, with the exception of Rouncefield and Tolmie who provide rich accounts of how reading takes place in the domestic space, and especially among family members [4]. As they point out, reading is a richly embodied process, and there are many cues that indicate reading habits, such as how books are held, where they are placed, and how they are stored. Indeed, Ross notes that "reading is in fact motivated and sustained by social relations and embedded in a social context…we need to pay more attention generally to the communal and social aspects of the information encounter and build opportunities for collaboration among users into system design" [27]. As e-readers become more prominent, there become more opportunities for designers to leverage the social context in which reading takes place.

3 Study

The present study is an attempt to unpack some of the social relations that exist around reading, with the goal of moving towards more specific design directions. As mentioned above, there are numerous types of social reading activities available for study. In our case, we became interested in how pairs of co-present readers can read together from a single device. This scenario, unexplored in the literature, is a familiar activity – parents may read with their children, or we may read a book with a loved one as a way to pass the time. In order to better understand how this particular type of partnered reading occurs and is changed by the presence of an e-reader, we conducted an interview study and deployed a custom e-reader application for the iPad called ALLT (Accessible Large-print Listening and Talking).

3.1 Participants

Eleven participants were recruited by distributing flyers to community libraries, academic buildings, and by posting online classified advertisements. To be eligible for the study, participants were required to have an established practice of using an e-reader to read together with a friend or family member. Participants owned e-readers of many varieties, and several different types of relationships were represented in the sample (Table 1). All participants rated themselves "frequent" readers.

Table 1. Participants

ID	Current e-Reader(s)	Age	Gender	Occupation	Partner(s)	Sessions
P1	Kobo Touch	22	F	Student	Cousin (7 yrs.), friends	3
P2	Kindle	23	M	Volunteer	Boyfriend	6
P3	iPad 2	30	M	Sales	Girlfriend	6
P4	iPad 2, Sony e-reader	62	F	Retired	Husband	14
P5	Kobo Touch	24	F	Student	Mother (learning English), sisters	7
P6	iPad 1	21	F	Student	Boyfriend, sister	6
P7	iPad 2	19	F	Student	Mother, brother	7
P8	Kindle	27	M	Student	Mother (learning English), son	6
P9	iPad 2	19	M	Student	Sister, friends	6
P10	Kobo	20	F	Student	Friends	6
P11	Kindle, Android phone	44	M	Actor	Volunteer at community center	2

3.2 System

All participants used a custom prototype e-reader iPad app called ALLT (Figure 1). While participants had experience using their usual e-readers, distributing a common platform to all participants offered two benefits. First, it standardized the experience across participants and allowed us to make reference to a common user interface during analysis. Second, it offered the ability to collect additional data through an embedded on-device questionnaire.

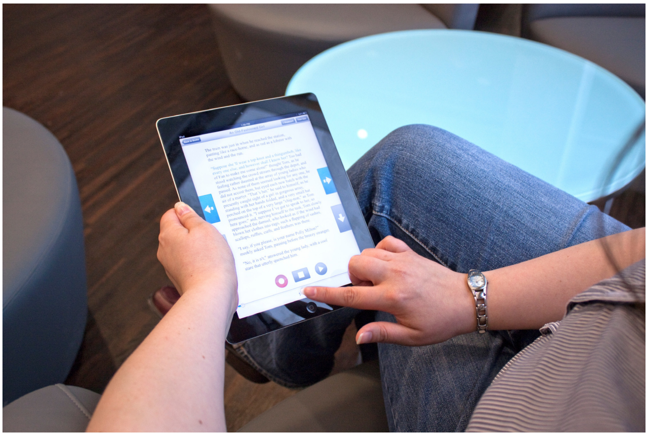


Fig. 1. ALLT e-reader prototype in use

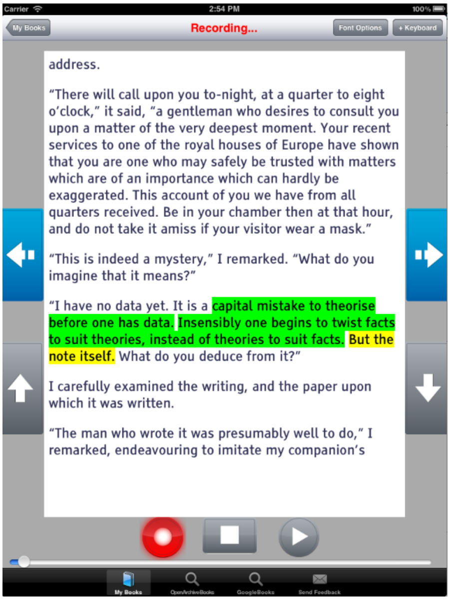


Fig. 2. ALLT user interface. The left and right arrows turn pages, while the up and down arrows control highlighting of text. Voice playback controls are located along the bottom. The current sentence is highlighted in yellow and user recorded sentences are highlighted in green.

ALLT contains a standard set of e-reader features, including: page forward/backward, font size increase/decrease, white on black/black on white contrast modes, table of contents listing, and a slider indicating book progress (Figure 2). ALLT also provides a mode where audio can be recorded through the iPad's microphone. Audio is then synchronized with a highlighted sentence in the book. After the user is done recording the audio for a sentence, they must push a button on the touch screen to progress to the next sentence, and ALLT continues to record. In this way, arbitrary lengths of selected text can be synchronized with user-recorded audio.

ALLT can then later play back the audio track for a selected book. Playback can be performed in one of two modes: using a synthesized text-to-speech (TTS) engine, or by playing back synchronized user-recorded audio. If user-recorded audio is available, ALLT will play it; otherwise, it falls back to the TTS engine (which our participants sometimes called the “computer voice”). This record and playback mode was included in order to support activities where individuals chose to read aloud together, or to create a recording that could be accessed later [11].

ALLT can download books from two sources: the Open Library, and Google Books. Downloaded books are then reformatted before being presented in the ALLT interface. Although books downloaded from outside sources (such as Amazon) were not available in the prototype for copyright reasons, the provided sources contain a large selection of public domain reading material suitable for pleasure reading.

3.3 Method

Each participant participated in a 2-week long deployment study. Participants were initially met in the laboratory where they were interviewed about their existing e-reading habits and preferences. The participant was then given an iPad 2 with the ALLT software pre-installed. If the participant already owned an iPad, the software was installed on their own device.

Participants were instructed to use ALLT with a reading partner for 6 different sessions over the course of 2 weeks. We asked participants to read for at least 10 minutes per session to ensure that participants had at least 1 hour of accumulated usage. Aside from these instructions, participants were encouraged to use ALLT in a way that felt comfortable and mimicked their present use of e-readers. Immediately following each reading session, ALLT prompted the user to complete a post-reading questionnaire.

The questionnaire collected data about who was reading during the session, what they were reading, and the relative utility of features of the system, along with any additional comments that participants wanted to share. At the end of the 2 week period, participants completed an interview with the researcher about their experiences using ALLT. This interview asked participants to comment on the features of ALLT and the process of using an e-reader to read with a co-located partner.

In total, approximately 8 hours of interviews were audio recorded, yielding 96 pages of transcribed material. Transcripts were then analyzed by using affinity diagramming to generate 110 initial concepts. Concepts were then reviewed by the research team and grouped based on thematic connections.

4 Session Characteristics

Across all 11 participants, 70 post-session questionnaire responses were successfully completed. Session times ranged from 10 minutes (minimum) to 60 minutes (maximum). The average reading time was 19.9 minutes, with a standard deviation of 10.6 minutes. Of these sessions, 24 (34%) were completed alone, with the remainder (46) with a reading partner. Sessions alone were, on average, 21.8 minutes, while sessions with a reading partner were on average 18.9 minutes (an insignificant difference, $p = 0.28$). Of the 46 sessions with a reading partner, the participant read aloud more than the reading partner in 14 sessions (30%). In 9 sessions (20%), the participant and partner read aloud in equal amounts, while the partner read more than the participant in 9 sessions (20%). In the remaining 14 sessions (30%), partners read silently.

Each participant read 1 or 2 different books during the study, yielding a total of 17 unique titles. Ten of the 11 participants read works of fiction during the study such as The Call of the Wild, Pride and Prejudice, Grimm's Fairy Tales, and The Adventures of Sherlock Holmes. P3 and P10 also read two non-fiction titles: The Idiot and The Art of Controversy. P10 was the only participant to exclusively read non-fiction.

The relationship between reading partners varied from participant to participant, and included family (parents, siblings, spouses), boyfriends/girlfriends, and in one case, a co-worker. In one instance, P1 read aloud to two children, while all other cases involved only the participant and one other reading partner.

5 Interview Findings

We now report on the findings from our interviews and provide richer characterizations of the reading sessions outlined above. We also provide suggestions of what these findings could mean for the design of e-readers and e-books.

5.1 Reasons for Reading for Pleasure

Entertainment. In the study we chose to use an Apple iPad as the platform for our ALLT software. The iPad is not a dedicated e-reading device, and can provide many forms of entertainment beyond reading, including games, movies, YouTube, and so on. Because entertainment options are more readily available than ever before, we saw that participants sometimes chose other options instead of reading. For example, P11 took ALLT to a community senior center where he volunteers to read.

"Well it definitely brings people together because of the technology... but the seniors asked if there was poker on it...That's a plus." – P11

This quote shows that reading for pleasure is part of a larger set of potential forms of entertainment. Reading on a multi-function device like a tablet opens up new ways of interacting across devices and systems. For example, P5 downloaded and read Sherlock Holmes in all 6 sessions. When asked why he chose that title he responded, *"I just started watching Sherlock [the television show] so I decided 'why not?'"* For designers of e-readers and e-books, there are many opportunities to take advantage of reading's positioning within an entertainment ecosystem. E-books might reference games and videos also available on the device. Thinking about reading as part of a broader home entertainment could lead to compelling situations where devices work together to deliver new entertainment experiences (e.g., a television that plays mood music and background visuals depending on what book is being read).

Entertainment does not occur strictly on the device. P11, an actor, recorded a narrative that he then played for an audience at the senior center. Reading aloud prompted participants to try on voices and add depth to the reading through performance.

"Because you're concentrating and reading out loud… I was a little more conscious of adding drama to it." – P4

While most e-readers do not support the ability to record and synchronize voices with text, ALLT's support of this feature led to other playful behavior. Because ALLT falls back to TTS when there is no user-recorded voice available, participants could juxtapose the artificial computer voice and their own sounds, as with P7:

"I was blow-drying my hair and I pushed the voice recording button…It was like white noise for that section, and first it was the computer voice, and then the white noise – blow-drying – and then computer voice again and that's funny." – P7

Because e-reading for pleasure is not necessarily beholden to the same goals of reading for comprehension, speed, or summarization, the range of purposes for the technology can also grow. Incorporating devices not usually associated with reading – like cameras or microphones – gives users the opportunity to "mess around" with different aspects of the system for entertainment and blend these aspects into the greater activity of reading for pleasure. Tablet games that incorporate multiple players have been relatively successful in the past, and perhaps reading applications could be improved by taking cues from gaming.

Edu-tainment. As described in Table 1, two participants used their e-reader in order to read with their mothers who were learning English as a second language. One other participant used the e-reader to read a book aloud with children learning to read. In both of these examples the primary activity was reading for pleasure, but at the same time, there was an educational component. While there is a rich literature on dedicated devices for second-language acquisition and for childhood literacy, these are outside of the scope of this paper [29, 30]. Instead, we focus on the specific ways that ALLT permitted participants to engage in pleasure reading in new ways.

Building on the idea of reading alongside television, P5 described one way her mother used her e-reader to learn English through entertainment.

"My mom, she just came to the country. Her English is ok but it could be better. They have a lot of things that are meant for primary school kids, but it's easier to read with her. Or sometimes we will get books about a teenage show – Pretty Little Liars. It's easier for her to watch it and read it and compare. She actually writes down stuff and shows me and my sister to look it over and see if she did a good job and can understand the difference between the book and the TV show." – P5

As shown above, reading for pleasure can be one technique for acquiring new language skills. While our sample included few sessions of reading with children, 5 different participants suggested that reading aloud as an educational program for kids would be an effective use of the recording feature. In particular, the ability to record voices creates a useful model that allows learners to check their pronunciation.

"Both of us would read the same paragraph. Also she is trying to get rid of her accent. So when she hears me she is… trying to say it the way I am saying it." – P5

These statements support the possibility of e-reading for both entertainment and education. P3 and P4 used the voice recording feature to annotate their e-books to mark specific passages for later review. This would suggest that reading for pleasure can sometimes involve deeper processing of text, and might benefit from annotation functionality in the same way that reading for work or school purposes might [5, 12].

5.2 Reading as Relationship Work

Reading, like any social activity, is influenced by pre-existing relationship dynamics. Decisions about what, how, and when to read are controlled by how the relationship works. For example, when asked who chose what content to read, P2 said *"I chose what to read. A kind of overarching theme for our relationship,"* indicating the carryover from pre-existing relationships into the reading activity.

Reading together on an e-reader came more easily to some participants than others depending on the relationship. Three participants (P2, P5, and P6) reported using an e-reader together with a partner improved parts of the reading experience. P2 and P5 enjoyed opportunities for "literary discussion" they otherwise would not have had.

These kinds of conversations, as byproducts of reading, can form the basis for stronger relationships. While this kind of benefit is not limited to e-readers, this is a potential place for designers to innovate. For example, e-readers might track a reader's place in a story and suggest discussion topics in order to stimulate conversation.

Gifting and Digital Work. Giving someone else a book as a gift is a common practice. Our findings showed that this kind of behavior persists in e-reading environments. While most participants selected reading material together with their partners or chose books of interest to themselves, P2 and P7 reported browsing the collection of content available in their e-readers for topics that would be of interest to their reading partners as well, and downloaded books for their partners to read later.

"To save my boyfriend some time and effort I tried to look through all of them and downloaded maybe 10 or 15 books that I thought he might also like." – P2

Similarly, P11 reported printing out paper copies of e-books that he was reading in order to share them with friends. E-readers allow users to download a baffling number of books; sifting through books and providing them to others in a preferred format constitutes a form of benevolent "digital work." As noted in the literature, technology can sometimes reduce effort and in so doing, devalue the meaning of the exchange [31]. Designers might consider how downloading and preparing a book to be read can be considered an act of caring. A system might allow a user to modify or extend the digital text (e.g., adding audio narration) in order to personalize the book as a gift.

Equality, Negotiation, and Ownership. Participants remarked on how they negotiated their reading by engaging in common relationship-building practices that affirmed equality. For example, P2, P3, P6, and P7 would alternate back and forth, taking turns reading aloud. P2, P4, P5, and P7 demonstrated a similar form of equality and negotiation around their e-reading by browsing and selecting the reading material together. In all other situations, the participant/owner of the iPad would choose the reading selection (or, as mentioned above, on behalf of their reading partner).

With print books, ownership of the book is usually well-understood. In this situation, however, where books are jointly acquired and jointly read, these principles of ownership become blurred. P1, P6, and P7 reported that they chose what to read because they were the owners of the iPad. For other participants, the iPad was a shared device – P3 was forced to share with others because they only had one tablet. In these situations, the same device might hold multiple books that "belong" to multiple people. Additionally, multiple people might be reading the same book at different times.

Many current e-reading systems are designed based on a model where there is a single user, a single device, and an associated account with an online retailer (e.g., a user has an Amazon account and reads on his Kindle). In reality, multiple people may use the same device and share the account. In initial interviews, P2, P5, and P11 reported that digital rights management software impeded their ability to share books with household members during their e-reading. Bringing a book "into the house" is now changed as a result of e-readers. Use models where each person has an account, or where each person has their own device, do not always exist.

5.3 Pragmatics

All participants reported that reading with a partner was something they do less frequently than reading alone, and that reading together was not always ideal. In this section we detail some of the practical concerns that participants raised regarding their reading experiences. When asked if he enjoyed reading with a partner, P3 noted:

"It's definitely a social experience. I don't know if I can say it's more fun. It's more like reading, based on the person, it's a very solemn pastime." – P3

Indeed, while all participants said that overall they prefer to read alone, they all also saw the value of partnered e-reading in specific cases (e.g., reading to a child, or to someone in hospital). In these kinds of situations there is an existing social and physical setting that makes them conductive to reading. In all cases there was a perceived need to fit reading into existing household routines [4]. Where these routines did not already exist, participants needed to carve out the time from their schedules. This was often done by adding them to existing events, such as reading together before bedtime (P2), after dinner (P6), or after classes (P7). However carving out these times could be difficult. P4, for example, commented on the overhead involved in reading together.

"It's hard unless you book it, almost, make an appointment to do it. Like 'I'll be there in a few minutes' and then you're sitting there waiting. That was the hardest part of reading together." – P4

Because the point of reading is for pleasure, P1, P4, P8, and P10 all reported that they would read on their own and then allow their partner to catch up later, rather than deal with scheduling an appointment. While the questionnaire data showed that reading partners stayed stable during the study, P10 noted that this scheduling could be exacerbated by the challenges associated with finding a reading partner.

Systems that reduce the overhead with scheduling could potentially be helpful here, as would systems that allow people to find reading partners more easily. For example, routine learning systems might track of family and friends' reading patterns and use that to suggest a set of times to read together [32].

Pacing. P3, P5, P6, and P7 remarked on the difficulties they encountered in setting and maintaining an appropriate reading pace with their partner. One of the two partners would read faster than the other and would have to wait for their partner to finish before turning the page. Participants responded in different ways:

- *Use TTS to establish a reading pace*. P7 would turn on the TTS playback in order to establish a shared reading speed for both partners.
- *Reading aloud to maintain reading pace*. One of the two partners would read aloud in order to establish the shared reading pace. P1, P2, P5, P9, and P11 also noted that the person reading aloud would turn the page.
- *Reading silently and then confirming at the end of the page*. P6 and P7 described situations where both partners would read silently to themselves, and then verbally confirm that their partner was ready to turn the page.
- *Use sentence highlighting*. ALLT's voice recording feature highlights one sentence at a time, and allows users to advance sentences using the arrow keys. P1, P2, P5, and P6 used this highlighting (sometimes in conjunction with other methods) to provide a visual indication of their reading progress to their partner.

P2 and P6 encountered situations where their reading partner would read ahead independently, and used the voice recording as a way to catch up.

"If we were sharing a book and he was ahead of me… I could just hit play and it would be like he was reading to me when he was at work." – P2

Participants suggested that novel hardware configurations could address pacing.

"The option to split the screen… so if two people are reading together if the screen was able to split one person based on the other person's reading speed you can flip ahead or back, otherwise you're still looking at the same screen, you know." – P3

Multiple devices for reading activities has been explored before in work settings [14, 22], and future work might explore how multiple displays fare for co-located reading for pleasure as well. Additionally, user interfaces that can better support the communication of pacing through a page or a book could be beneficial; for example, future systems might be able to automatically identify what sentence is being read, and leverage this information to automatically adjust reading pace and page turns.

Interest, Choice, and Distraction. A common theme among participants (P2, P3, P7, P9, and P11) was that their reading interests differed from those of their reading partners. Unlike work or school where the material is often determined by an instructor or work outcome, reading for pleasure introduces choice. If one does not enjoy reading fantasy novels, then there is no need to do so (and indeed would be antithetical to the activity of reading for pleasure). Additionally, the variety of reading materials for pleasure is much greater than for work purposes. Without the demands of school or the workplace, there is relatively little "glue" holding together adults who are reading together (aside from the value found in socialization).

Difficulties maintaining interest in reading together was noted by several participants. Five participants described occasions when they lost interest during the session, and would become distracted with other activities or fall asleep.

"[My] two little sisters, they have a short attention span. They were really quiet and eventually they sort of fell asleep." – P1

In addition to losing interest, some participants felt that partnered reading introduced too many distractions. P2, P3, P4, and P6 felt that having someone else commenting on the text as they read made it difficult to concentrate. P4 believed that the mere presence of another person distracted her from her reading. This corroborates the findings from the questionnaire, which indicated that partnered reading sessions were shorter in duration than individual reading sessions (although not significantly).

E-readers may potentially handle this situation in a number of ways. For example, a lack of a shared book of interest can be a good opportunity for the system to recommend texts that could be of interest to both readers (e.g., by consulting past reading history). E-readers could also more specifically support partnered reading by introducing prompts for different partners to perform different tasks, as in collaborative video games. For example, the e-reader could indicate when to switch reading aloud, or present challenges and awards for continued reading.

Awareness. Reading together requires that both partners have an awareness of where their partner is in their own reading. Unlike print books, e-readers may make it more difficult to identify what household members are reading, where in their books they are, and what their intentions are for the book (e.g., leaving a book in the bathroom vs. the kitchen table) [4]. Current e-readers mark the last page read and return to that page when the book is reopened. When multiple people are reading the same book, however, this kind of bookmark forces the user to flip forwards or backwards to find their spot again. While this has some usability problems for users who do not share a device [20], this may help to provide some context about others' reading activities.

Additionally, because e-books are "invisible" in the sense that they do not have a clearly identifiable physical form, members of the household could not always tell what books others were reading. To handle this situation, P4, P5, and P6 designated particular books as being "for together" and others "for alone." In this way, if one person wanted to read when the other did not, they could continue on their own until both partners were ready to read their shared book.

"I came home once and she was definitely reading something, but it wasn't the book we were reading together, just because then we wouldn't be in the same spot and that wouldn't be fun." – P6

This type of reading practice where multiple people are reading multiple texts is what Pearson calls "parallel reading" [23]. Applications that tie into social media such as Facebook have been developed in order to support awareness in parallel reading environments. For example, the iPad app Subtext allows groups of users to share annotations to books as they read, and browse what other users are reading [33]. Kobo's Reading Life feature automatically publishes information about what the user is reading to Facebook friends [26]. Both of these applications also permit users to post to conversations regarding the material, taking the discussion of the text online.

While these kinds of apps can help to support awareness, there are additional features that might be helpful for co-located partnered reading. Being able to see how far someone else has read in a book can be one way to support awareness. Social reading applications might also consider how to support co-located reading when a device is shared among multiple people (e.g., handling multiple accounts).

Device Interaction. Aspects of the hardware and software affected the partnered e-reading experience in both positive and negative ways. P8 and P10 found the iPad's comparatively large screen to be helpful for partnered reading, so that both people could read the screen more easily. P2 thought that the device was too small to share, and would have preferred a larger screen. P3 suggested the system have adjustable margins so that line lengths were shorter, and shifting from line to line could be made easier. P7, P8, and P9 all suggested that the application flip from portrait to landscape so that the iPad could be held sideways and placed across two laps.

While overall there was a desire for larger screens, this may result in a tradeoff with respect to holding the device. P4 and P10 found that holding the device in a position that was comfortable for both reading partners to be physically awkward and tiring, although they liked the fact that e-readers didn't require them to continually hold the book open to maintain the page, as with some printed books. P7 suggested that the e-reader should come with a stand so that it could be placed on a table (although add-on stands for the iPad are already commercially available). Four participants (P2, P4, P6, and P7) remarked on the need to adjust the font size and contrast in order to make reading more comfortable, while P5 and P6 additionally mentioned the need to find a location that had suitable lighting for the reading task. Importantly, the preferences for settings like font size and contrast could be different for each person.

How the device is held affects the way that screen interaction occurs. P8 noted that the touch screen of the iPad made it ideal for partnered reading. For page turns, some participants would pass the device back and forth, and whoever was holding the iPad would turn the page. P6 reported that in their reading sessions, whoever was sitting to the right would turn the page, since they were closer to the right edge of the screen.

These observations suggest opportunities to improve the design of e-readers for partnered reading. For example, reading settings (font size, contrast, etc.) could be stored for each of the two reading partners. When the device is then passed from one person to the other, the appropriate settings could be swapped in by pressing a button or by detecting the passing motion using an accelerometer. Screen layouts could also be optimized for different positions, such as reading from a stand or one's lap.

6 Implications for Design

While we have touched on some potential changes to e-readers in the themes above, we now present a distilled list of design implications from the findings. We note that these implications stem from our small-scale deployment study with ALLT and may not be suitable for generalizing across all device platforms and populations.

Support group discovery of shared reading material. Whether e-books are downloaded through a PC or to the e-reader directly, systems for acquiring books should support social reading activities such as giving books as gifts, identifying material that others might be interested in, and helping people to explore mutual reading interests.

Allow books to be personalized and passed on. Finding books of interest to others and passing them along can be a way to show thoughtfulness. Designers should consider how personal touches can be imbued into e-books, such as including custom recordings or notes, in order to facilitate giving or receiving an e-book as a gift.

Support coordination of reading activities. Readers may be in the process of reading multiple books at the same time – some of these might be read with others, while others are read alone. Systems should make reading progress visible in order to support the coordination needed to engage in partnered reading activities.

Provide short texts and divide long text into chunks. Partnered reading sessions in our study were shorter than individual reading sessions, and were subject to interruption and distraction. Providing brief reading materials and a clear sense of progress through a text can help readers stay on track and give opportunities for breaks. Longer materials such as books should be clearly divided into subsections.

Accommodate multiple reading paces. Some people read faster than others, and systems that are able to adapt to differing paces may better support partnered reading activities. For example, a system might have a split screen option that allows for two readers to read at difference paces while still reading together.

Settings should be quickly swapped in and out. With multiple people using the same device to read, settings for font size, contrast, and so on should be easily swapped out to suit the preference of the current reader. This could occur through a hotkey or by sensing the motion of passing the device from one person to another.

Consider hardware sizing, positioning, and fatigue. Users adopt many postures and positions when reading together. Stands, cushions, and other device add-ons can improve the comfort of a partnered reading experience.

Consider how reading fits into other forms of entertainment. Reading for pleasure is only one way to spend leisure time together, especially when games, TV, and movies are all available on tablet computers. Promote integration of reading into other forms of entertainment to enhance the shared reading experience.

Support creativity in the act of reading. Reading is a form of entertainment and has performative aspects to it, such as creating voices. Systems should help readers bring a story to life by providing tools to enhance the drama and storytelling, and bear in mind that reading can be the basis for "silly" forms of entertainment.

7 Conclusion

e-Reading is an increasingly common household activity, and reading for pleasure is a contributor to the adoption of e-readers. Right now, there is a trend towards each person having their own e-reading device, but sharing e-readers in order to engage in partnered reading is a practice that still has its place in the order of the household. Reading aloud with friends and family can be a way to pass the time and enjoy the company of others. Designers of e-readers and e-books should consider this rich type of activity in order to better support partnered reading.

In this paper we have touched on some of the opportunities for e-readers to improve this practice by making it easier to schedule, maintain, and enjoy partnered reading. We have found that there remain issues concerning awareness, presence, and pacing that can make e-reading together more difficult. At the same time, reading for pleasure is a way that people can connect and find entertainment value. Systems should support the creative and social aspects of reading together, while at the same time addressing physical and user interface characteristics that can make reading together uncomfortable in order to improve the quality of partnered e-reading.

Acknowledgements. We'd like to thank the participants, members of TAGlab, and GRAND.

References

1. Rainie, L., et al.: The rise of e-reading. Pew Research Center's Internet and American Life Project (2012)
2. O'Hara, K.: Towards a Typology of Reading Goals. Rank Xerox Research Centre (1996)
3. Bloome, D.: Reading as a Social Process. Language Arts 62(2), 134–142 (1985)
4. Rouncefield, M., Tolmie, P.: Digital Words: Reading and the 21st Century Home. In: Harper, R. (ed.) The Connected Home: The Future of Domestic Life, pp. 133–162. Springer (2011)
5. Adler, A., et al.: A diary study of work-related reading: design implications for digital reading devices. In: Proc. CHI 1998, pp. 241–248 (1998)
6. Sellen, A.J., Harper, R.H.R.: The Myth of the Paperless Office, p. 231. MIT Press (2003)
7. Sellen, A., Harper, R.: Paper as an analytic resource for the design of new technologies. In: Proc. CHI 1997, pp. 319–326. ACM (1997)
8. Tashman, C.S., Edwards, W.K.: Active reading and its discontents: the situations, problems and ideas of readers. In: Proc. CHI 2011, pp. 2927–2936. ACM (2011)
9. Schilit, B.N., Golovchinsky, G., Price, M.N.: Beyond paper: supporting active reading with free form digital ink annotations. In: Proc. CHI 1998, pp. 249–256. ACM Press/Addison-Wesley Publishing Co. (1998)
10. Raffle, H., et al.: Family story play: reading with young children (and elmo) over a distance. In: Proc. CHI 2010, pp. 1583–1592. ACM (2010)
11. Snelgrove, W.X., Baecker, R.M.: A system for the collaborative reading of digital books with the partially sighted: project proposal. In: Proc. BooksOnline 2010, pp. 47–50. ACM (2010)
12. O'Hara, K., Sellen, A.: A comparison of reading paper and on-line documents. In: Proc. CHI 1997, pp. 335–342. ACM (1997)
13. Landoni, M.A.: The active reading task: e-books and their readers. In: Proc. BooksOnline 2008, pp. 33–36. ACM (2008)
14. Morris, M.R., Brush, A.J.B., Meyers, B.R.: Reading Revisited: Evaluating the Usability of Digital Display Surfaces for Active Reading Tasks. In: Proc. TABLETOP (2007)
15. Price, M.N., Schilit, B.N., Golovchinsky, G.: XLibris: the active reading machine. In: Proc. CHI 1998, pp. 22–23. ACM (1998)
16. Marshall, C.C.: Annotation: from paper books to the digital library. In: Proc. JCDL 1997, pp. 131–140. ACM (1997)
17. Amazon. Amazon Kindle (2012), `https://kindle.amazon.com/` (cited September 12, 2012)
18. Nook. Nook, `http://www.barnesandnoble.com/u/nook/379003208` (cited September 12, 2012)
19. Kobo. Kobo e-Readers (2012), `http://www.kobo.com/ereaders/` (cited September 12, 2012)
20. Pearson, J., Buchanan, G., Thimbleby, H.: HCI design principles for ereaders. In: Proc. BooksOnline 2010, pp. 15–24. ACM (2010)
21. Thayer, A., et al.: The imposition and superimposition of digital reading technology: the academic potential of e-readers. In: Proc. CHI 2011, pp. 2917–2926. ACM (2011)
22. Chen, N., et al.: Navigation techniques for dual-display e-book readers. In: Proc. CHI 2008, pp. 1779–1788. ACM (2008)
23. Pearson, J., et al.: Co-reading: investigating collaborative group reading. In: Proc. JCDL 2012, pp. 325–334. ACM (2012)

24. Marshall, C.C., et al.: Introducing a digital library reading appliance into a reading group. In: Proc. JCDL 1999, pp. 77–84. ACM (1999)
25. Kaplan, N., Chisik, Y.: Reading alone together: creating sociable digital library books. In: Proc. IDC 2005, pp. 88–94. ACM (2005)
26. Kobo Reading Life (2012), `http://www.kobobooks.com/readinglife` (cited 2012 September 11, 2012)
27. Ross, C.S.: Finding without seeking: the information encounter in the context of reading for pleasure. Infor. Processing and Management 35(6), 783–799 (1999)
28. Nell, V.: Lost in a book: The psychology of reading for pleasure, p. 336. Yale University Press (1988)
29. Liu, M., et al.: A Look at the Research on Computer-Based Technology Use in Second Language Learning: A Review of the Literature from 1990-2000. Journal of Research on Technology in Education 34(3) (2002)
30. Levy, M., Stockwell, G.: CALL Dimensions: Options and Issues in Computer-Assisted Language Learning. Teaching English as a Second or Foreign Language - Electronic Journal 11(2) (2006)
31. Lindley, S.E., Harper, R., Sellen, A.: Desiring to be in touch in a changing communications landscape: attitudes of older adults. In: Proc. CHI 2009, pp. 1693–1702. ACM (2009)
32. Davidoff, S., Zimmerman, J., Dey, A.K.: How routine learners can support family coordination. In: Proc. CHI 2010, pp. 2461–2470. ACM (2010)
33. Subtext (2012), `http://www.subtext.com` (cited 2012 September 11, 2012)

The Mysterious Whiteboard

Clemens Nylandsted Klokmose and Olav W. Bertelsen

Department of Computer Science, Aarhus University, Denmark
{clemens,olavb}@cs.au.dk

Abstract. This paper raises the question of why electronic whiteboards are not ubiquitous. The paper provides a design-oriented analysis of traditional as well as electronic whiteboards in the context of collaborative and individual activities. We offer a novel perspective on whiteboards for collaborative activity based on a survey of the electronic whiteboard literature, a series of interviews with users of traditional whiteboards, and concepts rooted in Activity Theory. We identify a number of characteristics of the non-electronic whiteboard that are important to understand and preserve in the design of electronic whiteboard systems. Most importantly, we argue that the strength of non-electronic whiteboards is a combination of their simplicity and stability as well as a discontinuity between material on and outside of the whiteboard. We argue that the non-electronic whiteboard has uses and properties, which will require an electronic substitute to differ fundamentally in design compared to our traditional personal computing devices as well as most designs seen today. We present a set of themes for design of future electronic whiteboard systems that emphasize limitations as a main design principle. We conclude with three principles for design: The idea of installation rather than application; the principle of supplementing rather than replacing; and finally the principle of embracing and enhancing discontinuities.

Keywords: Electronic whiteboards, non-personal computing, activity theory, discontinuities, collaboration, science.

1 Introduction

Whiteboards and blackboards are ubiquitous in lectures halls, offices, laboratories and meeting rooms in most teaching institutions and workplaces. This suggests that whiteboards[1] have some generally valuable inherent qualities. However, whiteboards have obvious physical limitations in the form of space, persistence over time, and the possible content they can hold. Therefore, since it was feasible to size up an adequately advanced user interface to mimic the use of the whiteboard to the approximate scale of a board, researchers and industry have tried to develop hardware and software to replace the physical whiteboard. While two of Mark Weiser's famous trinity of devices (Weiser, 1991) *have* become ubiquitous, namely the *tab* (in the form of smartphones) and the *pad* (in the form of tablets) *boards* or electronic whiteboards

[1] For simplicity we use the term *whiteboard* to refer both blackboards and whiteboards.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 37–54, 2013.

are not ubiquitous today 20+ years after some of the first experiments with board sized computing took place at Xerox PARC. It is easy to argue that the hardware is not quite yet there; that a 100” slim, high-res display with low power consumption and state of the art multi-touch and pen interaction is currently not a reality within a standard company or educational budget. Yet, there are areas where electronic whiteboards, with their current limitations, supplement or replace physical whiteboards successfully. SMART boards[2] and similar technologies have been particularly popular as presenter tools in schools[3]. In hospitals we find examples of successfully replacing whiteboards used for coordination with interactive displays (Bardram, 2010).

We hypothesize that it is not only because of hardware limitations that electronic whiteboards are not ubiquitous in our workspaces, offices, laboratories and meeting rooms. Hardware limitations should not blind us from the fact that the activities involving whiteboards have fundamental differences from the kind of activities we traditionally support with personal computers. For pads and tabs to become a success as general purpose personal computers required a whole new ecology of operating systems, user interfaces and applications – an ecology geared towards the particularities of those devices. However, whiteboards are not personal. It would be naïve to believe that interactive surfaces ubiquitously hanging on our walls replacing or supplementing physical whiteboards will be a reality before a software ecology akin to what we have seen evolve over the recent years for pads and tablets will emerge.

In this paper, we develop a new set of design-oriented perspectives on electronic as well as non-electronic whiteboards. We do so based on our reading of the history of research into electronic whiteboards, in combination with a series of recent interviews with scientists on their whiteboard usage. We conceptualize the whiteboard, its characteristics, qualities, and its role in praxis through activity theory. We collect a set of themes, each with a number of challenges, we believe will be essential to address in a future ecology of software for electronic whiteboards. Our focus is the whiteboard as a tool for individual and collaborative work in offices and laboratories instead of the whiteboard as a presenter tool in the classroom.

2 Whiteboard Systems and Research

2.1 Whiteboard Systems

Much of the seminal work on electronic whiteboards took place at Xerox PARC in the late eighties and early nineties. *Colab* (Stefik et al., 1987) was one of the first systems that directly tried to imitate the qualities of whiteboards. Colab explored the interplay between desktop computers and a wall-display under the paradigm of WYSIWIS (“*What You See Is What I See*”). Colab contained three applications

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

[3] For a critical review of electronic whiteboard use in schools see e.g. Smith et al. (2005) or Slay et al. (2008).

reflecting traditional meeting activities: Creating presentations, outlining arguments and sketching. However, interaction in Colab was still mouse and keyboard based from desktop computers in the meeting room. Commune (Bly & Minneman, 1990), another PARC project, would link multiple horizontally oriented interactive drawing surfaces together over network to explore collaborative drawing and sketching. The *Liveboard* project (Elrod et al., 1992) introduced electronic whiteboards similar to what we see in classrooms today, and the Liveboards played the role as the third device of the initial ubicomp prototypes. The Liveboards ran a software suite called *Boardwalk* that provided a whiteboard application with free hand drawing support, and the ability to handle multiple *sheets* through a flip-chart analogy. Hence, drawings on the electronic whiteboard could be stored and reloaded. Findings in the project indicated how an electronic whiteboard is a very different beast than traditional desktop computers. For example it was observed that the Liveboards would not be used regularly if the users had to turn them on. Hence the PARC security staff was instructed to turn them on every morning. Secondly that users were less inclined to spend time solving software related problems when at the board: *"While users may tolerate the intricacies of UNIX in the privacy of their offices, we have found that in group settings, people are much less willing to take the time to solve software mysteries"* (Elrod et al., 1992, pp. 606). *Tivoli* (Pedersen et al., 1993) was a software project running on the Liveboards, and a continuation of the whiteboard application of BoardWalk. Tivoli explored different interaction techniques suitable for electronic whiteboards such as gestures and was designed to allow up to three users to interact simultaneously at the same board. Building on the previous experience with the BoardWalk software, Tivoli deliberately did not provide handwriting recognition, as it was assumed that this would be disruptive in a meeting context and make the users self-conscious. A central goal with Tivoli was to balance functionality and simplicity that would allow walk-up-and-use by novice users–which turned out to be a major challenge. *Flatland* (Mynatt et al., 1999) explored a "computationally-enhanced whiteboard" where the system would perform automatic actions on the content on the electronic whiteboard such as auto-segmentation and let the user rearrange and resize these segments. The user could apply *behaviors* to the segments to for instance create a ToDo list that would allow checking off items, or to write traditional handwritten calculation and apply a calculator behavior that will compute the result, and update the result if the calculation changed. Flatland would also let the users explore the history of the content on the board by dragging a slider back and forth in time. Unfortunately the effect of the ideas in Flatland in actual use was never documented.

The PARC electronic whiteboard systems targeted non-domain specific support for meeting activities. *Knight* (Damm et al., 2000) on the other hand was specifically developed for object-oriented design and diagramming. Knight allowed users to mix informal sketches and formal UML diagrams on an electronic whiteboard, the latter through recognition of hand-drawn UML elements. Hence acknowledging the need for moving between the formal and informal in a design situation. This mix seemed to have a positive impact on diagramming object-oriented systems. *Calico* (Mangano et al., 2010) is, like Knight, an electronic whiteboard system for software developers.

Yet instead of supporting more formal diagramming, Calico supports sketching of software, and instead of recognizing shapes as UML elements, Calico allows its users to define their own drawing primitives called *scraps*. Calico groups strokes similar to Flatland and supports multiple canvases, similar to Tivoli, yet Calico provides a fixed set of spatially organized canvases that the user can zoom out and get an overview of. A comparative study by Mangano et al. (2010) indicates that Calico does support a software design activity better than a traditional whiteboard.

One good example of structured whiteboards traditionally used for institutionalized planning and coordination can be found in hospitals. Recently, at least in numerous Danish hospitals, clinical whiteboards have been replaced with large networked interactive displays (Bardram et al. 2010). These kinds of electronic whiteboard systems are very different from the general-purpose electronic whiteboards developed at PARC. Each screen is tailored specifically for a given activity in the context of a given hospital for instance to show an interactive overview of the occupancy and association of clinical personnel to e.g. a patient ward. Configuration and reconfiguration of these displays involve highly trained consultants.

Another type of electronic whiteboard systems is whiteboard capture systems that rely on a traditional physical whiteboard, but use digital cameras to capture the content. ZombieBoard (Saund, 1999) applies a ceiling mounted digital pan/tilt camera to capture the content of two whiteboards. Writing special marks on the whiteboard controls the capture mechanism, e.g. to choose whiteboard to capture or area of a board to scan. ReBoard (Branham et al., 2010) is another example of a capture system but where ZombieBoard relies on the user to explicitly control the capture mechanism through markings on the board, Reboard captures everything that is put on the whiteboard and makes it available through a web interface. Branham et al. studied the effect of fitting ReBoard to a group of whiteboard users' whiteboards, and observed how in fact practice changed. Users become less afraid of deleting content from their whiteboards and would more often wipe the board clean, knowing that everything was stored. Yet they also observed how more care was taken in what was written on the boards, as users knew they might retrieve it and share it with colleagues in the future; *"ReBoard can thus be seen as creating tension between the ephemerality and persistence of board content"* (Branham et al., 2010, pp. 82).

Interestingly the application area of the majority of the electronic whiteboard systems described above were meeting situations and collaborative work activities – yet in present day commercialized electronic whiteboard systems are mainly used in teaching as a presenter tool connected to a personal computer such as the teachers laptop.

2.2 Studies of Whiteboard Use

Historically blackboards and especially whiteboards are a relatively new invention[4], however throughout the 20th century they have become a ubiquitous instrument for knowledge transfer and knowledge creation. The literature on *how* blackboards and whiteboards are actually used, *why* they are used and *what* they are used for is

[4] The class-room chalkboard was invented in the early 19th century by James Pillans (e.g. mentioned in Pillans (1856)) and the whiteboard was invented in the 50s and started to appear on the market in the 60s.

relatively sparse. The first electronic whiteboard systems papers did not report on studies of actual whiteboard and blackboard use to inform the design of the systems, but relied on casual observations and common knowledge regarding the use of whiteboards. In the last decade there have, however, been a handful of HCI papers studying how whiteboards actually are used to inform the design of electronic whiteboards.

Mynatt (1999) studied the daily whiteboard use in offices over a two-week period through snapshots of the boards and interviews with the users. The goal of the study was to capture the affordances of physical whiteboards under the assumption that an electronic whiteboard will fail if such affordances are not transferred. She especially focuses on long-term use rather than shorter-term use for meetings and teaching.

Mynatt observes how whiteboards in general are used as a *working space* rather than a *production space* as one of her interviewees puts it. The content of whiteboards is typically highly context dependent and getting the content off the board e.g. to share information from the board following a discussion is a source of frustration. Mynatt observes how the content of whiteboards clusters, and a whiteboard can be involved in multiple activities simultaneously, she observes how people scavenge space leading to some boards with hotspots that change a lot while other content is preserved. Mynatt emphasize how space constraints are a weakness of whiteboards, yet that traditional desktop interfaces with multiple documents wouldn't be appropriate for electronic whiteboards. However, the conclusions of the paper does not go beyond stating that developing an electronic whiteboard system is difficult.

Tang et al. (2009) study how people employ whiteboards to transition between related tasks. Through the use of surveys and interviews the authors observe how the use of whiteboards transcends the boundaries of a classical 2x2 groupware matrix of independent/collaborative – synchronous/asynchronous use. The content of whiteboards relate to multiple activities, and the authors claim that whiteboards support transitions between independent and collaborative activities. The authors identify four types of whiteboards through their study: public, semi-public, personal, and notification whiteboards. Public whiteboards belongs to no one and is primarily used for synchronous activities. They are most often wiped clean after use. Semi-public whiteboards, e.g. in a shared office, exhibit similar use as a public one, yet they are sometimes used for storage and a common praxis for writing on them may be established. Personal whiteboards, located e.g. in a single office, is used for a large variety of tasks and only close colleagues are invited to write on the board. Notification whiteboards are purely personal used for reminders or task lists.

Tang et al. summarizes their findings with a set of implications for design of electronic whiteboards:

- Whiteboard practice is largely enabled by the conception of whiteboards as contextually located containers for visually accessible information
- Providing users with expressive primitives will allow them to flexibly generate meaningful applications themselves
- Supporting transitions on electronic whiteboards means designing functional primitives rather than applications
- Designers can rely on the situated nature of interactive displays to determine which primitives are appropriate for that context.

Summing up: Activities at a whiteboard are very different from activities at a desk. The content of whiteboards can be part of multiple activities, often collaborative and what whiteboards are used for is very content dependent. Mynatt refer to whiteboards as *working spaces* rather than *production spaces*.

3 The Interviews

To get a better understanding of actual use of whiteboards we conducted a series of situated open-ended interviews applying what we refer to as “hit-and-run ethnography”.

Table 1. The type of whiteboards the interviewees describe they use. Office whiteboards are indicated personal/shared when they are located in a personal office but used collaboratively with colleagues in extended time periods. The majority of the interviewees describe public whiteboards in meeting- and classrooms hence these are omitted from the table.

	Office	*Laboratory*	*Other*
Interviewee 1	Personal	-	-
Interviewee 2	Personal/Shared	-	-
Interviewee 3	Personal	-	-
Interviewee 4	2 x Personal	-	-
Interviewee 5	Shared	-	-
Interviewee 6	-	-	Shared/Public
Interviewee 7	Shared	Shared	-
Interviewee 8	-	-	-
Interviewee 9	Personal/Shared	-	-
Interviewee 10	-	Shared	-
Interviewee 11	-	Shared	-
Interviewee 12	Personal	Shared	-
Interviewee 13	Shared	-	-
Interviewee 14	-	-	Shared

The interviews were conducted at the campus of Aarhus University where we approached colleagues at their offices and laboratories, asking them to show us their whiteboards and participate in an interview about their use of them. We did not announce the interviews before hand. We had prepared a short interview guide and a flyer informing about our project and us. We visited the departments of Physics, Mathematics, Geoscience, Bioscience, and Computer Science. The colleagues were positive and inviting. Only in a few cases, e.g. when a person was giving a lecture shortly, they opted out. The interviews lasted between 10 and 20 minutes.

In total we interviewed 14 people about their use of whiteboards, all fulltime employed at Aarhus University. We initially asked them to describe the whiteboards they used in their daily work.

Table 1 summarizes the types of whiteboards the interviewees described they used. Interviewee 6 mentioned a shared semi-public whiteboard in a hallway coffee lounge, and interviewee 14 mentioned a shared whiteboard onboard of a marine biology research ship (both indicated as other). Interviewee 8 was unique in reporting that he didn't use whiteboards at all in his daily work except occasionally for teaching. Interviewee 10 and 11 were interviewed in their laboratories, while the rest in their offices. *None* of the interviewees mentioned using electronic whiteboards when we asked about which whiteboards they had in their work environment, or when we at the end of the interview asked them about what they would want whiteboards to be in the ideal world.

3.1 The Use of the Whiteboards

A number of our interviewees used whiteboards in a very traditional fashion. The whiteboards of interviewees 1, 4, 7 and 9 were used with math heavy notation, typically collaboratively with colleagues or students discussing a proof of a theorem, a mathematical argument or data analysis (example in Figure 1). Interviewees 2, 3 and 5 used their whiteboards for traditional meeting activities such as project planning or paper outlining. Interviewee 12 used it purely personally and kept very neat and aesthetically pleasing notes on the board representing ongoing work. Interviewee 13 shared a whiteboard with his colleague and office mate. The whiteboard was used to represent ongoing experiments with notes and printouts of plots held by magnets. The whiteboards described by interviewee 10 and 11 exhibited the most elaborate use. The whiteboards were located at experimental setups in a physics lab and displayed information relevant to ongoing experiments and the equipment. The whiteboard of interviewee 11 was extremely dense with information and we were told the overall structure of the information on the whiteboard hadn't changed for more than 6 years, while it was updated on a daily basis (Figure 2).

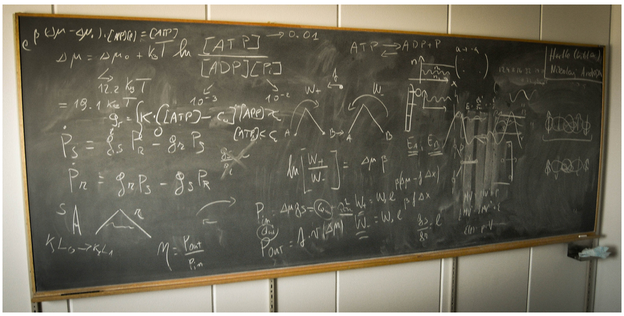


Fig. 1. The "whiteboard" of interviewee 9, a theoretical physicist

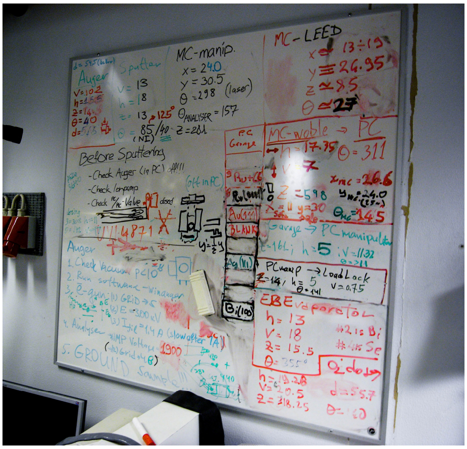


Fig. 2. The whiteboard of interviewee 11 located at an experimental setup in a physics laboratory

Two categories of use of whiteboards emerge from the interviews. The whiteboard as *information container*, where the main function is to show information with some purpose, and as *graphical workspace* where the main function is to develop, or transform something. We saw two aspects of the content of a graphical workspace; as *ephemeral externalization*, supporting collaborative creativity and problem solving, or as *in production*, where the results of the activity is seamlessly sustained onto stable physical form, a product. We saw information displays that could be described as simple information *storage*, information displays that aggregated *knowledge* and information displays where the content was part of a *coordination* of activity or *governance* of surrounding state.

Interviewee 9, who used the whiteboard for theoretical physics (as seen in figure 1), described how what was written on the board was not always directly post-processed e.g. in a written note but instead the result was a new understanding. Interviewee 7 described how he would write out an equation from a textbook to review and discuss it with students. The content was in both cases ephemeral; after it had been written and discussed it could be deleted. In both cases the board is used as a graphical workspace, but the pattern of usage is as *ephemeral externalization*. Interviewee 9 describes how something that is written on the board *"… either it's something that is clarified and then it is stored up here (points at his head) until it is written in a paper. Otherwise the idea is elaborated on a piece of paper, where it can turn into a note on 2-3 pages …"*. Interviewee 12 described how notes on a whiteboard could later be used and written into a paper, and interviewee 3 described how papers would be outlined and project plans sketched on the board that could feed into actual papers and project plans. The transition was either handled by manually writing down notes from the whiteboard or more often by taking a photo with a smartphone and then typing up the notes at a later time. Here the content of the whiteboard represents an early stage of a product of some sort. Hence we refer to this use pattern as being *in production*.

In the case of interviewee 10 and 11, the whiteboard was used as an information display, however, displaying information that in some form was available other places (equipment displays or laboratory notebooks) yet it was made readily available on the whiteboard near the experimental setup. Hence, we refer to this as *knowledge aggregation*. Interviewee 10 explained: *"… so I guess we kind of mostly use it as a… mostly for reference and also for keeping track of some, where samples are in the equipment. So eh, yeah obviously we have labbooks, but then as the years go on we have a big stack of these and somewhere hidden in them is some very important information like these diagrams (points to the whiteboard) [unintelligible] and eh, it becomes really a pain to look for these things in the labbooks"*. Figure 2 also displays information pertaining to the state of the experimental setup, e.g. the placement of samples in compartments and it is essential that when the whiteboard is updated, so is the physical setup and vice versa. We refer to this aspect as *governance*.

The other type of information display is using the whiteboard for storing information – this may be phone numbers (interviewee 7) or a list of who has borrowed books from the bookshelf (interviewee 1). Interviewee 2 describes how currently on his board is a project brainstorm that was written 3 weeks earlier in a meeting, and that they continuously have returned to and extended in follow-up meetings. Hence, the whiteboard was used in a combination of production, information storage and coordination of a project activity. This implies that the categorization is not static. In fact we also observe how an equipment sketch on interviewee 11's whiteboard started out as an ephemeral externalization but remained stored on the whiteboard to explain newcomers about the equipment. This is very well in line with Tang et al.'s (2009) observations that whiteboards serve as means for transitioning between activities.

The electronic whiteboard systems in the literature (see section 2) have an almost exclusive focus on the whiteboard as a graphical workspace used in production. While this is an important aspect of the use of whiteboards we see in our interviews that it is not the complete picture. The clinical electronic whiteboards (Bardram 2010) are an exception in that they are information displays used by the clinical staff for governance of e.g. the state of a ward and coordination of clinical activities.

4 Demystifying the Whiteboard

To demystify and conceptualize the use of whiteboards, we will turn to concepts derived from Activity Theory. An important characteristic of the whiteboard is that content is not transferred seamlessly to and from it. Bertelsen and Bødker (2002) discuss design of computer artifacts in terms of *discontinuities*, pointing to three classes of discontinuities. The first one is between experience and desire and is the fundamental discontinuity in design between what exists and has been experienced, and what is to become. In the context of whiteboards and electronic whiteboards the two next discontinuities are mostly relevant. Between parallel rooms, is a discontinuity between differing perspectives on, or purposes of, an object (or artifact). A representation on a whiteboard, typically has an existence before it gets onto the whiteboard as well as after. On the whiteboard emphasis may be on finding new solutions, while when the representation later is transferred into a document aimed to be shared and archived, exact conformance to conventional formalism may be the most important aspect of the representation. Between interpretation and implementation is the particular discontinuity between a formalized representation interpreted in a social setting, and the same representation when executed by a machine. In the later case we are dealing with strict causal relations between representations and changed state of a part of the world whereas the representation prescribes more loosely in the former. *"It is the discontinuity between de-contextualised principles and concrete historical practice; between artefact and situation; between technical implementation and cultural interpretation."* (ibid pp. 410-411). These discontinuities are seen as important resources that should not merely be bridged but rather be maintained and cultivated in design. And we will in the context of this paper argue that support for persistent discontinuity is an important quality of the whiteboard.

Historically, information technologies have blurred and removed discontinuities. While it is a clear advantage of modern computer based text production tools that separate typesetting and re-typing of manuscripts from version to version can be avoided, it is also problematic that parts of a text, with early phrasings etc. can survive from an early draft to the final publication.

When using a whiteboard something is actively transformed into writing/drawing on the board. This transformation is discontinuous. An idea takes an initial form, a formula from a textbook is rewritten and transformed into joint basis for a teaching situation, and the terrain of a problem at hand is sketched for the purpose of the situation. After the whiteboarding event, results may be carried away in memory, or they may be transformed into notes, or even refined and turned into a product for further processing. In all cases, again, we see a discontinuity.

There is a fundamental discontinuity between parallel rooms when using a white board; you go into the "whiteboarding room" where there are limitations to how much detail you can have, and there are limits to how much damage you can make because everything will be transformed later. The discontinuity between interpretation and implementation is also important to understand in relation to the whiteboard. When working with mathematical formalisms, several interviewees (e.g. 4 and 9) would use ad hoc notations hoping that they were not taking too many "liberties". Some time that would work and some times it would not. The important thing to observe is that the discontinuity between the writing/sketching of free ideas on the board and something that conforms to the formalisms, or is a direct implementation outside the board is what enables creativity and exploration on the whiteboard.

Thus, in design of electronic whiteboard systems, it is important to also maintain or redesign the right discontinuities.

Our classification of whiteboard uses can be rephrased in terms of the relation to the surroundings and the involved discontinuities.

Information storage does not depend on maintained discontinuity. A phone number, or information of who borrowed a book, should be complete.

Ephemeral externalization involves a quite abrupt purpose-driven discontinuity into the whiteboard, where only parts of formalisms or even ad hoc formalism is noted and worked with, similarly there is most often a substantial amount of processing following the whiteboarding session.

In the cases of *in production, coordination* and *governance* the relation in and out of the whiteboard is more complicated because both cases depend also on some degree of continuity. For these cases we need to understand the dual character of whiteboard writing as being both objects worked on and tools mediating action on something inside or outside the whiteboard.

Mynatt (1999) refers to the whiteboard as a working space rather than a production space. While we agree with Mynatt to a certain degree, the whiteboard can in fact be part of a production space, but what is written or drawn on the whiteboard almost never is the final product; an outline of an argument has to be written into actual text and a sketch of a figure has to be reproduced for print. This is very different from most modern computing where we constantly work on the actual "products" whether written documents, graphics, presentations or videos without these discontinuities.

In Activity Theoretical thinking the driving force for development are contradictions. Activity systems are permeated by contradictions, e.g. between the tool at hand and the object of work. Thereby, contradictions are constant sources of instability and development (Bertelsen & Bødker, 2001). In understanding collaborative inscription in artifacts Zander (2007) introduces the concept of *syntonic seeds* as inscriptions sublating contradictions by negating and containing them. Thus, the reminders placed around the lab of interviewee 11 could be understood as memory aids negating the contradiction between what should be collectively remembered about the experimental setup and the settings necessarily being hidden behind insulation, yet preserving attention to this contradiction. This kind of containment resembles how contradictions are maintained across a discontinuity. Interviewee 9 kept a sketch of an argument on his whiteboard to remind him to continue the discussion, thereby negating the insufficiency of attention.

According to Tang et al. (2009), Mynatt (1999), and our own studies, whiteboard content is typically pivotal in a number of activities. Tang et al. describe how a team-leader's whiteboard mediates planning, awareness and team discussions, similar to how the whiteboard in the laboratory of interviewee 11 serves as a means for representing the ongoing experiment, documenting it, and communicating the praxis in the laboratory. In the case of the team-leader the content of the whiteboard, a project plan, sublates a contradiction between a normal human's mental capabilities and the complexity of a big project, but at the same time it may sublate the contradiction between her and co-workers understanding of the progress in the project. With inspiration from the concept of syntonic seeds (Zander op cit.), we identify three key aspects of the whiteboard.

- The writing on the whiteboard has a double character containing (and possibly sublating) a contradiction between the form of representation on the whiteboard and outside the whiteboard.
- The contents is simultaneously and/or sequentially a mediator in one or many activities and an object of one or many activities

- Contents often oscillate between being a mediator and object of activity. To return to the previous example; the project plan oscillates between being the object of the team-leaders documentation of the project and being a mediator in discussion with the team.

Mynatt (1999) observes that the content on whiteboards is heavily context dependent. This means, in our vocabulary, that the content exists in its relation to something outside the whiteboard – something that is corrected, controlled, externalized. The content on the whiteboard is most often serving its purpose due to the discontinuity between whiteboard and other representations. The whiteboard in figure 2 may serve as nothing but an archeological artifact if the lab is shut down, only hinting at the kind of activities that took place in the lab. To move the content beyond the whiteboard requires re-mediation or requires the content to be de-contextualized e.g. when taking a sketch of an argument from a whiteboard and writing it into a paper.

The whiteboard is a *discontinuous artifact*; information has to be remediated both to and from the board. This is both a strength and a weakness. It is a strength as it provides an opportunity for critical reflection on what has been produced or is about to be (re)produced on the board. It is a weakness when the discontinuity is a source of error, e.g. when the whiteboard is used for governing critical state but updating it is forgotten. It is furthermore an inherent weakness that the discontinuity to and from the board may be a laborious for the user as it involves reproduction, and given that the content is context dependent; information may be lost in this process (e.g. when photographing and wiping a whiteboard clean after a meeting and not processing the content into written notes or similar right away).

As for electronic whiteboards there is a significant difference in whether the board is a self-contained computer or it is merely a display and input device for a personal computer. The latter is the current way of using electronic whiteboards; a teacher connects her personal computer to the electronic whiteboard in the classroom and uses its capabilities to make ready-made material interactive and dialogical e.g. through highlighting, annotation or hide and reveal (Mercer et al., 2010). This approach bridges the discontinuity to and from the whiteboard; content prepared on the personal computer can directly be presented on the electronic whiteboard and what changes are made during a presentation can be stored immediately and be distributed and reused. However, the electronic whiteboard goes black when the teacher disconnects the computer and leaves the room, and data is not shared beyond the teacher letting the students manipulate the content she makes available on the board for a limited amount of time.

5 Themes for Design of Future Electronic Whiteboard Systems

In this section we outline four themes that we find important to understand and consider in relation to the design of future electronic whiteboard systems. We discuss them in the light of our empirical findings, the literature and the conceptual basis of activity theory. The themes we want to point to are lifecycle of content, transition into and out of whiteboards, the relation between formalism and immediacy, and the impact of digital materiality.

5.1 Lifecycle

We saw a variety of content lifecycles on the whiteboards. Somebody would write contents onto the board, and this content would serve a purpose for a while until it finally would die by being whipped out or simply loosing its meaning. Some of the whiteboards we saw had data on them that had been there for a long time, but the general picture was that data lived a shorter time.

In the office of interviewee 9, the whiteboard had short lifecycles by functioning mainly as an ephemeral externalization during collaborative problem solving. According to the interviewee, writing was only left on the board after a session, because he would not spend the effort to erase it. However, writings from a session that did not lead to the desired solution weeks earlier were still “surviving” on the board, possibly because they now served as a continuous reminder about this unsolved problem. In that way a seamless transformation from ephemeral externalization to a kind of information display had taken place.

A much longer lifecycle is seen in the case of the lab whiteboard of interviewee 11. It has been persistent over years, with a layout that has been added to over time. It is used for making things explicit, for coordination, and for conveying rules and established practice, as well as for recording and sharing the state of the ongoing experiment. Learning to work with the whiteboard is important when new students and new colleagues get to learn to work in the lab.

At the state of the lab whiteboard as we saw it, it was mainly a container in all the three aspects of that. When we visited the lab the whiteboard functioned as container in all the four forms. It was mostly for governance and coordination, by displaying important parameters and states of the experimental setup, rules of conduct and relevant constants. It was, to a large extent knowledge display by the presentations of experimental set up and overall mindset of the lab. To some extent it was information storage where parameters were stored. In a longer perspective, contents on the lab whiteboard had evolved or oscillated between the different roles described in 3.1. An example is the black drawing approximately in the middle of figure 2. It was originally created as an ephemeral externalization when explaining the experiment to a visitor, but it remained on the board as a knowledge display.

Whiteboard use is constantly reconstituted and negotiated. In particular it is important to observe the seemingly seamless transition of content between the roles. The physical form remains the same but the status and interpretation changes radically, and thereby the flexibility depends on the lack of hardwired formalization leaving openness to interpretation.

The lifetime of whiteboard contents seems to be characterized by seamless transitions between interpretations. However, most contents die after a while, and has something been deleted it can never return to the whiteboard in the exact same form.

5.2 To and from

We argue above that the whiteboard offers two main discontinuities that are fruitful for a broad range of purposes. The whiteboard normally has no structure, no limitation and no support for interpretation. When content gets written onto the whiteboard a process of selection, abstraction and reduction takes place and a mutual understanding between collaborators is established. In the same way, processing takes place when information is transformed out of the board. A horizontal line can mean the duration of a project, or something else. In the context of a specific session the meaning is sufficiently clear, but after the session it will depend on participants memory, interpretation and mapping back onto a more universal format. In other words, a process of *de-contextualizing* is involved with the transfer out of the whiteboard. Such de-contextualization can happen by writing the work on the whiteboard into some formalized notation, or by transforming it into coherent prose. It is important to acknowledge the value of these discontinuous transformations even if they seem to be time consuming and a source of inaccuracy.

In some cases a formal link between whiteboard content and the surrounding world can be established. In the lab of interviewee 11, we saw how the experimental setup was recorded on the board. Particularly, we saw the governance aspect of information containers, and to some extent the knowledge container could be extended into a live data connection. In the same way, it could make sense to offer support for the analysis of data sets on the electronic whiteboard. Conventions for the linkage between data in the electronic whiteboard and their counterpart in other media have to be settled for each area of application. In some cases it may be fruitful to have live data in the whiteboard. In other cases data should be static.

At the other extreme we saw examples of knowledge containers, serving merely as reminders, being constructed by mounting paper onto the whiteboard, thereby circumventing the discontinuity into the whiteboard. E.g., interviewee 2 had an early version of a project plan mounted on his board, interviewee 14 had a print of the first data from the experiment he and the colleague in the office were running.

Many of the interviewees reported that they would take pictures of their whiteboards, typically with their phone cameras. It was not clear, however, how these snapshots were used beyond being backups, or records for the unlikely case that the content could be used later.

5.3 Immediacy and Formalism

It is important that the installation is ready and running. Otherwise the overhead will make users go to other media to support the work they want to do. We saw good examples of this in the literature (e.g. Pedersen 1992). In particular for the large part of the interviews where uses could be characterized as ephemeral externalization, it seems important that there is no startup time, otherwise it would not be feasible when they “just had to scribble something down…”.

This need for simplicity and the reluctance to accept any overhead was expressed by Interviewee 11 when he wished for a better organized layout of the lab whiteboard (fig 2). He did not hesitate to explain principles for a better layout. Still he did not seem to be willing to spend one hour re-drawing the board. *"Physically, so there is no reason why, so if I had time I would move things around in a different way because like there are things that you need more frequently and you would like to have them there and things that you have less frequently you would like to have them like on the corner or somewhere else."* (Interviewee 11, #00:14:28-5#)

In situations where a formalized notation exists, it would make sense to support the transition from ephemeral externalization to graphical workspace in production more systematically like in the Knight system (Damm et al. 2000). This transition could also be supported simply by supporting the transformation of free form graphics into structured graphics for further processing.

5.4 Digital Materiality

In ordinary whiteboards content is bound to a physically existing board. With electronic whiteboard systems it would be possible to move the content around between physical boards, as well as replicating it between boards. While portable workspaces could be an advantage, it is worth drawing attention to the way in which limitations, context dependency, and discontinuity are important features of the whiteboards, as we know them. Content on whiteboards continuously change roles and evolve over time–hence the traditional approach of well-defined file formats of digital data may not make sense. It may be necessary to revisit ideas like compound documents (as OpenDoc (Macbride & Susser, 1996)). Yet supporting governance of external state (e.g. something simple as controlling and displaying the angle of a motorized laser in an experimental setup through a widget) from the whiteboard would bring the digital whiteboard content beyond just being documents.

6 Design Principles to Explore

In the following we present three design principles for future electronic whiteboard systems.

6.1 Installation, Not Application

We observe that all our interviewees have idiosyncratic, yet over time established, ways, of using their whiteboards. Their whiteboards are physically bound to rooms where certain activities take place. If a goal is to maintain the immediacy of the physical whiteboard when creating an electronic counterpart we discourage thinking of the board as a general-purpose computer with multiple applications. An important reason for that is the general reluctance to spend time on setting up and negotiation observed in the literature and in our studies.

We suggest an alternative design strategy based on the concept of *installation*, meaning that the software running on the electronic whiteboard is permanently configured, and that it supports seamless invocation of the active components.

This strategy is useful for supporting specific praxes where institutionalized formalisms exist. Both theoretical physicists and software architects need boards for sketching, yet the kind of augmented support they could benefit from differ. The software architects may want support for turning free from drawings into formalized diagrams with possibilities for grouping and auto-rearrangement functionality. The theoretical physicists, on the other hand, may want handwriting detection and notation completion.

People dedicate certain whiteboards and certain areas of whiteboards to specific kinds of activities, and the equivalent should be possible in an electronic counterpart. This is the approach of the medical electronic whiteboard systems as described e.g. by Bardram (2010). The tools and information available on a board in an emergency ward differs significantly from what is available in the patient ward since the activities the boards are embedded in are different. Tang et al. (2009) implies a similar design philosophy of providing the users with functional primitives and using the contextualized nature of the whiteboards to decide which to use. However, it is an open question who will take care of the (re)configuration. In the case of the medical electronic whiteboards it is handled by trained consultants with insight into the clinical praxis, however such an approach would not necessarily be possible or economically feasible in a research setting.

The concept of installation, in particular the whole economy and division of labor around creating, installing and updating, should be investigated further and experimented with.

6.2 Supplement and Augment, Don't Replace

It is difficult to see how an electronic counterpart can reach the immediacy and ease of use of a physical whiteboard especially when it comes to supporting rapid ephemeral externalizations. Yet, as we see in our study, what was perceived to be an ephemeral externalization may change its role and become something more permanent. In the physics laboratory of interviewee 11 it would be meaningful to have a combination of a traditional whiteboard with a capture mechanism alike ReBoard (Branham et al., 2010), together with a digital surface for displaying sensor readings or storing and displaying (digitalized) drawings made on the physical whiteboard. An interactive electronic display is meaningful in the case where the information have a governing nature, since it will allow to electronically enforce a mapping that otherwise had to be done by hand.

6.3 Embrace Discontinuity

Discontinuity in some situations is a strength, yet remediating content from the whiteboard is laborious work. Our interviewees frequently photographed whiteboards to persist the content. Our hypothesis is that photographing the content of a whiteboard for later remediation may be a disservice to one self, because knowing that the content is persisted safely induces a reluctance to actually deal with it. We encourage exploring the design space of a capture mechanism of physical whiteboards where captured content only is stored for a limited amount of time, and the user is reminded that she has unprocessed whiteboard content waiting to be remediated. This means creating artificial limitations on the medium, which we believe is necessary to maintain one of the central qualities of a whiteboard namely that when content is erased it is gone.

7 Conclusion

Returning to Mark Weiser's three devices: Tabs, pads are here but they required a new software ecology to be successful – one would assume the same of the last device, the wall size device.

In this paper we have reported on a study of whiteboard use in a scientific research setting. We have observed a range of different types of use, and analyzed how discontinuity is an important aspect of much successful use of whiteboards. We have analyzed our findings based on activity theory, and the existing literature and identify 6 roles contents on whiteboards can have.

Our main finding is that the discontinuities in and out of the whiteboard can be a strength. This finding implies that the focus on seamless integration seen in large parts of the literature may need to be modified or supplemented by context specific concepts of the value of breaking the continuity of objects being worked on.

We have pointed to four themes to consider in future design of electronic whiteboard systems. Firstly, we have discussed how content on whiteboards shifts seamlessly between functions and roles during its *lifetime*. Thereby, flexibility and openness to adaptation of notation is important. Secondly, we point to the transformations of content *to and from* the whiteboard, in particular we point to how content is contextualized on the whiteboard and how transferring it out of the board involves de-contextualization. Thirdly, we point to the combined issue of *immediacy and formalism*, emphasizing that a most important aspect of the whiteboard is its constant availability, and that we will have to take that kind of immediacy into account when supporting formalized notation. Finally, we draw attention to the necessary consideration of what kinds *digital materiality* to support.

We conclude the paper by introducing three principles to explore in future electronic whiteboard systems. Firstly, we suggest that software on electronic whiteboards should be understood as an *installation*, i.e. a stable configuration that is readily available and only changes with very long intervals. Secondly, we suggest that designers should aim to *supplement and augment* the existing non-electronic whiteboards, rather than risk loosing the intricate patterns of use. Finally, we suggest that designers should aim to *embrace discontinuities* offered in the use of non-electronic whiteboards, and distinguish between unwanted and wanted ones.

Acknowledgements. Thank to colleagues at the Faculty of Science and Technology, Aarhus University, who served as informants. Thanks to James Eagan, Susanne Bødker and Liam Bannon for comments and discussions. This work has been supported by the Carlsberg Foundation and the Participatory Information Tehnology Centre.

References

1. Bardram, J.E., Hansen, T.R.: Peri-operative Coordination and Communication Systems: A Case of CSCW in Medical Informatics. In: Workshop on CSCW Research in Healthcare: Past, Present, and Future (2010)
2. Bertelsen, O.W., Bødker, S.: Cooperation in massively distributed information spaces. In: Proc. ECSCW 2001(2001)
3. Bertelsen, O.W., Bødker, S.: Discontinuities. In: Floyd, C., Dittrich, Y., Klischewski, R. (eds.) Social Thinking - Software Practice. MIT press, Cambridge (2002)
4. Bly, S.A., Minneman, S.L.: Commune: a shared drawing surface. In: Lochovsky, F.H., Allen, R.B. (eds.) Proceedings of the ACM SIGOIS and IEEE CS TC-OA Conference on Office Information Systems (COCS 1990), pp. 184–192. ACM, New York (1990)
5. Branham, S., Golovchinsky, G., Carter, S., Biehl, J.T.: Let's go from the whiteboard: supporting transitions in work through whiteboard capture and reuse. In: Proc. CHI 2010, pp. 75–84 (2010)
6. Damm, C.H., Hansen, K.M., Thomsen, M.: Tool support for cooperative object-oriented design: gesture based modeling on an electronic whiteboard. In: Proc. CHI 2000, pp. 518–525 (2000)
7. Elrod, S., Bruce, R., Gold, R., Goldberg, D., Halasz, F., Janssen, W., Lee, D., McCall, K., Pedersen, E., Pier, K., Tang, J., Welch, B.: Liveboard: a large interactive display supporting group meetings, presentations, & remote collaboration. In: Proc. CHI 1992, pp. 599–607 (1992)
8. MacBride, A., Susser, J.: Byte guide to OpenDoc. Osborne McGraw-Hill, Berkeley (1996)
9. Mangano, N., Baker, A., Dempsey, M., Navarro, E., van der Hoek, A.: Software design sketching with calico. Proc. ASE 2010, 23–32 (2010)
10. Mercer, N., Gillen, J., Staarman, K., Littleton, K., Twiner, A.: Interactive Whiteboards: Does New Technology Transform Teaching? In: Lud, L., Rasmussen, S. (eds.) Learning Across Sites: New Tools, Infrastructures and Practices. Routledge (2010)
11. Mynatt, E.D.: The writing on the wall. In: Proc. INTERACT 1999 (1999)
12. Mynatt, E.D., Igarashi, T., Edwards, W.K., LaMarca, A.: Flatland: new dimensions in office whiteboards. In: Proc. CHI 1999, pp. 346–353 (1999)
13. Pedersen, E.R., McCall, K., Moran, T.P., Halasz, F.G.: Tivoli: an electronic whiteboard for informal workgroup meetings. In: Proc. INTERCHI 1993, pp. 391–398 (1993)
14. Pillans, J.: Contributions to the Cause of Education. Longman, Brown, Green, & Longmans. London, UK (1856)
15. Saund, E.: Bringing the Marks on a Whiteboard to Electronic Life. In: Yuan, F., Hartkopf, V. (eds.) CoBuild 1999. LNCS, vol. 1670, pp. 69–78. Springer, Heidelberg (1999)
16. Slay, H., Siebörger, I., Hodgkinson-Williams, C.: Interactive whiteboards: Real beauty or just "lipstick"? Comput. Educ. 51(3), 1321–1341 (2008)
17. Smith, H.J., Higgins, S., Wall, K., Miller, J.: Interactive whiteboards: boon or bandwagon? A critical review of the literature. Journal of Computer Assisted Learning 21(2), 91–101 (2005)
18. Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S., Suchman, L.: Beyond the chalkboard: computer support for collaboration and problem solving in meetings. Commun. ACM 30(1), 32–47 (1987)
19. Tang, A., Lanir, J., Greenberg, S., Fels, S.: Supporting transitions in work. In: Proc GROUP 2009, pp. 149–158 (2009)
20. Weiser, M.: The computer for the 21st century. Scientific American 265(3), 94–104 (1991)
21. Zander, P.-O.: Collaborative Process Change by Inscription. Lund, KFS i Lund AB (2007)

A Tabletop System Using Infrared Image Recognition for Multi-user Identification

Shota Suto[1] and Susumu Shibusawa [2]

[1] Ibaraki University, Graduate School of Science and Engineering, Nakanarusawa 4-12-1, Hitachi, Ibaraki 316-8511, Japan
12nd304x@hcs.ibaraki.ac.jp
[2] Ibaraki University, Dept. of Computer and Information Sciences, Nakanarusawa 4-12-1, Hitachi, Ibaraki 316-8511, Japan
sibusawa@mx.ibaraki.ac.jp

Abstract. Many tabletop systems have been developed to facilitate face-to-face collaboration and work at small meetings. These systems often require users to attach sensors to their bodies to identify their positions, but attaching a sensor to one's body can be bothersome and annoying, and user position and posture may be restricted depending on where the sensor is attached. We have proposed a technique for estimating user position in a tabletop system by image recognition and implemented a tabletop system having a user position identification function incorporating the proposed technique. This technique first obtains touch points and hand-area information from touch operations performed by the user, and establishes an association between the touch points and hand from those positional relationships. Since the direction in which a hand is extended can be derived from that hand's touch information, the position of the user of the touch points belonging to that hand can be estimated. As part of this study, we also implemented a photo-object manipulation application, which has a function for orienting a photo object to face the user based on the results of the above user-position estimation technique. We performed an experiment to evaluate the position identification rate, and found that the proposed technique could identify user position with high accuracy.

Keywords: Tabletop system, Image recognition, FTIR, Multi-touch, User position identification, Area extraction.

1 Introduction

A tabletop system facilitates face-to-face collaboration at small meetings or any setting where people come together to work as a group. It enables all users to manipulate or examine displayed information and enables a variety of input operations to be executed by touching displayed information and performing gestures. Research surrounding tabletop systems with these features has been active including studies of interactive techniques [1], [2], development of information display technologies [3], [4], and the use of tabletop systems for supporting collaborative work [5], [6], [7].

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 55–62, 2013.

In a tabletop system, objects can be oriented in various ways by touch gestures performed by multiple users, which means that a user may find it difficult to understand the text or photo of an object that is currently not facing in the user's direction. In response to this problem, several techniques have been developed to identify user positions by attaching sensors to chairs or the users themselves [8], [9], [10], and these techniques are used to automatically modify the orientation of objects according to the position of the user manipulating the objects. For example, Diamond Touch [9], registers beforehand each user and the user's position by having each user sit on a conductive sheet for user-identification purposes, and uses this information to determine the position of the user whenever the user is identified. However, identifying user position by attaching sensors to chairs or people can be troublesome, and time must be devoted to learning how to use sensor equipment. User posture may also be restricted depending on where the sensor is attached.

On the other hand, research has been performed on an interactive system that extracts images of body extremities using image recognition technology so that physical movements performed by the user can be used as input operations [11]. This kind of interactive system using image recognition negates the need for wearing a sensor thereby enabling users to use the system in a free and natural manner.

We propose a technique for estimating user position by image recognition in a tabletop system and construct a tabletop system incorporating this technique. This technique negates the need for wearing a sensor and removes restrictions depending on where the sensor is attached. This system uses a frustrated total internal reflection (FTIR) multi-touch panel and obtains touch points and hand-area information from user touch operations using image recognition. It establishes an association between the touch points and hand from those positional relationships. Since the direction in which a hand is extended can be derived from that hand's area information, the position of the user of the touch points belonging to that hand can be estimated. This study includes the implementation of a photo-object manipulation application that enables users to manipulate photo objects on the tabletop system by touch operations. The application has a function for orientating a photo object to face the user according to the results of estimating user position from touch gestures.

2 Proposed System

2.1 System Configuration

For this study, we designed a tabletop system with a multi-touch panel capable of touch-point recognition by the FTIR method [12]. This system features an infrared Web camera installed beneath the table to capture infrared images on the tabletop by picking up the acrylic panel on the table via a mirror. The system also includes a projector connected to a personal computer (PC) to display photo objects to users by projecting images onto the acrylic panel from underneath the table. In addition, tracing paper is pasted onto the acrylic panel to act as a screen and users are presented with information by having the projector project images onto this tracing paper. System configuration is shown in Figure 1.

Finally, to obtain information on hand area, an infrared light is installed on the ceiling above the table. Since a hand on the tabletop will block infrared beams emitted from this light, an infrared shadow corresponding to the hand will form. The Web camera is used to pick up this shadow effect.

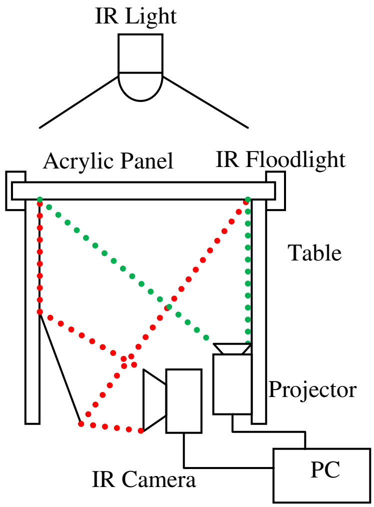


Fig. 1. System configuration

2.2 Overview of User-Position Estimation Technique

In a tabletop system, a user typically extends a hand from the edge of the table to manipulate some object. Accordingly, if it were possible to determine the direction from which the hand including touch points is being extended, the position of the user generating those touch points could be estimated.

In FTIR-based touch-point recognition as used here, the system picks up touch points as white light owing to the diffused reflection of infrared light. It also picks up the area occupied by a hand as a shadow generated by the blocking of light emitted from the infrared light installed on the ceiling. At this time, the system uses brightness values in the captured image to differentiate from the background image in two ways. Specifically, it extracts touch areas having a higher brightness value than the background and an area having a change in brightness value as a hand area. This extracted hand area includes the touch areas as a subset. Thus, when focusing on certain touch points, a hand area that includes those touch areas as a subset certainly exists, which means that the direction from which that hand is being extended can be determined from that hand-area information. As a result, the position of the user associated with certain touch points can be estimated since touch points and user position can be indirectly connected through the direction of extension of the hand to which those touch points belong. Extraction of these key areas and estimation of user position are outlined in Figure 2.

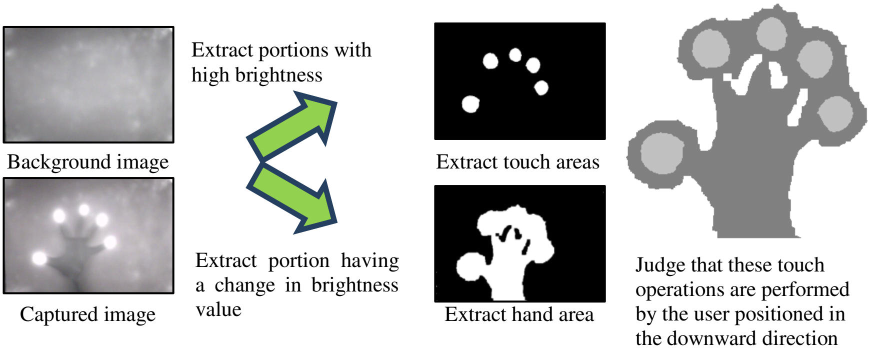


Fig. 2. Extraction of key areas and estimation of user position

2.3 User-Position Estimation Model

The FTIR touch-point recognition system picks up touch points as white light and the hand area as a shadow of infrared light. Establishing a difference with the background image enables the touch areas and the hand area that includes those touch areas to be extracted. An image of a touch area and that of a hand area extracted by background differentiation are shown in Figures 3 (a) and (b), respectively.

(a) Image of a touch area (b) Image of a hand area (c) Positional relationship among, *TA*, *TP*, and *HS*

Fig. 3. Extraction of touch-point and hand areas and their positional relationship

The positional relationship among touch area (*TA*), touch point (*TP*), and hand shadow (*HS*) is shown in Figure 3 (c), where the outer circle and inner circle at the fingertip of *HS* corresponds to *TA* and *TP*, respectively. This technique determines *TP* by calculating the center of gravity of *TA*. When we define hand-area *Hand* as the union of *TA* and *HS*, *TP* is an element of *Hand*.

The proposed technique first investigates the attribution relation between *TP* and *Hand*. It next determines from which edge in the image the *Hand* to which *TP* belongs to is extending. It finally estimates user position by drawing a correspondence between that direction and the touch point.

2.4 User-Position Estimation Technique

More than one hand area may exist on the tabletop at any one time. To recognize individual hand areas, the technique labels areas having connected pixels and assigns the label *L* to each hand area.

Then, when focusing on a certain touch point *TP*, that a hand area *Hand* that includes the coordinates of *TP* as an element exists. The results of labeling can therefore be used to extract label $L[TP(x, y)]$ corresponding to the coordinates of *TP*. Here, $L[TP(x, y)]$ is equivalent to label $L[Hand]$ including *TP*. The hand area corresponding to the targeted touch point can therefore be identified by referring to $L[TP(x, y)]$.

Referring to Figure 3(b), it can be seen that *Hand* consists of a continuous area connected to a certain edge of the image. Accordingly, if that edge can be determined, the direction from which *Hand* is being extended can likewise be determined.

3 System Implementation

Our prototype tabletop system has a height of 70 cm, a panel 100 cm × 90 cm in size, and a display manipulation range of 60 cm × 50 cm. The tabletop is shown in Figure 4.

We also implemented an application for manipulating photo objects by touch gestures. This application reads in image data as photo objects and displays them on the tabletop system. It treats touch gestures made by the user as input and generates results in response to those gestures. Touch gestures are listed and described in Table 1. Change direction is an operation which makes the object face the user when touching it with three fingers. An example of the change-direction gesture is shown in Figure 5.

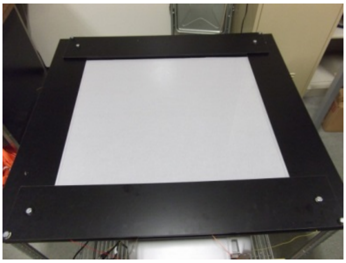


Fig. 4. View of tabletop

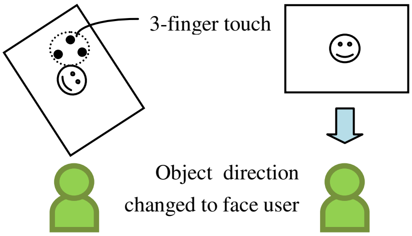


Fig. 5. Change-direction gesture

Table 1. Touch gestures

Operation	No. of Touches	Description
Move	1	Move object
Zoom in/out	2	Change object size
Rotate		Rotate object
Change direction	3	Change object's direction to face user

4 Evaluation Experiment

4.1 User-Identification Evaluation Experiment

To assess the accuracy of the user-position estimation technique, we conducted a subjective experiment using the 3-finger change-direction gesture. After briefing four subjects on how to perform this gesture, we asked each to perform the gesture ten times on the tabletop system in each of the up, down, left, and right directions. At this time, we recorded actual user position and system-estimated position in order and determined position identification rate by comparing the values obtained. Specifically, denoting the number of times this gesture was performed as d_{act} and the number of times that actual user position agreed with system-estimated position as $d_{correct}$, we computed position identification rate by Eq. (1).

$$\text{position identification rate} = \frac{d_{correct}}{d_{act}} \times 100\ [\%]. \quad (1)$$

The experiment was performed in the evening after sunset considering the possible effects of infrared light contained in sunlight. The prototype tabletop was installed in the center of a room and two infrared lights were installed on the ceiling above the table. Each light was 90-cm long incorporating six equally spaced infrared LEDs and the two lights were set 70 cm apart. The distance from the ceiling to the tabletop panel was 185 cm.

4.2 Results and Discussion

Average identification rate for the change-direction gesture by four subjects for each of the four tabletop directions and overall average for all directions are shown in Figure 6.

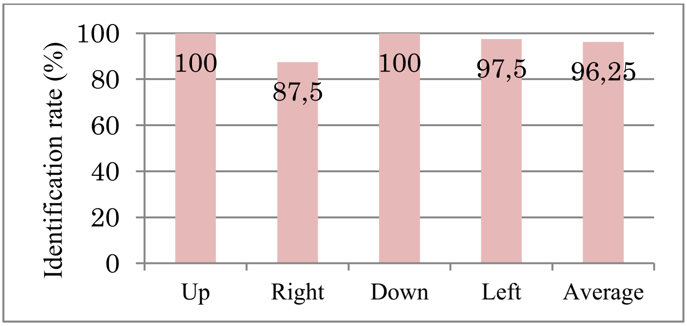
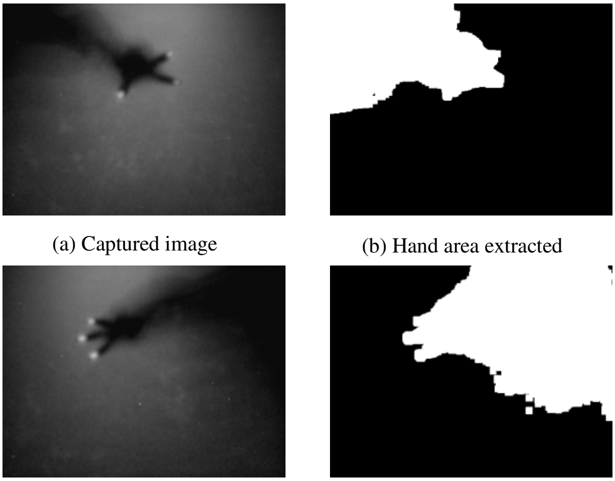


Fig. 6. Identification rate for change-direction gesture

Average identification rate for the change-direction gesture was found to be about 96%. Examining these results, it can be seen that hand extension from the up and down tabletop directions could be accurately identified. There were times, however, when hand extension from the left and right tabletop directions could not be accurately identified. Examples of incorrect hand-extension identification from the left and right directions are shown in Figure 7. Images for a change-direction gesture from the left are shown in Figures 7 (a) and (b) and those for a change-direction gesture from the right are shown in Figures 7 (c) and (d).



(a) Captured image (b) Hand area extracted

(c) Captured image (d) Hand area extracted

Fig. 7. Change-direction gesture from left and right directions

In the examples of Figure 7, the number of pixels in the hand shadow area crossing an edge are higher in the up direction than in the left and right directions. As a result, the system erroneously judges those tabletop operations to be those of a user positioned at the top of the tabletop (in the up direction). This is because the touch-point coordinates on the captured image cannot necessarily be used to correctly obtain the touch-point coordinates on the acrylic panel.

5 Conclusion

We proposed a technique for identifying user position in a tabletop system through image recognition and introduced a prototype tabletop system incorporating this technique. We also implemented a photo-object manipulation application for running on the tabletop and conducted a subjective experiment to evaluate the accuracy of the proposed technique in estimating user position. In this experiment, we recorded actual user position and system-estimated position when subjects made change-direction gestures and calculated the position identification rate by determining the rate of agreement between actual and estimated positions. We found that an average position identification rate of about 96% could be achieved in the case of change-direction gestures, which demonstrates that user position can be estimated by the proposed image recognition technique. The technique is a simple technique for identifying the user position and negates the need for wearing a sensor.

Errors in identifying user position occurred in the system as a result of camera image compensation. An offset between the manipulation area and display area on the tabletop prevented the system from correctly recognizing how a hand shadow area was crossing an edge making it easy for erroneous judgments to occur. We considered that using a wide-angle camera would be able to capture the entire tabletop panel head-on, which should eliminate complicated image-compensation processing by the system and suppress this phenomenon. In future research, we plan to evaluate the position identification rate during simultaneous operations by multiple users and to add gesture functions making full use of user position identification. We also look to apply the proposed technique to face-to-face collaborative work systems and to evaluate its usability in such an application.

References

1. Clifton, P., Mazalek, P., Sanford, J., Rébola, C., Lee, S., Powell, N.: SketchTop: Design Collaboration on a Multi-touch Tabletop. In: TEI 2011, pp. 333–336 (2011)
2. Haller, M., Brandl, P., Leithinger, D., Leitner, J., Seifried, T., Billinghurst, M.: Shared Design Space: Sketching Ideas Using Digital Pens and a Large Augmented Tabletop Setup. In: Advances in Artificial Reality and Tele-Existence, pp. 185–196 (2006)
3. Yamaguchi, T., Nacenta, M., Sakurai, S., Itoh, Y., Kitamura, Y., Subramanian, S., Gutwin, C., Kishino, F.: An Evaluation of Perspective Correction Regarding Locations of a User and Displays. IEICE Trans. on Information and System J91-D(12), 2746–2754 (2008)
4. Sakurai, S., Kitamura, Y., Subramanian, S.: A Tabletop Display with Visibility Control Using Revolving Polarizer. IPSJ Journal 50(1), 332–343 (2009)
5. Funamoto, S., Sugiyama, K.: Support and Evaluation for Group Discussions in Seminar Type Lectures Using Table Top Interface. In: Proc. of the 6th Knowledge Creation Support Systems Symposium, pp. 119–126 (2009)
6. Ohashi, M., Itou, J., Munemori, J., Matsushita, M., Matsuda, M.: Development and Application of Idea Generation Support System Using Table-Top Interface. IPSJ Journal 49(1), 105–115 (2008)
7. Goh, W.-B., Shou, W., Tan, J., Jackson Lum, G.T.: Interaction Design Patterns for Multi-touch Tabletop Collaborative Games. In: CHI EA 2012, pp. 141–150 (2012)
8. Watanabe, K., Takeuchi, T., Inoue, T., Okada, K.: Face-to-Face Collaboration System that Identifies Individual User. IEICE Transactions J91-D(12), 2755–2764 (2008)
9. Dietz, P., Leigh, D.: DiamondTouch: A Multiuser Touch Technology. In: UIST 2001, pp. 219–226 (2001)
10. Annett, M., Grossman, T., Wigdor, D., Fitzmaurice, G.: Medusa: A Proximity-ware Multi-touch Tabletop. In: UIST 2011, pp. 337–346 (2011)
11. Mamiya, H., Sato, T., Fukuchi, K., Koike, H.: Tabletop Entertainment System with a Finger Opening and Closing Gesture. In: WISS 2007 (2007)
12. Han, J.Y.: Low-cost Multi-touch Sensing Through Frustrated Total Internal Reflection. In: UIST 2005, pp. 115–118 (2005)

Performing Online and Offline: How DJs Use Social Networks

Mayur Karnik, Ian Oakley, and Valentina Nisi

Madeira Interactive Technologies Institute, University of Madeira
Campus da Penteada, Funchal, Portugal
{mayurkarnik,ian.r.oakley,valentina.nisi}gmail.com

Abstract. Music and online Social Network Sites (SNS) are closely intertwined in popular culture, but we know relatively little about how performers use and take advantage of such social systems. This paper investigates this space by exploring how professional DJs leverage SNS in their work. It adopts a long term mixed-methods ethnographic approach encompassing semi-structured interviews, supported by studio visits and participant observations. Results revealed that DJs used SNS for connecting to their audiences; promoting their work; receiving peer feedback; discovering content and keeping abreast of their field; and organizing and coordinating events. We further interpret our findings in the context of issues DJs highlighted about their professional practice and technology, and our observations; and draw out design implications for future music orientated systems and services.

Keywords: Ethnography, DJs, social networks, participant observations.

1 Introduction

Social Network Sites (SNS) are a key part of the success of modern performance artists, helping small or emerging acts connect with fans, advertise their shows and provide access to their media [1]. Rapid growth in SNS usage along with technological advancements in computational power, mobile technologies and network bandwidth, present these artists with new opportunities and challenges. For example, unlike in the past when music artists were relatively inaccessible to fans and media, today they directly seek engagement with their fans and patrons using a number of channels available through SNS [2]. Indeed, MySpace, one of the earliest SNS to achieve popular critical mass is strongly associated with its role as a platform for bands to promote themselves [3]. Conversely, music audiences today use multiple and inter-linked sites online, reflecting the changing ‘shapes’ of online communities [4]. It comes as no surprise, therefore, that the most watched and liked content on YouTube today belong to musical phenomena such as Justin Bieber and PSY (Gangnam Style) [5], both arguably products of a fame that has been powered largely by social media (e.g. in the form of tribute videos). Easy access to technologies aimed at content creation and distribution has popularized the technique of ‘micro-celebrity’ [6,7], which is the practice of using social media to develop and maintain an audience.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 63–80, 2013.

Furthermore, a growing ecosystem of companies is beginning to provide dedicated social software tools to help performers understand, maintain and manage audience interactions [8]. However, despite the prevalence and importance of this issue for performers, there has been relatively little academic work studying the needs and practices involved in this user group's adoption and appropriation of social technologies. This paper aims to address a specific area of this gap by describing a mixed-methods ethnographic study of a group of resident DJs. The study focuses on their use of technology in their everyday lives, and in particular to their professional use of SNS.

We argue that DJs, solo performers who participate in an almost inevitably local and heavily technology driven popular art form, are a particularly interesting user group for this study. We focus on DJs for a variety of reasons. Firstly, we observe that DJs, with their electronic and digital instruments, have a more direct and deeper engagement with technology, as compared to traditional musicians or bands. Secondly, DJs have to constantly depend on social engagement, not only offline with their audience during their performances [9], but also online with other DJs and for promoting their work in order to build and maintain their profiles as artists [10]. Thirdly, DJs rely on discovery of music to feed their work rather than the more traditional creative and individual activities that other types of artists typically engage in. Furthermore, they need to be sensitive to trends, styles and techniques not only in a single genre, but across a range of types of music, as required by the different gigs they play; thus implying the need to be externally focused and engage with a range of different communities to stay current. These characteristics make DJs an interesting and important group to study in order to explore the socio-technical gaps in the context of artists. Studying such a tech-savvy and socially active segment of artists can have wider implications for our understanding of the social aspects of music [11], and further lead to insights into the wider HCI issues faced by artists and cultural workers.

This claim is supported by the richness of activity captured in Ahmed et al.'s [10] recent ethnographic study of DJs at work, that highlighted the overall complexity of the work practices, which go considerably beyond performance to encompass activities such as collecting and collating musical libraries through to preparing specific gigs and set pieces, to engaging in intensive publicity and promotion. They briefly touch upon the role of SNS in a DJ's promotional activities and highlight some of the issues that SNS present. However, we argue for a more in-depth examination of DJs' use of SNS through an understanding of their motivations for using SNS and the concomitant challenges. The current paper thus extends their insightful analysis with an exclusive focus on, and a more detailed study of, the role of social networks in the work of DJs.

In sum, the goal of the paper is to provide a deeper understanding of the role of SNS in a DJ's everyday life. It is structured as follows: a brief overview of related work; description of mixed-methods employed towards a long-term study of a small community of DJs; a description of the findings, focusing particularly on the motivations for SNS use and issues germane to their practice; followed by a discussion leading to design implications for future music orientated social systems.

2 Related Work

Music plays an active role in the construction of our everyday social lives [12]. It lends us means to express, understand and share our identities [13]. Tanaka [14] argues that social computing coupled with artistic creativity can point out 'ways in which technological evolution can be assimilated directly in cultural production, ultimately leading to possible new forms of musical content.' Drawing on Goffman's seminal work on social interaction [15], he compares early conceptions of human-machine interaction with musical ensemble performance and suggests that music, as a cultural practice, has the potential to contribute to a deeper understanding of interaction [14].

HCI researchers have described a number of studies of, and systems for, the DJ community. Spurred on by the advent of the Internet, smartphones and the digitization of music, early research in this domain focused on tools and audiences. Gates et al. [9] classify these initial works as interactive nightclub technologies that were essentially audience-centered applications, DJ-centered applications, and applications for interaction between the DJs and their audiences. These applications broadly attempted to take advantage of sensors, mobile devices and communication technologies. They took the form of playful applications, performative spaces, automation and mixing tools, and systems based on bio-feedback [9]. Ahmed et al. [10] present an overview of more recent studies and prototypes that employed a variety of modalities (e.g. wireless, mobile, haptic, and multi-touch) but were mainly limited to the nightclub setting.

More recently, however, researchers have begun focusing on ethnographic approaches in order to understand the DJs' overall context of work. Gates et al. [9] focused on how DJs gathered information about, and interacted with their audience in nightclubs, while Ahmed et al. [10] give a detailed account of how the DJ's work extended beyond performance to include 'the work to make the performance work': collecting, preparing, performing and promoting. They found DJs to be active on several SNS and forums in order to promote themselves and acquire bookings, but faced general challenges in integrating these activities. Many more traditional social aspects of SNS were reportedly viewed with suspicion – basically as a source of biased (universally positive) feedback from fans.

This viewpoint is counter-intuitive – SNS have a long history as a tool for performers to reach audiences. For instance, the early success of MySpace was strongly linked to its role as a platform for bands to promote themselves and engage with their fans and audience [3]. Beer argues how the presence of the musicians catalyzed social connection amongst fans on MySpace [16]. Sargent [17] carried ethnographic studies of local musicians in Charlotesville and found that active social networking was crucial to their music distribution and promotion. Her studies demonstrate how musicians cultivated audiences through social capital both online and offline. Music fans have been found to use multiple and multi-linked network sites as Baym's early work on Swedish independent music fans demonstrates [4]. More recently, Baym explored online music audiences from the perspective of the artists and presents a rich overview of research on fans and fandom [2]. Similarly, Marwick and boyd [7] discussed how social media is utilized by public figures to maintain the practice of celebrity despite the underlying power tensions of these interactions. Researchers have proposed online systems that enable DJs to broadcast their work [18] and shown ways in which the activities of performers rely on digital resources and social networks [19]. Furthermore, Facebook, the currently dominant SNS, recently launched a 'listen' feature that enables performers to easily share content from their pages [20]. On the other hand, Ping, a music-based SNS from industry heavyweight Apple, recently announced its closure [21]. The diversity of these experiences and features suggest that the interaction between performers, SNS and audience is complex, multi-level and poorly understood.

The remainder of this paper aims to contribute to understanding of this issue by presenting an ethnographic study of a small community of DJs, with a specific focus on their use and engagement with SNS.

3 Methods

A mixed-methods approach was adopted in order to develop a nuanced understanding of the social milieu of a DJ's everyday life. This comprised of 1.5 to 2 hour semi-structured interviews conducted with 6 DJs supported by six to eight observational site visits to the clubs and environments in which these DJs performed and four visits, each spanning three to four hours, to the personal home studio of one of the DJs. Taken together, these activities allowed the capture of a detailed picture of both the private and public environments in which the DJ's work took place. Additionally, as part of a process of participant observation [22, 23], the primary author of this paper took on the role of an apprentice or amateur DJ / VJ and performed several gigs with three of the DJs over a period of one year. This extent of participation was motivated primarily by the need to get a closer access to an otherwise inaccessible group and secondarily owing to the primary author's own interest and experience in audio-visual performances. These activities are unpacked and described in detail in the following sections.

3.1 Semi-structured Interviews

The interviews were conducted between March 2012 and June 2012 with a representative group of six DJs working in Funchal, a small and relatively isolated city (approximately 150,000 residents). Five held residencies, or permanent performance arrangements, with major local clubs. Aged between 27 and 35, their professional experience spanned six years (three individuals), ten years (one individual) and 17 years (two individuals). They were also highly engaged with the wider aspects of musical culture - three authored or produced music while one had previously owned a record store and served as a partner in a small record label. They were not provided with compensation for their time.

The semi-structured interviews explored three questions:

- What were their daily routines like in the week and during the weekends?
- What technological artifacts did they use, personally and professionally, in their day-to-day lives and why?
- What social networks did they use, why and how?

At the end of the interview, respondents were encouraged to make more open-ended comments. The interviews led to more than 10 hours of data and were transcribed verbatim. In line with Hook et al.'s [24] process, the transcripts were then subjected to Thematic Analysis [25].

3.2 Studio Visits

One of the DJs volunteered to participate in a more in-depth study. Two visits were made to his home studio, followed by visits to his work sessions at his resident nightclub, followed by further two visits to his home studio. This process was conceived to provide detailed insights into his full workflow, studio environment and the overlap between his personal and professional life. Photos and videos were taken during these sessions for later analysis. Care was taken to keep these recording activities discrete to ensure it had minimal interference or influence on behaviors.

3.3 Participant Observation

Overt participant observations of the DJs, conducted with informed consent, were made periodically over a one-year period. These observations took place at performance venues and spanned a full range of work activities: preparing, warming up, performing and transferring the decks to a follow-up act or retiring for the night. The observations were recorded through photos, videos and notes. These observations were further supported by club owners who provided weekly images (through their Facebook Pages) for a period of several months. Extending these efforts, the primary researcher adopted the role of an amateur DJ / VJ in order to connect with the community, gain a deeper understanding of the performers' context and engage more empathetically with the participants and their roles. Such truly participatory observation is a valuable and established part of ethnographic practice. In this case, it involved making seven performances (four as a DJ, three as a VJ) in different venues with three of the DJs over a period of one year.

4 Findings

The interviews were transcribed and subjected to a process of inductive Thematic Analysis [25] using TAMS Analyzer, an openly available qualitative analysis software tool. A list of meaningful units were extracted from all the transcripts and subjected to an inductive coding process, resulting in a list of 85 codes. These codes were further clustered and organized into different categories that reflected the various subjects that the respondents talked about, for example SNS, performances, routines, personal time, etc. Owing to the focus of the study, SNS emerged as a higher-level category and we further scanned through these items and extracted insights that comprised of motivations for use and challenges encountered.

In the results section below, we first summarize and discuss these insights, giving particular attention to data about SNS, in conjunction with insights from our participant observations. Due to the scope of this paper, we limit our reporting of our observations to the social dynamics embedded in their work environments.

4.1 Routines

As expected, the routines of DJs differed for weekdays and weekends. While weekends were reserved for performing at their respective club venues, weekdays were characterized by social life, leisure and secondary work activities such as searching for new content, making music, preparing gigs and networking. Weekend work routines typically started and finished late, causing a sleep pattern that extended to their entire week. Owing to these erratic patterns, the DJs adopted ad-hoc routines through the week, moving between preparing for weekends and producing music; reflecting an inclination to maintain flexibility necessary for a creative profession.

MC: *"So I dedicate lots of time researching the tracks to get music… and then it really depends on my mood because I make music as well. Some days I am not feeling particularly inspired. So I dedicate my day to research only… some other day I wake up and think I am going to make music, so I will dedicate my time (only) to that."*

However this would often lead to frustration over a perceived lack of control and consequently to attempts at getting more organized.

FL: *"Because I work at night, my sleeping schedule is a little bit... I don't know... a bit different, mixed up. I work on the weekends, but on the daily weekdays I simply cant just go to bed at midnight… I can't I can't! I just can't sleep... I sleep at 2 or 3 in the morning, that's the best I can do."*

G: *"I used to wake up at 3 o clock in the afternoon, but now I learnt to wake up at 8."*

4.2 Artifacts

A wide variety of artifacts were used in work. Many were technological; all the DJs performed with computers and used various storage media (CDs, memory sticks, hard-drives) to host their music. Dedicated accessories (turntables, mixers) were widely used to control and access their digital libraries. Smartphones were used for communication (mail and SNS) and several participants also used music identification apps. Overall, DJs exhibited a high rate of acceptance to changes in technology as far as professional equipment was concerned. They appreciated the inter-operability of this equipment and particularly its easy integration with laptops via modern software applications catering directly to the tasks of live performance and music production.

F: "*I work with Abletone live… you can connect with various plugins, USB, you can get VHC; it's a world of music, an elephant!.. I have MIDI controller which is from M-Audio. I have all the DJ components… I use Traktor; last year I bought Traktor Controller S4 deck.. which you can connect with your computer and play which is great."*

MC: *"I used some samplers in the past. When I was playing vinyl or anything, I could record live what I was playing… like a short loop and would remix it live. If you have a song, you could transform it, remix it… if there was some part the crowd liked, I could record it and then play it over at the end of the show just to get a little creative. I used to use it like that. But as things were changing, there were new programs designed for playing for DJs… like Tractor, Native Instruments… there were programs where we could use MP3 to play. So I made this change 3 years ago. I use Traktor from NI and I use my Macintosh, my Mac Book. I have all my music so I connect my hard drive where I keep all my music."*

Respondents valued the practical performance related aspects of digitization of music and the attendant miniaturization of media formats, offering benefits such as portability, convenience during travelling for performances, and a higher degree of control over their planning.

H: *"5-6 years ago to go to a gig to [xx city] I had to carry 25 - 30 kilos of music. And then in the plane you can only take 20, every time you go you have to pay extra... with vinyls sometimes you arrive there, the bag doesn't arrive... the bags disappear and nowadays I can go only with my laptop, with few vinyls and it's done. These days, with ten kilos you can make a party."*

M: *"I carry also my Macintosh everyday because I have my projects, my music, my plans all inside my Mac. When I go for gig, my Macintosh. I don't carry music physically, I don't carry CDs or vinyls, I play with computer, Traktor... I don't like to carry a lot of things. I am very practical; I don't need a lot of things. I can give you an example; if I go on a trip, I carry light, pack light... two pairs of shoes, two pairs of trousers, two pairs of t-shirts."*

Social networks and peer communication had distinct associations with artifacts, with one respondent describing how friends he knew solely through social networks helped him not only in recommending a new equipment but also by buying and shipping it to him. Another respondent described at length an array of networking related apps that he carried on his smartphone. Almost all the respondents, barring one, owned smartphones and used the same for emails, social networking and social media consumption.

F: *"I like to share (about) new stuff, new hardware stuff with my friends... and they share as well; for example, I bought my keyboard via Internet… from some guys in Berlin; they ordered for me, that's the advantage of social network."*

FL: *"I work with a computer and software applications. I work with that and I have lots of music. If I play it all together, I will play for 5 or 6 days... I have a Blackberry with some apps for text messages, battery status... a player for YouTube, a chat for Facebook, Shazam the music guesser… the one that I really use is the Blackberry messenger, we can chat with BB community for free."*

4.3 Motivations for Using SNS

The DJs used social networks extensively for both personal and, primarily, professional reasons. All had accounts on Facebook, SoundCloud and Resident Advisor while 4 used Mixcloud, 2 used Twitter and 5 had, largely inactive, MySpace accounts. As local celebrities, they reported high numbers of friends on Facebook with a large fan base supplementing typical cohorts of family and “real” friends. Indeed, three of them maintained two or more Facebook accounts and four ran their own Facebook Page.

Our study reveals that DJs used SNS for interaction and connection, promotion, feedback, discovering and learning, and for organizing and coordinating events. A detailed discussion of their use of SNS is presented below.

Interaction and Connection. DJs used social networks to communicate with two specific groups: local fans and regular club patrons; and a set of remote artists, performers, fans and potential promoters in other cities or countries. This second activity, perhaps due to a lack of other channels to achieve it, was viewed as having higher value.

FL: *“First of all I want to share (my music) with people here in [city]... my first priority. I put them on my Facebook or [club]'s Facebook and the second objective is... it's the world of course, because I am on an island I have to reach the people that are outside this island.”*

DJs highlighted how they differentiated between fans and friends on their social networks. Fans were often courted through the DJs releasing their own content (either novel tracks or mixes) online or via recommendations and endorsements of other artists or songs. The DJs largely perceived these activities as part of a responsibility to educate and expose their audience to high quality new music.

H: *“On Facebook I have 50% people I know, others who know my work or know someone who knows my work or simply so many times it happens, you post a YouTube video or a set I made... 20-30 guys and add me only for that because they identify themselves with that... ok this guy to be my friend. At the end of the day, this thing of Facebook friends its not like... the friends word is not right, its not friends… It’s people who have same interests… the starting engine of friendship maybe, but not friends.”*

FL: “*The videos I upload on my Facebook through YouTube is to show the people some good music and try to teach them, because the nightlife and clubbing are a little bit of weird than what it used to be... the people don’t have the feeling they had a few years ago.”*

They also engaged in post-performance dialogues with audience members where they exchanged (almost universally) complements and tagged and commented on posts or photos. One DJ, however, pointed out how, on the downside, SNS also sometimes enable performers with better public relation skills to be more popular than the ones with talent.

MC: *"Before you had to 'do' something for people to like your work or you had to have great music that nobody else had… Or you had to be a really good selector or a really good mixer… but now you need to be... well, if you are a good public relation guy, you don't need to have talent or anything. It's just a bit annoying sometimes."*

DJs used dedicated features of music-centered SNS (e.g. SoundCloud) to share music and interact with performers playing in similar genres; exchanging content with peers was commonplace. SNS messaging complemented these media focused activities. One DJ noted that the widespread use of SNS has fostered and simplified a vibrant culture of "remixing" where DJs receive, reinterpret and re-release each other's content.

F: *"These days we are getting more and more remixes on the track because of the interaction of the people on the sites… DJs… these days you can make music, put it online and then share with other side of the world to make remix... ten fifteen years ago, we were listening to only single and original tracks... there are more remixes... its about the technology that allows you to share and collaborate with other artists."*

Promotion. The DJs valued SNS for promoting their work; such digital activity was reported to have simplified, reduced the cost of and, ultimately, replaced previous approaches such as distributing flyers or demo CDs. Specific activities they engaged in included posting flyers and schedules; maintaining shared music charts (on sites such as Resident Advisor) to showcase and recommend their work; and posting or linking to their media across a wide range of SNS services. Some of the older DJs had moved from music oriented SNS like MySpace to more generic ones like Facebook due to higher visitor traffic. Two DJs used the Facebook "Listen" feature [20] on their pages on that site and others shared content via third party plugins (e.g. Bandpage and MixCloud).

M: *"I used MySpace in the beginning, like 6 years ago... I used it a lot. Now I go there like once in a month. I put my gigs and not too much. I don't put too much information. Now I am using more Facebook. I have my 'Bandpage' and it works better than MySpace. I see the statistics from the page… like at MySpace after 6 years I have like 4000 or 5000 visitors, its what I have per month on Facebook. It's different."*

However, beyond these commercial services, some of them highlighted the importance of being present in other communication channels, for example most of them talked about the value of a personal web presence carrying a biography, schedule, sample media and collecting together all SNS contact details. Furthermore, four DJs played regular sets on local radio stations and one on a well-known Internet radio station (Proton). To support these activities, the DJs reported distributing podcasts of this material. Indeed, two reported broadcasting their work immediately after finishing a set using recording tools integrated into SoundCloud. Their need to cross integrate profiles and content between traditional channels on the Internet and social networks was quite evident in their responses.

Feedback. The DJs all reported that feedback on their work was a key motivation for using SNS. Detail, granularity and feedback from peers were especially valued. For instance, SoundCloud a pro-service that enables site visitors to comment on specific instances of a track was repeatedly praised. Often these comments would be made by other professionals in the form of suggestions about how the track could be improved.

H: *"On SoundCloud, you can have feedback in the very 'second' they like and sometimes its good if you have some work-in-progress some tracks… and the guys can say oh its good but over here you can change... you go to the studio and ok he has a good point of view, lets make this change."*

One DJ mentioned how he looked up the profile pages on SoundCloud to assess whether commenters on his material were "qualified" to provide a worthwhile critical opinion. Feedback from fans on sites such as Facebook, however, was viewed as invariably positive and, thus, of relatively low professional value, a finding that mirrors Ahmed et al. [10]. That said, the DJs reported this fan feedback was highly useful in gauging the quality of audience experience after a live performance.

The DJs highly valued analytics where these were available (e.g. Facebook Pages and SoundCloud). These typically offer numerical track-level feedback such as likes, plays and downloads but also highlighted the geographical demographics and patterns of specific user visits.

FL: *"I have 5000 plays on my SoundCloud. It's really a good tool, you have lots of statistics of things, my SoundCloud has been played in I don't know... about 50 countries, like India, your hometown I don't know... Venezuela, Argentina, the stats are very specified to that. And the fanpage on Facebook is really cool; you have access to everything. Everything! You posted a song, you can see how many plays it has, how many people see that, how many commented, liked and... everything. You can see the evolution of the fanpage on graphics... 100, 200, 2000, 3000, 4000… The number of people that for some reason unlike your page, you also have that one."*

However, meaningfully interpreting such trends was reported to be challenging and they were sometimes simply integrated into promotional activities: one DJ reported that his club tied special events to such statistics, throwing a party when they met a target for Facebook "likes".

Discovering and Learning. SNS also played a strong role in discovery of both content and tools. In terms of content, a range of practices was reported. These included monitoring online genre-specific charts, but also SNS activity such as trawling the walls and pages of other DJs. In such activities the perceived fame and influence of the DJs was an important criteria in assessing the value and worth of new tracks. It was also reported to be common practice to request specific track IDs and histories when commenting on online remixes, pinning down the sources and history of particular media content.

FL: *"The really good thing (on SoundCloud) is that people can interact in the middle of the music, leave a comment on minute 52 or 53 or minute one... ok this is great, track ID that kind* *of thing... you share information on that. It's really cool because you can always follow artists that are your inspiration or influence, you see them work and if they have a DJ set that had attracted you... you really, really like, you can go there and ask for the track ID, the DJ will answer or someone from the community that is following will answer too... that’s really, really cool.”*

In terms of tools, as mentioned earlier, DJs reported that hardware and software reviews and tutorials are widely posted on music SNS and such material was a key way to keep abreast of technical developments and innovations in the field.

Organizing and Coordinating. SNS such as Facebook were seen as a valuable tool for planning and organizing events. Beyond promotion of clubs, venues and parties via posting teaser sample content, the acceptance of event-invites and likes or comments was reported to provide valuable information on prospective attendance levels and enthusiasm for special themes. However, on the other hand, some found SNS suffering from a deluge of information about events and expressed frustration over how it was getting difficult to use for promoting local events.

H: *“Nowadays its too much, too many things... people use Facebook only to see friends... like you can ask people, they don’t care about the events because its too much, too many events... now its the opposite I think the best way to promote things is (by) speaking with the persons, calling up the person, giving him flyers.”*

5 Issues and Concerns

This section summarizes the wider concerns the respondents voiced about their profession. Some of these come across as issues germane to their profession that are, or have the potential to be, supported by SNS usage in some ways, whereas others serve as reflection of the perceived threats and shortcomings of existing social technologies. DJs highlighted a need to collectively tackle issues of curation, perceptions of threats about exclusivity, issues in cultivating and maintaining connections with peers, and a loss of mobility due to economic factors.

5.1 Need for Curation

Some of the DJs viewed themselves as curators of traditional formats on one hand and contemporary music on the other. One DJ reported how some of the older DJs like him were taking it upon themselves to preserve the analog legacy of vinyl records, citing nostalgia, sound quality and content exclusivity as legitimate drivers. Implicit in his call for this preservation was a need to collectively ‘keep things alive’.

H: “*If you go to the vinyl or the other formats you can be special… still special because so many records are released and never arrive at digital shops. So lots of people, normally more older DJs are keeping the things alive... like ok, we always say oh vinyl is going to die but no... we are the guys who have to keep it alive, we cannot be like this.”*

This nostalgia also carried forward to music itself, with some of the DJs commenting on early exposure to these cultural legacies. It's worth noting how the notion of curation ties in closely with old artifacts such as 'grandfather's records' or to old mu-sic genres. One of the respondents related a DJ's role to that of a teacher with responsibilities to educate people about 'taste'.

FL: *"When I was a kid, I grew up listening to my grandfather's records... Eddie Grant, Boney M, ABBA... I really have a... I really have a strong connection to the 80s."*

M: "*I think that the DJs also work as a teacher. A DJ has a responsibility to have a good musical taste and to show people what good music is about. When you go to the doctor, you don't want to teach him how to fix you."*

These concerns directly reflected in their sharing behavior on social networks. For example, they shared their own music on Facebook as well as music from artists or labels they liked and followed. Some of them also shared charts that they created every month on Resident Advisor. DJs who hosted radio shows would also share their playlists or podcasts later on SoundCloud and Facebook.

5.2 Threats to Exclusivity

Almost all the DJs expressed strong concerns about the lack of exclusivity caused by easy access to music online, especially via social networks and online music stores such as beatport.com. They perceived this as a threat to their music selection activity for performances, suggesting tensions between sharing and maintaining secrecy in their 'sources'.

MC: *"Nowadays, especially nowadays when everyone is a DJ, everyone is playing music, very difficult to stand out… when I say stand out everyone is playing the same track. If I get a track on beatport I have to understand that lots of other people will have it. So I dedicate lots of time researching the tracks to get music."*

M: "*I don't use social networks to do my music research, I use some sites from... about music, like beatport, djdownload.com, Juno records… but mostly I use the label sites like compact, moon records... then I use some magazines / reviews, some DJ reviews."*

In this context, as also noted by Ahmed et al. [10], DJs considered gifting as a valuable way to receive exclusive content. DJs who frequently shared music on their Facebook pages or had their own radio shows or podcasts regularly received large amounts of free music from other lesser-known artists who wanted exposure and feedback to their own work. One DJ commented on how his research involves more 'listening' to tracks sent to him via emails rather than looking for new artists or music on the Inter-net.

Concerns of exclusivity also overlapped with issues of professional identity. Readily available tools and equipment with more user-friendly features also brought into question the skills that defined a DJ as an entity. Some of them voiced concerns about how technology now enables anybody to assume the role of a DJ.

F: "*Everyone is a DJ nowadays. When I was young, the main thing was having a band. Everyone had a band and I was the DJ… and nowadays everyone is a DJ because it's easy. You can have an application on your iPhone to run to a speaker and be a DJ for one night or two nights.*"

5.3 Connecting with Peers

Technology also had a huge bearing on the ways in which the DJs researched and bought music changed over time. Two of the DJs who had been in the profession for more than 15 years narrated tales of how they used to buy music in the past by going through catalogues in magazines and calling up the record shops in other bigger cities. They would spend hours on the phone listening to one record after the other, as the shop owner manually played out samples of tracks. Although they valued the present day online music stores for the afforded convenience, they missed the personal recommendations and social aspects of buying music that physical record shops had.

H: *"One thing I miss a lot is we don't have here record shops where you can go and ... that is the really social network, because every DJ is going to be there and you can be there 5-6 hours chatting, exchanging ideas that is one thing we don't have… Recently I was in Amsterdam, I spent about 5 hours in one record shop and I met 3 to 4 people I didn't know before, only because of the music… 'Oh you are going to buy this, if you like this also try this'… I think it's social, the big social thing for DJs and artists. Going to a record shop as physical space you know... everyone is going to be there and have the same passion as you... That helps a lot in actual networking."*

As mentioned earlier, DJs appreciated the role SNS played in maintaining their connections with peers, fans and record label owners. However, they also experienced the problem of having to maintain profiles in multiple SNS. Facebook and SoundCloud integrated very well according to them, but in general they perceived a lack of integration between most.

G: *"You cannot be everywhere, because the effect is opposite. You have to be in one or two (then) people will be interested. If you have 5 or 6 social networks, it's going to be messy. You cannot arrive to the people. If it's too big, things get too big and then you lose focus."*

Additionally, they also highlighted the need to narrowcast instead of simply broadcasting; this was perceived as a problem in the current SNS.

FL: *"You can record something you put it on a CD and give it in the hands of friends… Give it in their hands.. Otherwise you can record something and upload it on SoundCloud and you have the whole world accessing your files. Its very much; you can reach more people. The problem is that everyone can do it, so I am uploading and another 3 million DJs are uploading their work."*

Almost all the DJs highlighted the need for some of the local DJs to come together as a small community in order to maintain a certain quality of music in the clubs in the city. Some of them even complained about how some in the city were playing 'commercial stuff', a derogatory term in DJ-speak, in order to cater to 'younger audiences who didn't care about music'. On the other hand, they would also engage in activities showing mutual respect, for instance by visiting each other's gigs or commenting on each others' posts on Facebook, following each other and exchanging feedback on SoundCloud.

FL: *"Yes we have a small community. Most of the DJs already know SoundCloud, they have an account, they have (their) work (on it) and we always share... I share, they share… I follow them, they follow me and we have a small community."*

5.4 Loss of Mobility

DJs considered being invited to far off venues as a sign of popularity and success or as an opportunity to gain fame. However, they ascribed the current economic crisis as an impediment to mobility, their ability to travel around or be invited to big gigs, owing to the expenses involved.

H: *"It's expensive for clubs in other cities to pay for my flights. So many times (it's) happened that they couldn't hire me because of that… because of the travel expenses. It's really hard, I need to go more and more to other places… but with that money they can hire bigger names."*

In addition the local clubs also stopped inviting international DJs because of economic reasons and this, according to the DJs, adversely affected the local electronic music 'scene' since the club patrons weren't 'exposed to good stuff'.

F: *"The clubs must invite international DJs… they did that a few years ago."*

Some of the DJs tried to bridge these gaps through the use of SNS by sharing music from well-established international DJs on their pages or personal walls, attracting comments by fellow DJs and fans alike. DJs used SNS to follow their favorite artists and accessed their social media on sites like YouTube and SoundCloud, occasionally catching live streaming feeds of performances.

6 Observations

We present a brief account below of our observations about the DJs' social behaviors online and offline in order to support our findings from the interviews. DJs used SNS extensively to promote events from the clubs they held residencies in. As public faces of the clubs, their social networking blurred the lines between friends and fans, reflecting their unique role as resources for maintaining place-based social capital and as practitioners of 'micro-celebrity' [6,7].

As mentioned earlier the DJs used SNS to organize and coordinate events. Specifically, DJs engaged in a pre-event and post-event dialogue with club patrons using the Events feature on Facebook. It's interesting to note here that most of the clubs' Facebook pages were managed by the resident DJs themselves and hence they would be the ones responsible for creating the event page on Facebook for weekend parties and invite all the 'fans'. They would upload posters and share music videos from content communities such as YouTube or SoundCloud to offer a glimpse of the kind of music to be played.

Although a full account of the participant-observations is outside the scope of this paper, we present below a brief summary of our observations of DJs and reflexive notes from the primary author's first hand experiences playing as a DJ several times. We limit our description to the part that relates to SNS directly or indirectly.

6.1 Friends as Audience

Audience members who were friends of (or simply familiar with) the DJ exhibited a unique etiquette of patronage. They would display their relationship to the DJ by explicitly shaking hands, talking to them briefly or spending time standing or dancing adjacently. These individuals often acknowledged familiarity with, or appreciation of, specific tracks by various explicit gestures (nodding, raising hands, showing thumbs-up). They also frequently took pictures of themselves (sometimes with the DJ), uploaded this content to SNS and tagged themselves and the venue and/or DJ.

6.2 Audience as Friends

Social barriers between a performer and audience members were significantly lowered when the DJ was visible and playing on the same level as the dance floor (instead of on a raised platform). Some of the DJs appreciated this openness since it allowed them to interact with club-goers who approached to ask about a track ID or show their appreciation. Such individuals frequently added the DJ as a Facebook friend in the venue (using smartphones), or simply requested the DJ's account ID. In general, these were readily provided and the process was viewed as socially appropriate (unlike requesting a phone number).

7 Discussion

This paper discusses how DJs use social networks in their work. A substantial body of mixed method fieldwork was carried out. The synthesized results expose the quality and detail of a range of practices from DJs promoting themselves (locally and further afield), to connecting more closely with fans, to gauging their abilities and learning and discovering content. These raise many implications for designers of music related applications, services and social systems. We discuss these implications and outline some of the design directions that emerged from our data. These directions are also aimed at demonstrating the true value of adopting an ethnographic approach for studying such an area of HCI research.

DJs are prolifically social artists adopting a wide range of different social roles. Primarily, they view themselves not just as performers but also educators, finding and defining musical trends. This role as the disseminator of new music has inherent tensions as their ability to set trends partly relies on their ability to spot and gain access to them ahead of the crowd – although their main business is sharing music, to be successful, their sources need to retain an element of mystery. Furthermore, they receive a wide range of public feedback on their work and need to understand and manage it carefully – fans should be supportively thanked and encouraged while peers treated with attention and respect. These tensions suggest a range of novel opportunities for design. For instance, social tools that enable improved segregation of friends or acquaintances into qualitatively different groups (such as peers and fans) would enable them to more effectively respond to feedback – allowing them to narrow-cast rather than broadcast their activities and avoid problems of "context-collapse" [26]. Social music sites (e.g. last.fm, grooveshark.com) and personal music devices (e.g. iPod) could adopt features with a stronger emphasis on the role of curators and integrate recommendations, via charts and playlists, by local DJs.

DJs use SNS to understand their audience better (e.g. using analytics), engage in pre and post performance dialogue with them, and interact with them in nightclubs often augmenting their social capital by adding them as friends. This behavior can be supported more deeply by designing systems that allow DJs to acquire information about their audiences' preferences in music. Considering the fact that DJs 'play to the crowd' and invariably improvise their playlists [10] depending on the audience, it would be of substantial value to build systems that can offer music related meta data from the audiences' SNS profiles (e.g. artists they 'like') on an onsite 'real-time' basis to the DJ.

Beer [16], in his interview with Jarvis Cocker, observed how fans on MySpace interacted more with one another rather than with the musician, and how the presence of musician worked as a catalyst for fan-to-fan interactions. Future research can focus on understanding these phenomena better and explore the role of DJs as a catalyst for enhancing face-to-face interactions among the club patrons. Services such as Foursquare (https://foursquare.com/) that focus on such place-based social interactions can look at ways to integrate the role of place owners or inhabitants in enhancing face-to-face interaction. We observed how the audience integrated SNS with the physical venues the DJs performed in, such as photo-tagging and friend-adding behaviors in clubs. One possible design direction could be to design applications that integrate the music being played at a particular moment with the photographs, allowing the audience to capture memories in more expressive ways and enabling the DJ to have real time feedback or evidence of the impact of his selection on the mood of the audience.

We believe these findings will be useful for researchers and designers working in the area of DJ-audience interaction or in the larger topics of crowd computing and designing for spectators. We also believe that insights such as those presented in this paper can serve as the foundation for designers and developers to create novel integrated SNS that support real performer and DJ practices and needs. Extending beyond DJs, these findings can also inspire design of social systems around common interests that acknowledge the role of human beings embedded in places as resources for enabling social interaction, face to face and beyond.

A limitation of the current study is that all information about SNS is based on data reported in interviews or recorded in participant observation. A valuable extension to this work would be to build tools that capture actual SNS' network data from performers over a protracted period. This would help shed more interpretative light on the specifics of DJs' online behavior. Particular topics for attention should include how DJs use features such as Facebook Pages (and its Listen feature) or the granular commenting system on SoundCloud. As with any small study, the representativeness of the sample of DJs (a busy and relatively hard-to-access user group) observed could be called into question. Data from a larger and more diverse group, for example featuring several cities and DJs at different income brackets, would clearly add meaning and validity to the results. These issues aside, we believe this paper has exposed important aspects of how performers use SNS in their work activities. We hope this impels other authors to explore this interesting and under-researched area in the future.

Acknowledgements. This work was supported by Madeira Life, a project co-funded by the Regional Government of Madeira (MADFDR-01-0190-FEDER_001) and ZON Multimedia. We would like to thank the DJs for their participation and support.

References

1. Choe, S.H., Russell, M.: Bringing K-Pop to the West, New York Times (March 5, 2012)
2. Baym, N.K.: Fans or friends?: Seeing social media audiences as musicians do., Participations. Journal of Audience & Reception 9(2) (2012)
3. Boyd, D.M., Ellison, N.B.: Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication 13(1), article 11 (2007)
4. Baym, N.K.: The new shape of online community: The example of Swedish independent music fandom. First Monday 12(8) (2007)
5. BBC News: Gangnam Style becomes YouTube's most-viewed video, `http://www.bbc.co.uk/news/technology-20483087` (retrieved on January 22, 2012)
6. Senft, T.: Camgirls: Celebrity and Community in the Age of Social Networks. Peter Lang, New York (2008)
7. Marwick, A., Boyd, D.: To See and Be Seen: Celebrity Practice on Twitter. Convergence 1(7), 125–137 (2011), doi:10.1177/1354856510394539
8. Mashable.: 5 Superb Social Media Tools for Musicians, `http://mashable.com/2010/01/04/social-media-tools-bands` (retrieved on January 22, 2012)
9. Gates, C., Subramanian, S., Gutwin, C.: DJs' perspectives on interaction and awareness in nightclubs. In: Proc. DIS 2006, pp. 70–79. ACM Press (2006)
10. Ahmed, A.Y., Benford, S., Crabtree, A.: Digging in the Crates: An Ethnographic Study of DJs' Work. In: Proc. of CHI 2012. ACM Press (2012)
11. O'Hara, K., Brown, B.: Consuming Music Together. Springer (2005)
12. Denora, T.: Music in Everyday Life, 181 p. Cambridge University Press, Cambridge (2000)
13. Brown, B., Sellen, A.J., Geelhoed, E.: Music sharing as a computer supported collaborative application. In: Proc. ECSCW (2001)

14. Tanaka, A.: Interaction, Agency, Experience, and the Future of Music. In: Brown, B., O'Hara, K. (eds.) Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies, vol. 35. Springer, Dordrecht (2006)
15. Goffman, E., Joel, B.: Interaction ritual: Essays in face-to-face behavior. Aldine De Gruyter (2005)
16. Beer, D.: Making friends with Jarvis Cocker: Music culture in the context of Web 2.0. Cultural Sociology 2(2), 222–241 (2008)
17. Sargent, C.L.: Local Musicians Building Global Audiences. Information, Communication & Society 12(4), 469–487 (2009)
18. Shamma, D., Churchill, E., Bobb, N., Fukuda, M.: Spinning Online: A Case Study of Internet Broadcasting by DJs. In: Proc. C&T 2009, pp. 175–184 (2009)
19. Benford, S., Tolmie, P., Ahmed, A.Y., Crabtree, A., Rodden, T.: Supporting Traditional Music-Making: Designing for Situated Discretion. In: Proc. CSCW 2012 (2012)
20. TechCrunch.: New "Listen" Button On Facebook Musician Pages Instantly Plays Their Songs In Your Favorite Streaming App, `http://techcrunch.com/2012/04/17/facebook-listen-button` (retrieved on January 22, 2012)
21. Gilbert, J.: Ping Dead: Apple's iTunes Social Network Will Shut Down on September 30. Huffington Post (September 13, 2012)
22. Rode, J.A.: Reflexivity in Digital Anthropology. In: Proc. CHI 2011, pp. 123–132 (2011)
23. Johnson, R., Rogers, Y., Van Der, L.J., Bianchi-berthouze, N.: Being in the Thick of In-the-wild Studies. In: Proc. CHI 2012. ACM Press (2012)
24. Hook, J., Green, D., McCarthy, J., Taylor, S., Wright, P., Olivier, P.: A VJ Centered Explo-ration of Expressive Interaction. In: Proc. CHI 2011. ACM Press (2011)
25. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. Qualitative Research in Psychology 3(2), 77–101 (2006)
26. Marwick, A.E.: I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society 13(1), 114–133 (2011)

Tension Space Analysis: Exploring Community Requirements for Networked Urban Screens

Steve North[1], Holger Schnädelbach[1], Ava Fatah gen Schieck[2], Wallis Motta[2], Lei Ye[1], Moritz Behrens[2], and Efstathia Kostopoulou[2]

[1] Mixed Reality Laboratory (MRL), Department of Computer Science, University of Nottingham, Nottingham, UK
{Steve.North,Holger.Schnadelbach,L.Ye}@nottingham.ac.uk
[2] The Bartlett, University College London, London, UK
{ava.fatah,w.motta,moritz.behrens,ucftefk}@ucl.ac.uk

Abstract. This paper draws on the design process, implementation and early evaluation results of an urban screens network to highlight the tensions that emerge at the boundary between the technical and social aspects of design. While public interactive screens in urban spaces are widely researched, the newly emerging networks of such screens present fresh challenges. Researchers wishing to be led by a diverse user community may find that the priorities of some users, directly oppose the wishes of others. Previous literature suggests such tensions can be handled by 'goal balancing', where all requirements are reduced down to one set of essential, implementable attributes. Contrasting this, this paper's contribution is 'Tension Space Analysis', which broadens and extends existing work on Design Tensions. It includes new domains, new representational methods and offers a view on how to best reflect conflicting community requirements in some aspects or features of the design.

Keywords: 'tension space analysis', 'human factors', 'design tensions', 'design space', 'urban screens', 'networked urban screens'.

1 Introduction

The contribution to knowledge presented by the authors is a development of existing work on Design Tensions, offering new ways to view and apply the essential knowledge obtained from the user. This development is described in the context of ongoing work on an exemplar Networked Urban Screens project: Screens in the Wild [1]. This paper will discuss how initial work on Screens in the Wild helped to identify specific user-requirement tensions.

In the early stages of the Screens in the Wild project, the researchers had observed the emergence of such tensions. This started a process of reviewing existing thought on tensions and best practice to resolve them. The process of recognising that tensions may be associated with Networked Urban Screens projects and the review of existing literature is described in early sections of this paper. This is then followed by a key section introducing the proposed Tension Space Analysis. In a later section, the reader is introduced to Screens in the Wild in more depth. Finally, an initial attempt is made to apply Tension Space Analysis to the Screens in the Wild project.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 81–98, 2013.

2 Networked Urban Screens: A Special Case When Capturing Community Requirements

During the initial phases of the Screens in the Wild project (which is more fully discussed later in this paper), the research team became increasingly aware of a unique challenge. By connecting four urban screens into a network and embedding this system into several host communities, the importance of capturing, managing and reflecting community requirements (for that screens network) became obvious. In this section, the authors consider the conceptual progression from Urban Screens to Networked Urban Screens. In addition, a brief introduction is provided to the ‘in the wild’ approach to conducting research, as used for the Screens in the Wild project. This way of working is particularly relevant to the unique challenge of capturing community requirements for Networked Urban Screens. It is likely that an embedded research team will be exposed to more honest and in-depth feedback from the host communities.

2.1 Urban Screens

This section will now consider current research into (non-networked) urban screens. Large urban screens are increasingly becoming a part of the architectural landscape of our cities [2][3]. They are used for advertising, art, local information and global newsfeeds, as well as providing public viewing access to cultural and sporting events [4][5]. In this context it is often proposed that urban screens will somehow help regenerate communities, giving free access to culture and providing a focus for local social cohesion and global networking [6]. However, this technology is not without its problems. A recent report by the Commission for Architecture and the Built Environment (CABE) and English Heritage highlighted concerns for the impact on heritage and urban character, calling for installations to be sensitively designed [5][7]. In other research, issues were found with screens becoming a focus for groups of teenagers and perceived anti social behaviour [8]. Finally, their installation can lead to divided opinions amongst local communities. Previous studies have explored people’s social behaviour and relationships to urban screens. For example, Kate Taylor’s qualitative studies highlighted behaviours around the BBC screen in Manchester [9]. Many of the large screens in urban environments are stand-alone and offer a non-existent or very low level of interactivity for the user. Many are ‘pushing’ information, news or advertising at a passing public who are only invited to respond in a passive, receptive manner. Where networking is considered relevant for installations of this type, it might be used to enable a broadcast media model. For example, in the field of 'digital signage', where the advertisements on multiple large screens may be scheduled and updated remotely [10].

2.2 Networked Urban Screens

Beyond instances of screen networking for reasons of technical practicality, minimal consideration has so far been given to the creation of screen networks for the purpose of sharing content experiences across communities. Some examples of content-sharing screen networks are briefly described below. Building on the standalone Urban Screens previously described, possibilities for connecting screens are becoming a topic of interest [11]. Such networks offer new possibilities for social interactions and emergent encounters to arise [4][12].

Some projects have suggested using screens to augment the connectivity of remote communities using videoconferencing technologies (eg 'Hole in the Earth' between Rotterdam and Indonesia and the 'Telectroscope' between London and New York). More recently, projects have been exploring connected cities using large urban screens (e.g. Connected Cities with 6 European cities and large screens and the transnational public sphere [13] connecting Melbourne and South Korea). This sense in which cities may be 'connected', is reflected in the multidisciplinary nature of the Screens in the Wild project described in this paper. Networked Urban Screens provide an interesting area not just for HCI researchers, but also for architects, ethnologists and anthropologists, wanting to explore the virtual space created at the liminal interface between the screens [1]. In addition, such multi-disciplinary groups will find interest in the ways that people move through and use the physical spaces around Urban Screens [5][14]. When considering the virtual space that may be created between Networked Urban Screens, it is possible to find useful literature in fields such as virtual environments and mixed reality. For example, researchers have already given consideration to the idea that mixed reality and virtual environments can be used to digitally extend the physical space that people interact within, leading to hybrid-spatial configurations. Consisting of multiple physical and virtual places, these configurations provide a new type of interactional framework. The impact of these on the inhabitant community may require study in the medium to long term [15], [16].

It should be noted that Networked Urban Screens may have multiple modes within which content experiences are shared by the user. Two of the fundamental modes are: synchronous (the content is shared in real-time, similar to an online multi-player action game) or asynchronous (the content is shared 'as and when', similar to social networks. For example: Facebook). Both approaches have advantages and disadvantages. For example, screen content running synchronously is highly dependent on users being available, simultaneously at more than one screen. Conversely, screen content running asynchronously may leave users at each screen unaware that they are part of a screen network and that other users may be interacting concurrently. For this reason, it is important that screen networks (or at least those where the focus is on exploring emergent social behaviour) have methods of continuous real-time interaction, whether the screens are displaying synchronous or asynchronous content.

In the Screens in the Wild project, this is solved through the use of a video link panel, which shows camera feeds from all of the screens (see later section).

2.3 Research in the Wild

The Screens in the Wild project is usefully considered against the background of a specific methodology which underlies this project: research 'in the wild'. This might best be summarized as the development of systems in real world settings. Research 'in the wild' is often described as a process of moving away from the controlled comforts of the research lab and the safety of prototype demonstrations that need only target research colleagues [17][18]. Instead, the research is placed in-situ, throughout all iterations of requirements capture, design, programming and observation. On this, Chamberlain et al. ask: "is it the case that lab-based studies, taking people out of their natural environment and designing in the lab without long term user engagement are no longer appropriate to properly understand the impacts of technology in the real world?" [17].

The relevance of research 'in the wild' in the context of this paper is that: embedding the research team and their developed artefacts into the target community, may lead to a more thorough understanding of how that community functions and how such functionality impacts community expectations in terms of the delivered artefacts. These expectations might relate to the artefact's form, function or meaning. With a researcher's increased awareness of the underlying dynamics in the community, greater insight into the web of conflicting user requirements may emerge. Therefore, research 'in the wild', by nature of its increased focus on the user community and their requirements, is in need of new methods to aid researchers. Tools are required to ensure that maximum benefits are gained from all of the new observational data collected.

2.4 Tensions

As the Screens in the Wild research team considered the implications of working 'in the wild' in the relatively unexplored field of Networked Urban Screens, a common theme emerged: 'tensions'. The communities had diverse and often diametrically opposing views concerning the screens network and its possible uses. The team's experience of capturing community requirement tensions will be discussed in more detail later in this paper. Having recognised the presence and importance of these 'tensions', it was decided necessary to further investigate existing ideas on managing requirement tensions in the design process.

3 Existing work on Design Space and Design Tensions

When conceptualising and designing a digital artefact or experience, the principles of Human-Centered Design suggest that researchers or practitioners should incorporate the requirements of the end-user community [19][20]. However, the process of requirements capture frequently reveals a complex network of competing values, wants, needs and limitations. The tensions that exist between these requirements may lead to 'design paralysis' [21, p.57]. In this situation, it is not surprising that a common response is to prioritize one or other extreme of a particular tension, rejecting the opposite extreme as less important. How such decisions are made, may often lack clarity or rigor. In User Centred Design (a forerunner of work on Design Spaces and Design Tensions), solutions to this problem (moving from design complexity to specification) are usually sought through scientific methods to aid prioritisation [22]. For example weighted values might be applied to requirement attributes, such as: technical feasibility, available resources, time constraints or popularity of a specific requirement within the studied user community.

However, in his discussion on the nature of design practice, Stolterman suggests that scientific methods are not the best tools for handling the complexity of design spaces. He says: "humans try to reduce complexity, to establish control, by making things simpler" [21, p.57]. The complexity surrounding the creation of an artefact or experience is often described as 'the Design Space' [23][24]. Earlier work considers the same phenomenon, under the name Design Rationale. However, the focus of much work on Design Spaces, assumes a need to simplify or filter the tensions, in order to prioritize user requirements. Tatar proposes an alternative analytical framework: 'Design Tensions':

> "*Design tensions conceptualize design not as problem solving, but as goal balancing. They draw explicit attention to conflicts in system design that cannot be solved, but only handled via compromise*" [25, p. 3].

Design tensions "differ from design spaces in that they do not set boundaries or simplify the problem but rather provide a framework for creating a space of relevance." [25, p. 413]. In Design Tensions, Tatar utilizes the older Value Sensitive Design methodology (and how it integrates value considerations) [26].

4 The Need to Broaden and Extend Existing Work on Design Tensions

There are clearly some limitations of the existing work on design spaces and tensions and these may now be summarized. With regards to Design Spaces, one issue is that they "are not easily extensible to describe a large number of dimensions or more complex interrelationships" [25, pp. 415-416]. Furthermore, it is assumed with Design Spaces that all requirements will eventually be reduced down to a set of essential artefact attributes. Therefore, some requirements will be labelled as 'less important' to the artefact. On the other hand, it was problematic for us that Design Tensions are often expressed in terms of tension between the user-community and the technology, rather than the requirement tensions within the user-community itself (which we would argue is often the key focus in Networked Urban Screens projects, such as Screens in the Wild). Also, while Design Tensions help us to understand that the simplification of conflicting user requirements may be undesirable, this work still suggests that not all user viewpoints can be reflected in the artefact or experience [25].

5 Tension Space Analysis

In response to the limitations described in the previous section, Tension Space Analysis is offered as a methodology to aid understanding of any tensions that may exist, surrounding the requirements for an artefact.

Extending the existing research on Design Tensions and Design Spaces, Tension Space Analysis focuses on tensions within the user-community itself, not between users and technology. Tension Space Analysis helps to ensure that tensions are represented more fully in the artefact, system or design, rather than being rationalized at the point of design implementation. A visual method of representation for design tensions is provided.

Table 1. Tension Space Analysis step-by-step

Through a variety of ethnographic and requirements capture methodologies, identify reoccurring tensions within the project or design.
Formalize these tensions in writing.
Consider whether these tensions apply generally, or may be categorized as only relating to one or more aspects of the project or design (for example system-only, or screen content experience only).
Visualize the tensions diagrammatically, using a representation that allows for rapid 'by eye' comparison. Note: at this stage you have a diagram showing the tensions and their dimensional extremes, but no aspects of the system have actually been mapped across this representation. This happens in the next stage.
With a radar chart representing the tensions as axial lines, plot on the current project or design. Use the tension extremes as questions: "is content locally or globally sourced? Is there any advertising? Etc". Plot on the radar chart accordingly.
Use the diagrams frequently, to compare system/content versions, when writing design briefs, when reviewing completed projects and at all stages in a project lifecycle. Use this as a guide to check that your current thinking reflects all of the requirement tensions present at the inception of the artefact. If the current artefact cannot itself practically address all requirements, consider how unrepresented tensions might be addressed. Can they be represented (or otherwise resolved) in later versions, updates, sub-elements or complementary artefacts?

The decided visual approach was to look for a relatively simple diagrammatic method to reflect the individual characteristics of a designed artefact (in this case a Networked Urban Screens system and its screen experiences). It was hypothesized that, using a visual method to compare two or more possible design candidates (or built systems, content experiences running on a system etc), would allow distinctive patterns to emerge. As the human eye is adept at perceiving differences between patterns, the intention was to provide a 'quick and dirty' tool to aid iterative design review and comparison. With the user community's requirement tensions always 'present' at meetings (through the diagrams), it becomes harder not to continually address the underlying tensions and to ensure that they remain reflected in ongoing design decisions. A ‘radar chart’ (see the later section, where an example implementation is presented) was chosen for this purpose, as it allows the representation of multiple requirements, each along an axial line, representing its dimensional extrernes. The Tension Space Analysis process is summarized in Table 1 (above).

6 The Screens in the Wild Project

Tension Space Analysis emerged from a Networked Urban Screens project called Screens in the Wild. This project considers a specific challenge: how best to integrate urban screens, a radical and potentially disruptive new technology, into the urban realm. As the research for the project is conducted ‘in the wild’, an essential element has been to establish a network of four urban screens, as a presence in the communities being studied. In fact, the four screens are based in two UK cities, two screens in each city. This network of screens spanning sites in two cities provides the base-platform for the core research of the project. Partners in the project include two universities and local authority / urban regeneration representatives. The entire setup provided a unique opportunity for the research to examine sub-regional interactions (between the communities within the town centres), as well as UK-wide. The Olympics provided an ideal cultural backdrop to the research.

The original research proposal envisaged that the screens would be placed in public areas, such as (the inside of) retail shop windows or accessible public service buildings. Following an active phase of researcher ‘outreach’ (and in some cases with the assistance of local authority partners), suitable venues were identified for all four screens. All of the screens are positioned inside the front windows of public space venues, visible from the street and/or public areas. The venues are: a public library, a community centre, an art space and a cinema. One aspect of this project’s academic study is to consider the screens network from an urban design perspective. Specifically, to develop a theoretical understanding of how the urban experience, mediated through networked urban screens, can be augmented to support communities and culture and the unique design issues related to the implementation of large public screens in urban space [27].

To further this agenda and prior to screen installations, members of the research team studied the built environment in the proposed venue locations. They considered how the urban spaces surrounding the planned screen locations were currently structured and used by different groups of the population. In addition, ethnographic methods were used to capture emergent social interactions in these spaces, before the screens were installed. There then followed a phased installation of all four screens, over a period of several months. During this time, initial screen content prototypes were being built and tested. From the urban design perspective, researchers were interested in the impact of the networked screens on people’s relationships with physical, social and interaction space. From the ethnographic perspective, researchers looking at all four networked screen locations (differing in their urban settings and the types of populations they support), were interested in identifying both outcomes that were purely site-specific and those that could be generalized across different sites. As prototype screen content was gradually deployed across the network, a process of iterative looping fed back captured user responses and requirements into refined prototypes. The knowledge captured from these iterations, also helped the researchers to generate ideas for potential new screen experiences that might be initiated at a grass roots level by the communities involved. A further strand explored generating ideas for screen content experiences generated by local artists, or artists-in-residence with experience in working with local community groups.

6.1 The Screen Experiences

The Screens in the Wild project developed a number of screen experience prototypes. These were implemented on a base system iteratively developed in the first phase of the project. The key aspects of this system (a full description goes beyond the scope of this paper) are as follows. Each 46” screen is mounted vertically (portrait mode) inside a partner venue, i.e. behind glass. A camera and speaker (both available 'off-the-shelf') operate through the glass, making the installation interactive. Each screen node also contains a networked Windows PC, which is remotely administered. The software system uses the UNION Client/Server infrastructure [28] to provide real-time multi-user functionality for web applications across the screen network. Client applications were implemented in HTML5/Javascript to run in a local copy of Firefox full-screen, with scheduler software switching between the experiences described below. To protect the system from undesired remote and local attention R-Kiosk is used in Firefox [29] and Secure Lockdown is used for the desktop [30]. All four nodes are administered remotely using TeamViewer [31].

Video Link

As general support for interaction and potentially also as a standalone experience, synchronous four-way video communication (no audio) was implemented using a combination of YawCam [32] and iSpy [33]. Both the video panels and the USB camera generating the video feed at each of the nodes have been placed at a low height to encourage interaction from and with children. Currently, video is being displayed as set of four video panes, towards the bottom of the screen (see Fig.1, below). The pane to the left always shows the local feed. It has been mirrored to increase its value as an attractor [34]. The other three panels show video from the remaining three connected screen locations. The Video link encourages synchronous multi-user interaction across the four screen nodes by providing a view of who is interacting and a way to acknowledge or even interact with the other party by waving and through gestures.

SoundShape

SoundShape provides for collaborative music making between users in different physical locations. It is inspired by the ‘Tenori-on’ electronic musical instrument, which featured a grid of LED pad switches that could be activated in a number of different ways to create sounds. The visual illumination of active pads produced patterns and displays to complement the music. In SoundShape, this idea is taken a stage further by linking networked urban screens to create a multi-user, networked musical instrument (see Fig.1, left).

Fig. 1. SoundShape shown in context at one of the venues (left) and a screen shot of ScreenGram (right). The video link is displayed at the bottom of each experience.

Each networked screen has its own SoundShape client. A central server remembers the state of each pad. If a user touches a pad (via touchfoil enabling through-glass touch detection), the server sends this new information to all of the networked screens. All participants see and hear the changes to the musical composition. SoundShape has a 5x5 grid. Each row represents a note. Each column represents 5 chronological steps in a continually looping musical sequence (1,2,3,4,5,1 etc). As notes are added, or removed (by touching the pads to turn them on or off), the musical sequence changes and develops. Active (currently sounding) pads have an animated illumination that flashes in time with the musical sequence. At the one screen where the chosen touchfoil technology does not work with the installed glazing, users can scan an onscreen QR code into their smartphone enabling their phone to work as an interaction device with the screen. SoundShape encourages synchronous multi-user interaction between the four screen nodes. People can share the creation of a single looping music sequence. They can also compete over who can change buttons faster and over which pattern should be on screen.

ScreenGram

ScreenGram (see above Fig.1, right) searches Twitter for tweets that have specific hashtags, which also have associated Instagram photos. It then retrieves the photos from Instagram and displays them as a constantly updating and revolving slide show. In its simplest form, ScreenGram can display the results for just one hashtag. Users can stand in front of the screen (or anywhere else, but in front of the screen they will get instant feedback), take a photo (or retrieve an image from somewhere else) and send it over Instagram, to Twitter, with the a hashtag of their choice. It usually only takes from 30 seconds to a minute to retrieve a new photo that has been tagged and for it to appear on the networked screen. The buttons on the right-hand side of the screen allow the user to display the results for other (currently relevant) hashtags. For example, the user can also choose to show the latest photos for: #summer, #olympics etc. It is very easy for the researchers (or in theory the community, venue staff, etc.) to change the hashtags that appear on these buttons. The buttons are ‘radio buttons’. So, clicking one will disable any already selected and each button toggles on (red) and off (clear). ScreenGram encourages asynchronous multi-user interaction between the networked screens and people elsewhere interacting through their phones. They can actively participate by posting images to the screen, and for this interaction they do not have to be at the screen location. At any of the nodes, people can also filter the onscreen content and watch the slideshow that is determined by which hashtag is currently selected.

7 Capturing Community Requirement Tensions in the Screens in the Wild Project

Before, during and after the initial screen content prototypes were in development, researchers from the Screens in the Wild project observed the communities located near to the venues. It is in the nature of research ‘in the wild’ that the distinctions between user, researcher and any other person in contact with an artefact become blurred. Certainly the role of the researcher as a dispassionate, objective ‘gatekeeper’ is highly questionable, when he or she is embedded into the community under study. In a sense, the researcher becomes a part of an ‘ecology’, shared with the local community and anyone else that interacts with or is in any way impacted by the presence of the introduced artefact. Some of the methods used included: observations, conversations, interviews and workshops.

Researchers conducted semi-structured interviews with managers of the organizations that allowed us to install a screen node in their premises. The purpose of these interviews was to better understand their perception or perspective of the project. Before the development and implementation, the researchers organised four workshops with members of the public at two locations in London. The intention of these workshops was varied. In some cases, the goal was to analyze the usability, in others to enable some production of local content for the screen, but in general the research team was seeking to receive feedback or suggestions of different kinds on the work completed so far.

In order to kickstart the development of screen content prototypes, the academic partners held a bodystorming session in a research lab. Similar to brainstorming, bodystorming encourages participants to move around, use their bodies and to use physical or paper mockups as stand-ins for the proposed artefact(s). In this case, functioning prototypes of the hardware and software system were set up in the research lab. Participants were requested to test what was built so far and suggest new ideas for development. Later on in the project, the stakeholders of the first venue to deploy at were invited to a similar session to give them a sense of what would be installed later in the project.

Researchers also participated in using the screen content prototypes, alongside communities, uploading their own images onto the screen nodes, sharing their perspectives and personal views.

Later, the team reflected on its own participation, producing field notes and summaries of their screen interaction sessions. This enabled the researchers to understand the context of each screen experience in a more profound way. The team's time spent as 'co-participants' better prepared them to respond to any experience-related concerns expressed by the user-community participants. Researchers conducted sixty-four structured interviews with people belonging to different social groups and communities in neighbourhoods where the screens are located. The people interviewed had a diverse level of engagement with the social life of the neighbourhoods.

7.1 Identified Tensions in Community Requirements

As the Screens in the Wild project still has several months left to run, formal evaluation has not been completed. That said, the conducted observations, interviews and workshops have already thrown up a fascinating raw dataset. This section describes the process of distilling the unstructured community feedback, down to a series of observed tensions.

Table 2. Sample opinions from the raw dataset

Venue owners have concerns over control of curation of content and scheduling.
"People are starting to get the hang of the schedule, especially the lights. Some comments on the schedule from the public: 'Can you turn the game on now?', 'Can you show me that picture from this morning?' suggest that people think we're in control of the schedule, and indeed that we control the screen as a whole."
On the Sound Shape experience "The lights game: this is the most popular app, with users of all ages, but especially appealing to pre-teen children. Some adults have asked what the point of the game is, but kids largely make up their own rules and get quite competitive. It's not unusual to see a group of two or three children playing with the squares all at once, with an adult looking bemused over their shoulder."
Participants emphasized the importance of having an open system, where anyone could upload pictures, curate and engage with the screens in personal terms. This involved less censorship from institutions and more user options for choosing/producing what they would like to see.
People revealed that they ignored public screens present at nearby local shopping centres or other public venues. They have learned to dismiss them as part of annoying highly-commercial advertising.

Table 2. (above) presents some sample opinions (both formal and informal) expressed by the user communities and venue owners/managers.

In the team meetings and discussions following the most recent round of observations, it started to become obvious that certain requirements (or ‘what they would like to see the screen network displaying in their community’ to put it more simply) were repeatedly expressed by the local communities. What struck the researchers was how frequently there was complete polarization of opinion on a single issue. For example, person1 says: “the last thing I’d want to see is advertising – keep it non-commercial” vs. person2 says, “it would be great for advertising local businesses”. Having recognized the existence of these ‘tensions’, it was interesting to record and consider them. Table 3. presents the observed dimensions of tension, labeled in natural language, as they had expressed in interviews and conversations.

Table 3. Identified requirement tensions

Moderated community	vs.	Open community
Auto scheduled	vs.	Community/user scheduled
Cities hosting screens too culturally similar, contrast not great enough...screens should be more geographically/culturally separated	vs.	Don’t know where cities are!
More local content	vs.	More generalised content
Local advertising	vs.	Non-commercial
Content not personal enough	vs.	Content too personal
Content with history, places, things and tangible objects important - stimulate memories	vs.	People, relationships important, current moment important - don’t care about things

It is important to stress that these are only some of the possible tensions that might exist within a project. They may be unique to the Screens in the Wild project and no claim is made to have identified a comprehensive set of tensions. There may well be some tensions that are common to all projects/artefacts/experiences, but it is assumed that each situation will generate its own unique ‘tension space’. In addition, tensions may exist across many sub-domains of the project (for example: community vs. technical, community vs. researchers, venue owners vs. venue users etc). At this stage, attention is only being given to tensions that exist within the user community itself (venue users, venue owners, local residents etc), as these are the groups physically impacted by the introduction of a screen network, into their local urban space.

Next, some of these natural language tension descriptions were translated into shorter more descriptive labels (partially for easier use in visual representations):

- **moderated community vs. open community** (unchanged): can content be added to the system (or individual screen experience, if applied to this) by the public, or is there some level of moderation?

- **auto scheduled vs. community/user scheduled** (unchanged): is the programming of the screen content (time of day, choice of experience etc) controlled by the researchers automatically (using a scheduler application) or in the hands of the venues or venue users?
- **locations sufficiently different vs. locations not different enough** (formerly: Cities hosting screens too culturally similar, contrast not great enough...screens should be more geographically/culturally separated vs. Don't know where cities are!): some users thought that the cities hosting the screens were too similar, others (from one city), did not know the location of the second city.
- **locally sourced vs. globally sourced** (formerly: More local content vs. More generalised content): is screen content created in the community, or could it come from anywhere (over the Internet)? Is this content about this community (or this screens network), or is it location-neutral?
- **(local) advertising vs. no advertising** (formerly: Local advertising vs. Non-commercial): should the network (or experience) be commercial or completely non-commercial?
- **personal vs. non-personal** (Content not personal enough vs. Content too personal): some users were very worried about sharing personal information. Where was it going? Who would see it? Others, wanted experiences to be more personal.
- **things & past vs. people & now** (formerly: Content with history, places, things and tangible objects important - stimulate memories vs. People, relationships important, current moment important - don't care about things): there was a noticeable gap between users wanting to see images of past times and events (often coupled with low levels of interactivity from them) and users wanting to see current (often social) content. Sometimes, this division reflected a difference between age groups, but not always.

8 Applying Tensions Space Analysis to the Screens in the Wild Project

Now, it is possible to apply the Tension Space Analysis concepts developed thus far to current elements of the Screens in the Wild project. Referring back to the tensions identified earlier in the paper, some of them are immediately more relevant to the baseline screen network system itself (without screen experiences), whereas other tensions are relevant to the screen content experiences.

The figures below, illustrate the mapping of the identified the tensions, as dimensions on two separate radar charts. Fig. 2. shows the dimensions primarily relevant to the baseline screen network. Fig. 3. presents the dimensions considered more relevant to the screen experiences. These radar axial radar charts may be regarded as 'empty templates', ready for the assessment of the system and screen experiences relating to the studied artefact.

Referring back to Table 1, once the empty axial line radar charts have been created, the next step is to superimpose the current project or design. To achieve this, it may

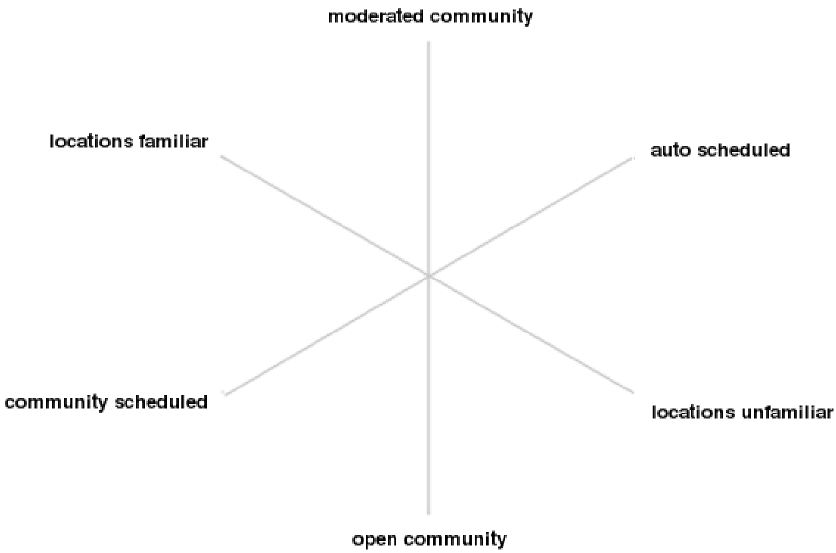


Fig. 2. The system-related tensions shown as dimensions on a radar chart

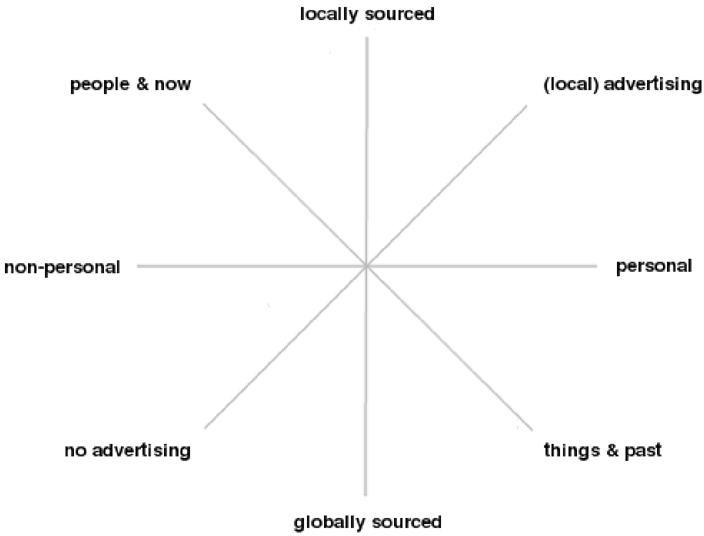


Fig. 3. The experience-related tensions shown as dimensions on a radar chart

be useful to consider the 'tension extremes' on each axial line in terms of questions (for yourself or the project team): "is content locally or globally sourced?", "is there any advertising?" etc. The results of this are then plotted onto the radar chart accordingly. The figures below show the tensions applied to both the baseline system (see Fig. 4.) and an example screen experience – ScreenGram (see Fig. 5.).

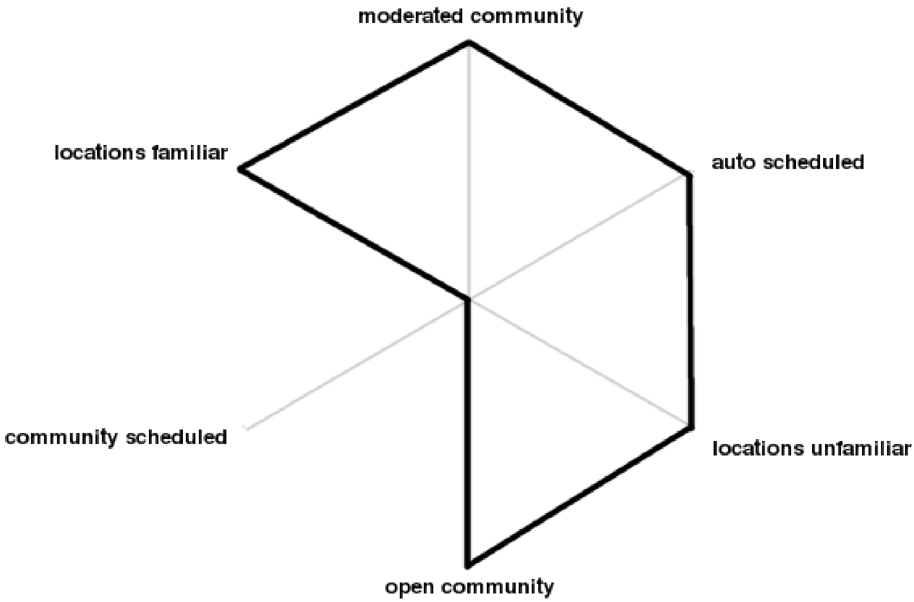


Fig. 4. Tension Space Analysis radar chart for the current screen network system

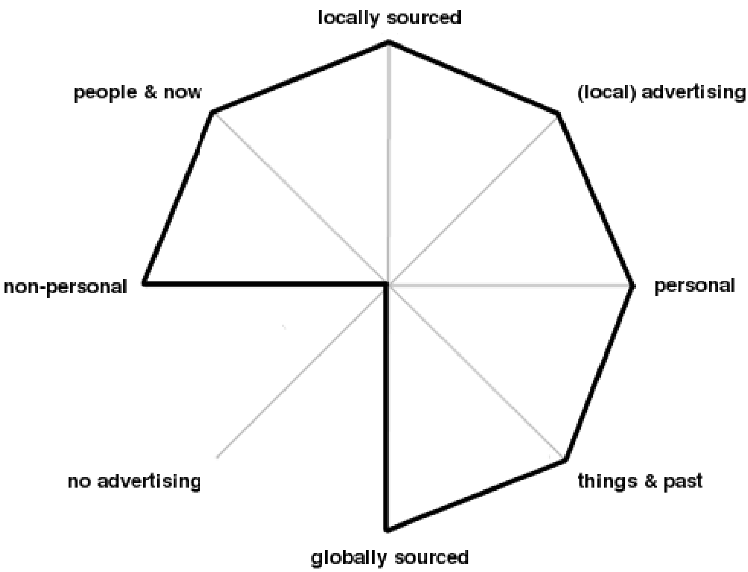


Fig. 5. Tension Space Analysis radar charts for the ScreenGram prototype experience

Looking at the radar chart for the current screen network system (see Fig. 4.), we can see that it is capable of hosting both open and moderated communities (dependent on the experience running), that content is automatically scheduled, that users have reported both that the cities hosting the screens are culturally familiar (often implying that more distinction would be preferable) and also that one of the other host cities is unfamiliar to them (often implying that they regard the differences as sufficient to make the system interesting to them).

Looking at the radar chart for the ScreenGram prototype experience (see Fig. 5.), it can be seen that most of the tensions are represented – a pleasing result!. The highlighting on the advertising tension line shows that users desiring 'no advertising' may be disappointed. This is because ScreenGram's content is open (it represents an example of the 'open community' axial extreme, shown as active for the entire screen network system – see Fig. 4.). Anyone can post an image to the screen network by applying a specific hashtag. Therefore, it is possible advertising might appear and the 'no advertising' end of the axial line is not highlighted.

9 Conclusions

Drawing on a live, 'in the wild' research project (Screens in the Wild), we have identified some of the design tensions that might be encountered when designing any network of urban screens (for both the system itself and its screen content experiences). These tensions are likely to emerge because of the different views that multiple stakeholders (located within very different communities) have about the use of such a screen network.

Tension Space Analysis proposes an approach for capturing and visualizing existing user requirements and ensuring that they are reflected in the design, or operational modes of the final artefact or experience. In Tension Space Analysis we would suggest that it is essential to keep alive an essence of all user requirement tensions within the artefact, even if particular aspects of each dimension are emphasized either in the final static artefact, or through different aspects/modes of the artefact's function. For example, in the field of Networked Urban Screens, each screen experience that is displayed across the screens offers an opportunity to reflect some of the tensions.

Therefore, it is not necessarily the main goal to reflect all of the tensions in the totality of the system. All tensions might be incorporated gradually over its lifetime. In addition, Tension Space Analysis specifically considers tensions within the requirements of user-communities, as opposed to Design Tensions, which primarily focuses on user community versus technology tensions.

The contribution made by Tension Space Analysis is to broaden and extend existing thought on Design Tensions, to include new domains and methods of representation. Specifically, Design Tensions are improved in two specific ways.

Firstly, the focus of analysis is shifted from tensions that exist between users and technology (as in Design Tensions), to the tensions that exist within the user community itself. Secondly, Tension Space Analysis introduces a visual aid to understanding user-requirement tensions. This will then be used at all stages in the project to discourage researchers/designers from prematurely discarding tensions before fully considering how they might best be represented in the artefact itself, either in the short-term or during its entire lifecycle.

With an increasing HCI focus on 'in the wild' research and user community engagement, there is a pressing need for approaches such as Tension Space Analysis. It is hoped that using techniques such as this will help researchers to document and then frequently revisit the full spectrum of community requirements, ensuring maximum representation of the target community within a delivered artefact.

Acknowledgements. We would like to gratefully acknowledge the support from the Engineering and Physical Sciences Research Council (EPSRC) through grant EP/I031413/1 (Exploring the potential of networked urban screens for communities and culture) as well as the continued support from our partner venues.

References

1. Screens In The Wild, `http://www.screensinthewild.org/`
2. Brignull, H., Izadi, S., Fitzpatrick, G., Rogers, Y., Rodden, T.: The introduction of a shared interactive surface into a communal space. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 49–58 (2004)
3. Struppek, M.: The social potential of Urban Screens. Visual Communication 5, 173 (2006)
4. Fatah gen Schieck, A.: Towards an integrated architectural media space. First Monday 4 (2006)
5. Fatah gen Schieck, A., Boddington, G., Fink, P.: Framework for the implementation of urban big screens in the public space (2009)
6. McQuire, D.S.: The media city: Media, architecture and urban space. Sage Publications Ltd. (2008)
7. Large digital screens in public spaces. Joint guidance by English Heritage and CABE (2009)
8. Fatah gen Schieck, A.: A tale of two cities. In: Urban Screens 2008, Melbourne, Australia (2008)
9. Taylor, K.: Programming video art for urban screens in public space. First Monday (2006)
10. Memarovic, N., Langheinrich, M., Alt, F.: Interacting Places — A Framework for Promoting Community Interaction and Place Awareness through Public Displays. In: Adjunct Proceedings of the Tenth Annual IEEE International Conference on Pervasive Computing and Communications (2012)
11. Davies, N., Langheinrich, M., José, R., Schmidt, A.: Open Display Networks: A Communications Medium for the 21st Century. IEEE Computer 45(5), 58–64 (2012); Special Issue on Interactive Digital Signage
12. Memarovic, N., Langheinrich, M., Alt, F.: The Interacting Places Framework – Conceptualizing Public Display Applications that Promote Community Interaction and Place Awareness. In: Proceedings of the International Symposium on Pervasive Displays (PerDis 2012). ACM, New York (2012)
13. Papastergiadis, N.: Spatial Aesthetics: Rethinking the Contemporary. In: Smith, E., Condee (eds.) Antinomies of Art and Culture, pogl., vol. 19

14. Brynskov, M., Dalsgaard, P., Ebsen, T., Fritsch, J., Halskov, K., Nielsen, R.: Staging urban interactions with media façades. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 154–167. Springer, Heidelberg (2009)
15. Schnädelbach, H., Penn, A., Steadman, P., Benford, S., Koleva, B., Rodden, T.: Moving office: inhabiting a dynamic building. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 313–322 (2006)
16. Schnädelbach, H.: Hybrid spatial topologies. The Journal of Space Syntax 3(2), 204–222 (2012)
17. Chamberlain, A., Crabtree, A., Rodden, T., Jones, M., Rogers, Y.: Research in the wild: understanding 'in the wild' approaches to design and development. In: Proceedings of the Designing Interactive Systems Conference (DIS 2012), pp. 795–796. ACM, New York (2012)
18. Rogers, Y.: Interaction design gone wild: striving for wild theory. Interactions 18(4), 58–62 (2011)
19. Chayutsahakij, P.: Human Centered Design Innovation. Institute of Design. IIT Technology (2000)
20. Vanden Abeele, V.A., Van Rompaey, V.: Introducing human-centered research to game design: designing game concepts for and with senior citizens. In: Vanden Abeele, V.A., Van Rompaey, V. (eds.) CHI 2006 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2006), pp. 1469–1474. ACM, New York (2006)
21. Stolterman, E.: The nature of design practice and implications for interaction design research. International Journal of Design 2, 55–65 (2008)
22. Garrett, J.J.: The Scope Plane: Functional Specifications and Content Requirements. The Elements of User Experience: User-Centered Design for the Web. In: Prioritizing Requirements, ch. 4. Peachpit Press (2002)
23. Card, S.K., Mackinlay, J.D., Robertson, G.G.: The design space of input devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, pp. 117–124 (1990)
24. MacLean, A., Young, R., Bellotti, V., Moran, T.: Questions, Options, and Criteria: Elements of Design Space Analysis. Human-Computer Interaction 6(3&4), 201–250 (1991); Special Issue on Design Rationale: Carroll, J.M., Moran, T.P. (eds.)
25. Tatar, D.: The design tensions framework. Human–Computer Interaction 22, 413–451 (2007)
26. Friedman, B.: Value-sensitive design. Interactions 3, 16–23 (1996)
27. Fatah gen Schieck, A., Schnädelbach, H., Penn, A.: Research in the Wild: Exploring the potential of Networked Urban Screens for Communities and Culture. In: Research in the Wild Workshop, DIS 2012, Newcastle, UK (2012)
28. Union Platform, By USER1, `http://www.unionplatform.com`
29. R-kiosk Firefox plugin, `https://addons.mozilla.org/en-us/firefox/addon/r-kiosk/`
30. Secure Lockdown, `http://shop.inteset.com/Products/9-lock-down-windows-7-with-inteset-secure-lockdown.aspx`
31. TeamViewer, `http://www.teamviewer.com`
32. YawCam, `http://www.yawcam.com`
33. iSpy, `http://www.ispyconnect.com`
34. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking glass: a field study on noticing interactivity of a shop window. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 297–306. ACM, New York (2012)

BinCam: Designing for Engagement with Facebook for Behavior Change

Rob Comber[1], Anja Thieme[1], Ashur Rafiev[1], Nick Taylor[1],
Nicole Krämer[2], and Patrick Olivier[1]

[1] Culture Lab, Newcastle University, Newcastle upon Tyne, UK
[2] Social Psychology: Media and Communication, University of Duisburg-Essen, Germany
{rob.comber,anja.thieme,ashur,rafiev,nick.taylor,
patrick.olivier}@ncl.ac.uk, nicole.kraemer@uni-due.de

Abstract. In this paper we continue work to investigate how we can engage young adults in behaviors of recycling and the prevention of food waste through social media and persuasive and ubiquitous computing systems. Our previous work with BinCam, a two-part design combining a system for the collection of waste-related behaviors with a Facebook application, suggested that although this ubiquitous system could raise awareness of recycling behavior, engagement with social media remained low. In this paper we reconsider our design in terms of engagement, examining both the theoretical and practical ways in which engagement can be designed for. This paper presents findings from a new user study exploring the re-design of the social media interface following this analysis. By incorporating elements of gamification, social support and improved data visualization, we contribute insights on the relative potential of these techniques to engage individuals across the lifespan of a system's deployment.

Keywords: Engagement, Facebook, Sustainability, Recycling, Gamification, Social influence, Persuasive technology.

1 Introduction

The disposal of waste is a mundane behavior. Being part of our daily routines, it does not request much of our attention. The impact of waste disposal on the environment however suggests that we should bring more awareness to our waste-related habits [35]. HCI has proposed a number of means through which we can motivate individuals, groups and society to engage in behavior change, including personal informatics [25], persuasion [14,16], gamification [7,23,22], social influence [5,27], and in small number of cases coercion [32,15]. Persuasive technology has become one of the key trends in this regard (see [3,10] for a review of the sustainability literature). Yet, in direct opposition to the lack of motivation we experience on a daily basis, persuasive technology proposes that designers should focus their influential potential on those who are already motivated, facilitating their paths to goal fulfillment [14].

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 99–115, 2013.

The question then arises of how we first entice people to be engaged in a process intended to change their behavior. In this paper we explore our continued research in this area with the BinCam system. The BinCam system is a two-part persuasive technology, with which we explore issues of engagement, persuasion and motivation for recycling and food waste behaviors. The system comprises a household landfill bin, fitted with a mobile phone in the lid, which captures images on closure and shares them on a dedicated Facebook application. Although the system aligns itself with sustainable HCI, our primary goal here is to understand and motivate engagement. In this light, we recognize that monitoring or engaging people in recycling and food waste behavior is not intrinsically motivating for most people. As such, we consider it a significant challenge for HCI, and society more generally, to contemplate the means through which we can engage individuals in such behaviors.

1.1 Context and Motivation of the Research

Engagement in recycling behavior is a significant challenge for individuals in the UK aged 18-35, who are largely unaware of the problems associated with inappropriate disposal of waste [35]. We previously explored this issue with regard to recycling and food waste behaviors from the perspective of habits [4] and as a problem of awareness and reflection [32]. In both cases we employed social media to facilitate engagement in behavior change. Research has shown the positive effects of including others in behavior change efforts [5,27] and related positive impacts of social media [12,15,23]. Social media provides a platform through which we can leverage a number of channels for behavior change, including personal informatics, social informational and normative influence [8], persuasive messaging and hedonic motivation.

Findings from our previous research revealed that, although individuals in the target group have strong positive attitudes towards recycling and sustainability, they often do not act towards these attitudes. While the data capturing part of our system served as a means to draw attention to waste disposal behavior, engagement with a social media application associated with the intervention was less successful. We noted particularly that individuals were not motivated to use the system beyond a short period of time. From our previous work, we are now faced with two questions: How can we further engage participants in discussion around recycling? How can we promote engagement with social media as a means to facilitate this discussion?

Driven by these questions, we explore the re-design of the social media During the re-design of the BinCam system, we have incorporated lessons learned from both the previous study, and through a critical analysis of how HCI understands and designs for engagement. We contribute an analysis of how engagement is understood in HCI; the strategies employed to support and promote engagement; an empirical analysis of engagement techniques in waste disposal behavior; and a critical reflection on engagement with, and use of, social media as a means to promote behavioral change.

1.2 Engagement

There is considerable diversity to how HCI and related disciplines have conceptualized engagement. There is also much confusion about what constitutes engagement, as it is often used interchangeably with notions of participation, immersion and attention. HCI has appropriated the concept of engagement ranging from the broad indicator of the quality of interaction to the cognitive measure of attention [29]. In the development of web applications, engagement is measured on at least three axes. As [24] put it: "*Successful web applications are not just used, they are engaged with; users invest time, attention, and emotion into them.*" Such engagement is most often measured on two planes – first as *behavioral engagement* measured through behavioral data such as mouse clicks, time spent on pages and applications, and secondly as *affective engagement,* as captured by measures of satisfaction and affective response in questionnaires.

Engagement has also been understood in education and learning as the 'physical and psychological effort' devoted to a task [1]. Importantly, it is understood to have quantitative and qualitative features, to occur along a continuum and to impact on the effectiveness with which individuals achieve their goals. While physical effort pertains to behavioral engagement, psychological effort relates to a notion of *cognitive engagement.*

The exertion of psychological effort has been further considered in HCI research in terms of flow or optimal experiences [6]. Flow experiences occur when an individual is fully (emotionally, cognitively and physically) immersed in a task. Such immersion is not felt as requiring significant effort and is intrinsically motivating. Although flow represents optimal experience, it would not be expected to occur across a sustained and mundane activity, such as waste disposal. Flow experiences may also limit self-reflection, suggesting a non-conscious *intrinsic engagement.*

The recent move towards richer accounts of HCI has shifted the focus of engagement to understand it as the meaningful interactions an individual has with an artifact or service, and the quality of the attachment a user has to an object. While less easily quantified, such meaningful interaction has become a critical factor in third wave HCI [2,28] and extends beyond affective engagement. Such work draws on an understanding of engagement as the meaningful and effortful reflection on activity in experience [9]. This *reflective engagement* involves the critical reflection on on-going activity, thinking over and through current actions as they occur.

There is also an increase in the application of social dynamics in the design of persuasive systems [27,32]. Such dynamics can impact on individuals' performance of behavior positively in increasing engagement (e.g., social facilitation), or negatively in decreasing engagement (e.g., social loafing). Thus, although it relies on a variety of interpersonal and personal factors, and incorporates elements of affective, cognitive and behavioral engagement, this can be collectively understood as *social engagement.*

Against the backdrop of this theoretical conception, the following presents a variety of strategies for engagement, commonly applied in persuasive HCI.

1.3 Strategies for Engagement

In attempting to promote engagement with online mental health interventions, Doherty et al. [11] suggest designing systems that are *interactive*, allow for *personal* experiences, provide *support* for the individual and facilitate *social* contact with others. Engagement strategies borrowed from the field of personal informatics [25] are mostly targeted at *rewarding* the user if the desirable behavior has been performed. Very little research in HCI has so far been dedicated to studying the effects of *negative reinforcements* to promote behavioral change [e.g., 15,23,32].

Interactivity. Interactivity relates to the providence of rich and varied experiences through the use of the system, which actively invite user exploration. This can, for instance, be achieved through diverse representations of peoples' behavioral data providing insights about their performances, whether they improved or how their behavior compares to others [11]. At its most basic, interactivity affords behavioral engagement; where the ability to perform actions and receive responses invites users to further engage with the system.

Space for Personal Appropriation. Personal experiences are often achieved through tailored designs based on the individual preferences of the user, which facilitates a sense of control as well as ownership [11]. Personal appropriation lends itself to affective engagement, where individuals perceive similarity to or ownership of an interactive system [33]. With personal appropriation, group identification could increase affective engagement through affording a sense of belonging, but also behavioral engagement through normative influences. That is, the presence of normative in-group behaviors may persuade individuals to engage in group-similar behaviors.

Behavioral Support and Reminders. In the context of health supporting interventions, the 'supportive' strategy is intended to improve peoples' adherence to a treatment program [11]. It is assumed that implemented (personal) support, such as a recycling coach or a recycling guide in our context, helps increase an individual's motivation to display or continue a certain behavior. This engagement strategy, however, is targeted at motivating behaviors that the individual is already familiar with, therefore only requiring support to be reminded of, or encouraged to perform the behavior [13]. The appropriation into a routine of behavior change appeals to a notion of behavioral engagement, but at the expense of cognitive engagement – where, as we have discussed elsewhere [4], the performance of behaviors becomes habitual.

Social Support and Social Media. Social engagement considers the importance of peer support to increase engagement with a system and to overcome motivational barriers to display a desirable behavior. It is therefore not surprising that social media sites like Facebook and Twitter have become increasingly popular platforms for the study of social support in the field of persuasive technologies [e.g., 21,22]. This engagement has massive potential for how we might design technologies for behavior change.

There is, however, also evidence suggesting that frequent use of social media is associated with a lower need for cognition [35]. Thus, although users of social media may have characteristics, such as a high need to belong to others, that make them more susceptible to persuasion, they may not be prone to persuasion through a direct route; that is, through the quality of information provided. Thus, research suggests that reflective and cognitive engagement are less likely to be associated with high levels of social media use.

Positive Reinforcement. Other strategies, often found with personal informatics systems [25], relate to how this data is fed back to the user. Most designs in this regard focus on strategies of *positive reinforcement,* presenting visual incentives to the user, to foster compliance with desirable behaviors. Such reinforcement drives affective engagement, which in turn may drive behavioral engagement. With *UbiGreen,* Fröhlich et al. [16] displayed a tree graphic to indicate green transportation activity, with the tree accumulating leaves, blossom or apples the more the individual uses environmentally friendly transportation. Persuasive designs in the field of sustainable HCI also commonly include visualizations of reduced energy consumption, carbon emissions or ecological footprints [17,20], or highlight money savings if the creation of waste is avoided [18].

Gamification and Achievements. In the context of behavioral change, gamification has been used to encourage positive behavior that the user would not normally engage in [30]. Gamification is defined as "*using game design elements in non-game contexts to motivate and increase user activity and retention*" [7]. Such engagement may tend towards intrinsic engagement, though this depends on how challenging and rewarding game elements might be. Gamification has been increasingly popular in both research and commercial systems. Design elements common to games, including scoreboards and badges, have been used to reward desirable activity [26], such as a regular use of an application.

Much like positive reinforcement, such game elements might increase affective engagement, where they are valued, indicate esteem or personal achievement. Even in professional contexts, the use of scoring systems has been demonstrated to increase use of an internal social networking site [34]. When combined with social networks that make these achievements visible, these features introduce an element of competition between, and playful awareness of others. For example, Foursquare (https://foursquare.com/) encourages users to check-in regularly by declaring the most active user in a location as 'mayor', but they can be replaced by another user if they fail to remain active. Thus, gamification also lends itself to social engagement, particularly in terms of competition and group identification. In these circumstances, a scoreboard can provide both a source of pride and a sense of shame when undesirable behavior is exposed.

Negative Reinforcement and Coercion. Far less research has examined the potential of negative reinforcements or coercion to promote behavioral change in HCI. Exceptions include research by Kirman et al. [23], who argue that behavioral change technologies should employ constructive aversive feedback alongside strategies for positive reinforcement to support the learning and maintenance of desired behaviors. Negative reinforcement in this context means that the performance of a behavior prevents or removes a negative response (e.g. a person may recycle to avoid disapproval by others). Engagement may be driven by avoidance of negative affect and through reflection on the actions that have led to negative outcomes. Foster et al. [15] have shown that a light form of coercion in the form of aversive feedback does not necessarily disengage users, as previously claimed by [5], but instead can function as a valuable component for achieving behavioral change. While ethical questions remain about the use of coercion, findings of this research revealed that aversive feedback can be a useful supplement in promoting behavioral change if designed carefully.

2 Interface Re-design

Following from the previous study with BinCam [4,32], a re-design of the system was undertaken following the potential and strategies to design for engagement and lessons learned from previous evaluations. Three strands of development were initiated, aiming to improve (1) system reliability (including WiFi and 3G connectivity); (2) feedback accuracy and frequency; and (3) overall engagement with the Facebook interface. Our focus here is on the third and final element, though the development of a more stable, reliable and trustworthy system contributed to a more robust experience for participants.

2.1 Design for Engagement

With the BinCam system, the Facebook application (short 'app') is the primary system front-end with the main goal to give users feedback on their recycling behavior and help them reflect on their own and other people's waste-related performances. Based on participants' experiences with the system we have suggested techniques to increase engagement with the Facebook application [32]. These included: a neat integration within the ecology of Facebook, challenges to promote group identification and competition on Facebook within and across different households, more frequent and varied visualizations for cognitive and reflective engagement, and improved opportunities to compare own waste-related achievements with other BinCam users. Below we position these within our framework for engagement and detail our strategies for redeveloping the app. The app offers a set of interactive elements to explore such as a *BinLeague*, including a variety of different visualizations of collected waste data, as well as creative *BinProfiles* of each BinCam bin in the system. *BinAchievements* are playful elements that can be gained through interactions with the interface, or engagements with specific recycling or food waste *BinChallenges* initiated by the *BinMan*.

BinMan. The BinMan is a virtual person on Facebook that is managed by an administrator of the BinCam system. The BinMan has a personal profile page on Facebook and personifies the BinCam system by posting recycling-related information on his wall, leaving comments, answering questions, and acting as a referee to the BinChallenges. The role of the BinMan is to improve the social component of the system and to facilitate users' social engagement with the system. As a social actor, he allows for the flexible and dynamic provision of support and knowledge, while simultaneously allowing for personalization and interactivity when responding to, or posting comments, thereby fitting into the ecology of Facebook.

BinLeague. The BinLeague was originally designed to give participants access to a record of the recycling activity of their household. Following poor precision in our previous studies using Amazon's Mechanical Turk, the tagging interface was re-designed to allow administrators access to the images and tag them for categories of *landfill*, *recyclable*, *compost* and *food waste* items. The BinLeague summarizes daily results for all bins in the system. Thus, it served as a personal informatics tool for reflective engagement and helped create a sense of in-group identification and out-group competition for social engagement. The page provides a variety of different visualizations of the scores, extending the previous BinCam interface design. As in the original interface, each score has a unique visual representation, e.g. the recycling score is represented with a tree sapling that grows taller the better the score. Daily statistics reflecting the bin usage for a specific day are presented as a 24-hour graph, with each thrown away item producing an incremental progression on this graph.

BinProfiles. Additional bin statistics are also displayed on the bin's profile page and contain information on the daily bin usage, graphing the number of items in the bin according to the four tagging categories outlined above. As an additional playful team-building experience, each household has been asked to choose a profile picture for their bin from a set of 18 images. Allowing for personalization and affective engagement, the profile picture personifies the system, so the BinCam becomes a mascot of each household. The start page of the BinCam app displays the list of bin profiles and showcases awards given for succeeding in the BinLeague and the BinChallenges.

BinChallenges. BinChallenges are managed manually by the BinCam administrator and delivered through the BinMan's news feed. The purpose of the challenges is to boost user interest when needed, by providing activities that might be intrinsically engaging, or which promote certain waste-related actions. Most of the challenges required participants to respond by being creative in using the system, for example: "The funniest message on non-recyclable waste wins". By promoting competition between households it might also be possible to increase social engagement with the system.

BinAchievements. The system of achievements defines a number of fixed goals for the user. It is aimed to increase user engagement with the system and to encourage more exploration of the interface. The achievement system is automatically administered and gives an immediate feedback to the user. All achievements can be divided into three major groups: regularly visiting the system, browsing images and leaving comments.

3 Evaluation

For a period of six weeks we deployed one BinCam bin in a total of six student houses in Newcastle upon Tyne, UK. Prior to the start of the study all members of the household were introduced to the system and completed a pre-study questionnaire on their food waste and recycling attitudes and behaviors, and the Facebook Intensity scale [12]. The Facebook Intensity scale is a measure of Facebook use, including measures of behavioral and emotional engagement. Example items include for instance “Facebook has become part of my daily routine” or “I feel out of touch when I haven’t logged onto Facebook for a while”. Following the study, participants were invited to either a focus group or an individual interview and completed a post-questionnaire. One member of each household was randomly selected for an individual interview and the remaining members took part in a focus group. In total, five participants completed individual interviews, with only one invited individual not responding. The six focus groups involved 27 of the participants, with one invited participant not attending. The individual interviews were carried out to gain a sense of how individuals perceived the system and particularly to consider how individual concerns might differ from those expressed in the group. Focus groups also allowed for the consideration of group dynamics and for the exploration of social and normative influences. In both conditions, participants related similar concerns. In the data presented below all names have been changed.

3.1 Participants

34 individuals completed the pre- and post- questionnaire for the study. The study sample was aged 18-27 ($\dot{x}$=21.12, sd.=1.93), of whom 20 were female. Two participants were in part-time education, with the remainder in full-time education. 17 were in the first year of third level education, three in the third year of an undergraduate degree, 12 were enrolled in Masters level education, and one undertaking a PhD. One student was an exchange student from an international university. All but one flat had students at different stages of education. Five households had 6 participants and one household had 5 participants. Three households were mixed gendered (3 female, 3 male) and three households were single gendered (one household of 5 males and two households of 6 females).

4 Results

4.1 Recycling and Food Waste Attitudes

As with our previous study, pre- and post- questionnaires revealed little change in participants’ attitudes to recycling and food waste. Participants, partially due to self-selection and social desirability, report strong positive attitudes towards sustainability from the outset. This leads us to reiterate our previous assertions that, within rational choice models, recycling and food waste might be better motivated by examining issues of awareness and perceived behavioral control. There were some significant changes in social aspects of recycling. Most participants for instance reported changes in the social aspects of waste disposal (e.g. “I ask other people for advice as to how I can keep food for longer”), waste disposal knowledge (e.g. “I think food waste is difficult to avoid”) and in feelings associated with waste disposal (e.g. “I recycle because I feel better if I do”).

4.2 Facebook Use

On average the participants had 449 friends and spent 90 minutes a day on Facebook in the week prior to deployment. The average Facebook intensity (FI) score for the sample was 3.52 (n=31, min = .91, max = 4.93), suggesting that the participants are above average Facebook users. Comparing changes in pre- and post-questionnaire items further suggests that Facebook use is inversely related to commitment to change (r=-.438, n=30, sig=.016) and contemplation of changing behavior (r=-.437, n=30, sig=.016).

In order to more closely examine possible relationships, Facebook intensity was also correlated with responses to recycling behavior. Questions correlating FI with recycling predominantly related to social concerns (e.g. “I listen to what my flat mates have to say”, “I ask other people for advice as to how I can keep food for longer”) and identity performance (e.g. “I try to conceal food waste that I dispose of”). FI was negatively correlated with concerns about the cost of food waste (e.g. “Throwing away food costs me money” and “I buy fresh food on special offer like buy one get one free or three for two”) and general concerns about food waste (“throwing away food bothers me”). FI also positively correlated with beliefs about local facilities (e.g. “We have adequate facilities in our local area to recycle”). In each case the correlations were moderate and although they generally point towards the findings of [35] no strong relationship between FI and recycling attitude is supported in this study.

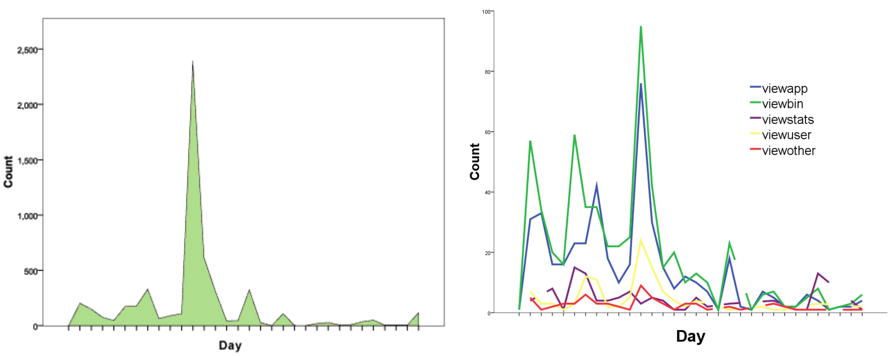


Fig. 1. & 2. Total daily activity (left) and daily levels for each activity, excluding viewing images (right). A large increase surrounds one user’s action 2 weeks into the project. Following this, there is a continued drop off in activity, with only minor increases coinciding with BinChallenges and activities at the end of the project.

4.3 Using the Interface

Of the 32 participants to complete the study, seven did not log into the Facebook interface. Behavioral engagement metrics for the remaining 25 participants were gathered from the BinCam application for a range of activities (see total activity in Figure 1). These were (with total number of actions, and percentage of overall activity): view application title page (444, 7.9%), view BinProfile (598, 10.7%), view BinPictures (4277, 76.3%), view user (115, 2.1%), view BinLeague (83, 1.5%), view daily statistics (22, 0.4%), view FAQ (12, 0.2%), and view BinChallenges (55, 1%). There were no significant relationships between scores on the FI scale and the use of the application or specific aspects of the system. There were no significant differences in access to or use of the system between households. Viewing BinPictures is a significant part of the overall activity. One individual, however, is responsible for 2124 image views, constituting 37.89% of all activity. This occurred predominantly in one sitting. During an interview he explained that he was motivated to do so to both gain achievements and to compete against another participant in a different household who had gained more achievement points than him. When figures for image viewing are removed, the application title page and BinProfile contribute over 78% of the activity (33.4 and 45% respectively). This suggests that the use of the system is somewhat limited to these features.

The majority of interaction occurs within the first two weeks with a peak in the second week. The peak of activity occurs as the participants become familiar with the application, after most participants have logged in, and as they begin to discuss it among themselves. Shortly afterwards, the level of activity drops off. In order to better understand the activity patterns, we can examine the daily activities, excluding image viewing (see Figure 2). Sophie summarizes this use pattern:

"*Erm... I would say... I logged in quite a bit in the beginning of the project... and then just as the weeks gone on I didn't bother anymore.... erm... I say...when one of the challenges went up...that made me log on again…*"

An analysis of variance and post-hoc tests suggest that the activity of only the top-most quartile was significantly different from the others for views on BinProfiles (F: 24.35, sig: .000), posting comments (F:8.34, sig: .001), visit application title page (F:39.83, sig: .000), view user profiles (F:8.28, sig: .001), and other activities such as viewing the FAQ (F:6.08, sig: .003). This suggests that a small number of users contribute significantly to the overall activity on the application and might be considered engaged with the system.

4.4 BinAchievements

A number of users were motivated to engage in searching for and gaining BinAchievements. Achievements were awarded for the engagement with the app in three categories: logging in, commenting and viewing images. In each case, achievements were incremented through progressively more engagement (e.g. view 1 picture, view 10 pictures, etc.). All participants who logged into the application received a BinAchievement for doing so. Users of the site received on average 4 achievements, or 58 achievement points, beyond the first achievement which was for logging in. There was no significant correlation between FI and achievements received.

The majority of BinAchievements were awarded in the first two weeks and only a small number of participants were motivated to gain achievements mid-way into the project. One user who was motivated to gain all achievements did so across one session. In general, however, the achievement trajectories suggest that only a small number of users were reminded of or engaged with achievements on the BinCam Facebook app in the 3rd and 4th weeks of the study. Brief resurgence of interest appears to be associated with a social BinChallenge and later the closing of the system.

5 Results from the Interviews and Focus Groups

The quantitative use data suggests that there was no sustained engagement with the Facebook application, but that some users experience intense engagement at the start of the study. We therefore look to qualitative interview and focus group data to understand why this might be the case and how, if at all, individuals had engaged with the system. Although our primary focus is on the social media interface, the ubiquitous system figures as a concern in how users' engage with the overall system.

5.1 Awareness, Guilt and Perceived Behavioral Control

Like in our previous work, significant findings from this study point to improved awareness among participants. In this study we gain a clearer understanding of how this occurs and how the ubiquitous system draws users to engage with their food waste and recycling behaviors. Specifically, we gain insight into the transition from under-awareness to mere awareness to engagement.

From Awareness to Routine. Participants are drawn to attend to the bin by its novelty, the awareness of it taking pictures, and the shutter sound of the camera when a picture is being taken, summarized by Eva:

"*Yeah, I think I was more aware as well cause like with it making the noise I think…It was the vibrating I think and also like… I don't know just because it's a different bin to what we have before. It was obvious like that gonna be uploaded.*"

As the presence of the BinCam bin leads to raised awareness, this awareness led to personal motivation to change:

Mary: *"…[the BinCam bin] makes you more aware and kind of want to do it right..."*

Participants experienced affective engagement with the system, feeling it morally correct to change their behavior. As the novelty of the system recedes over time, participants begin to lose some awareness of it as a persuasive system and therefore have diminished cognitive engagement with it. The bin itself though remains cognitively demarcated from the recycling bin:

Sophie: *"Erm... it kinda for one second got on my mind that bin was only stood for landfill kinda thing. It didn't... like I didn't think: "Oh it's going to take the picture now." I wasn't like listening out for the sound every time…erm….but erm at the beginning I was always thinking: "Right... ok... consciousness decision which bin am I gonna use…".*"

Thus, participants think of the BinCam as fundamentally different, and for landfill waste only and not as a recycling bin. Consequently, their engagement with the bin is behaviorally and cognitively different. This becomes a practiced routine and relatively unconscious knowledge. Further exploration reveals that this transformation from awareness to engagement may be both the forming of a habit and the avoidance of negative affect.

In our previous study, participants' sense of the system as one for behavior change had led to feelings of guilt, primarily about differences between attitudes and actual behavior. In this study, the drive to 'do it right' still led some participants to feelings of guilt. It became clearer in this study, however, that participants differentiate this guilt from feelings of shame:

Tom: *"…I did feel guilted in to doing sometimes but I never felt ashamed because my guilt preceded the shame."*

Feelings of guilt that arose were induced by personal reflection on behavior and motivated individuals to change behavior to reduce guilt. Thus, it is clear that the BinCam system, at least for some participants, promoted reflective engagement that resulted in negatively experienced affective engagement.

Persistent Awareness. One area where this diffusion of awareness did not occur was in the case of food waste. In all cases, the participants either did not have access to composting facilities, or would not be able to use compost. Consequently, although they felt guilty about food waste, there was little they could do:

Neil: "*I think we are really good at recycling in our house but in terms of food waste we are probably not so but... I think that's more with the university aren't providing... a compost bin."*

Thus, at every occasion where food waste was to be put in the BinCam bin, participants' attention was drawn to the behavior. Participants, who felt strongly about food waste found ways to adjust to minimize feelings of guilt or heightened awareness of inappropriate behaviors:

Sophie: "*it's made me just kind of just reduce my portion size and then think about how much stuff I'm throwing away and trying to catch things before they go out of date and stuff like that."*

These new strategies are likely to reduce engagement with the bin while increasing engagement in positive food waste behaviors. Furthermore, while this might reduce overall interaction and awareness of the system, it is likely to decrease the likelihood that the system's presence in food waste behaviors becomes routine. That is, the presence of the BinCam bin continued to draw attention to itself in food waste behaviors.

5.2 Gamification

With the participants we wanted to further explore their experiences with gamification. The achievements were designed to be discoverable, and the app provided information on how achievements could be received. For most of the participants, gaining achievements was first unintentional, following which they were motivated to find more. However, the value of the achievements for motivation quickly reduced, particularly where the activity involved was repetitive:

Peter: "*... erm... I got the achievement that were easy-ish and it took me a few minutes I guess. Erm and then some achievements like you said like the 250 [viewing pictures] it was just like the same as the 50 one but just more…It's kind of... cause it's just like a repeat of the same I guess I just thought...[it is not worth it]*"

For others, despite initial excitement, there was no further motivation to engage until after the study:

Clare: "*I wanted to get more.*
Jill: *Yeah. I should have gone back to get more.*
Clare: *And maybe log on more.*"

For some participants, the design for gamification of recycling was motivating, and, as previously mentioned, one participant was driven to contribute almost 40% of all activity by wanting to compete with someone else. However, for most users this was not the case and many did not feel they might ever be engaged with such an activity:

Jayne: "*I didn't really... I wasn't really interested in looking at what's in my own bin [...] let alone what's in other people's bin... or playing like "inter-bin-related games"*".

The challenges which drew most interest were those that involved some aspect of household team work, such as leaving a funny message in the bin or taking a picture of the group with the bin. Such challenges were appreciated by most participants, and they were among the only images to elicit cross-household activity:

Jayne: "*[...] We looked at pictures of other people's challenge... that joke challenge... we looked at that…We didn't look at pictures of people's like...'crap in the bin'*"

5.3 Facebook Ecology and Daily Routines

The decision to design for Facebook was based on its proliferation as a social and engaging platform. We had also, from our previous work [32], highlighted that the system must more closely align with the existing Facebook ecology. The behavioral engagement demonstrated, however, suggests that the application did not harness the benefit of being associated with Facebook. This appears to have been due to three difficulties in assimilating into the ecology of Facebook. First, some users simply do not use Facebook frequently. This prevented them from engaging with the BinCam app entirely. Second, some users did not use Facebook apps frequently.

Bill: "*I know... I well, to be honest I don't really use Facebook that much and I've never... I don't think I have ever actually use an app on Facebook or anything else*".

Finally, the configuration as an app impeded the extent to which messages from the BinMan were shared with users. Furthermore, most users suggested that the BinMan could post with higher frequency – on the one hand to increase visibility, and on the other to increase the amount of interaction between users and the system.

Although BinCam is designed to be situated within everyday practices in student households, some of the practices and routines of student households also decrease the possibility for social support and discussion around the system. In one household, the participants reflected on their routines, in saying:

Sam: " *... it's rare that we are all in together…I see you like once in three days…*".

Peter: *"[laughing] Same here. We cook in different times and stuff as well usually... so there wasn't really mentioned of it [BinCam].*"

6 Discussion

Users' engagement with the BinCam bin and Facebook application revealed mixed effects. The main impact of the BinCam system continues to be in raising awareness of recycling and food waste behavior. The audio cue from the bin serves as a reminder throughout their engagement, and as previously noted [4], acts as a post-actional cue for reflection. Thus the system draws attention to itself, which raises reflective engagement in the individual. This brought about a change in participants' behaviors where they reduced the amount of waste they produced.

The fact that some people do not engage with online and competitive games is not a new finding. The motivation to engage individuals with competitive game elements is, among other things, gender differentiated [19]. In critical literature on gamification [7], there is an assertion that gamification must mean more than simply awarding points and badges and showing these on leader boards. The empirical data on the use of BinCam suggests that this is the case. Although we were able to achieve two weeks of engaged use, there was little further use of the application. If we understand a 'game' as something in which we are challenged and must overcome challenges [7], then perhaps recycling is not such an activity. We have however focused on engagement with the system and not recycling itself.

The integration into everyday practices of the social media platform was not always successful. As stated, the use of Facebook and Facebook apps was not always within the routines of users' everyday behavior. In the case of waste disposal, simply being subsumed into everyday practice, particularly when those practices are habitual, means that it is difficult to create awareness or to change behavior. From this study we have examples of how designing to disrupt everyday practice both worked and did not work to create engagement. In the case of the audio cue from the bin, this was sufficient to disrupt the routine of waste disposal. While the post-behavioral audio cue did not change behavior in the moment, it created reflective engagement as it drew attention to the unconscious performance of it.

BinPictures were described as unappealing and lacking interest. Despite this, they received far more activity than any other aspect of the system – even when excluding extreme users. This is at least partly due to the influence of achievements and gamification. Thus, despite being potentially uninteresting, BinPictures had the most appeal as an interaction. Moreover, the mundane and particular nature of waste disposal meant that most individuals were not interested in viewing images of waste. There was little evidence here of either intrinsic or affective engagement. There was no real added value in seeing pictures from the bin, because people didn't care about them.

However, social challenges did increase participation, and many participants reported these challenges to be the most enjoyable aspect of the study. This is in line with [19] that games including meaningful social interaction can increase appeal. And although few participants sought support through the system, many participants reported discussing recycling issues within their household. This appears to be particularly important for the acquisition of recycling knowledge. Moreover, such sharing may expand the cultural knowledge [31] that underpins individuals' recycling knowledge and provides the means and skills to adapt to, for instance, new expectations about what can or cannot be recycled. It is therefore critical that research continues to explore the specific mechanisms through which competitive and non-competitive social engagement can be fostered in interactive systems for behavior change.

The use of the BinMan as a conduit for information on Facebook did not disrupt participants. In fact, several wished for more feedback and notifications from the BinMan. With the use patterns for Facebook, where participants logged in occasionally, comments posted could be easily overlooked and were arguably not of a high enough frequency. This is not to suggest that bombarding participants with messages will achieve better engagement, but that such interventions should be tailored to the practices and expectations of participants. More visible or direct notifications outside of the Facebook ecology might be more effective (e.g. e-mail, SMS messages, or a shared, open visualization in the home and near the bin). Thus the ways in which persuasive technologies explicitly draw attention to themselves needs consideration.

7 Conclusion

This paper presented a user study of our re-design of the BinCam interface on Facebook. Although most participants use Facebook (and other social media) they do so in particular ways, at particular times, and fit these activities around their everyday routines. The social and material practices that are shared among households do not directly, or necessarily, involve social media. This varied between households, some were more engaged than others, but across almost all of them, the use of the BinCam Facebook application was largely an isolated and lone activity, as were the activities of recycling and food waste.

When we review this data we are left with, among others, a recurring question: is recycling and food waste simply so uninteresting that we cannot get people to engage with it in the long term? We do not believe that this is the case, and we consider our research with the BinCam system to present some progress in this regard. Significant challenges of course still remain. In particular, we have highlighted the necessity for integration of multiple forms of engagement and feedback into everyday life as a central concern.

Acknowledgements. We would like to thank all our participants, Mat Kipling, and Gu-Young Kang. This project has been supported through the European FP7 Marie Curie Action Balance@Home project and the RCUK Digital Economy Research Hub SiDE: Social Inclusion through the Digital Economy.

References

1. Astin, A.: Student involvement: a developmental theory for higher education. Journal of College Student Personnel 25(4), 297–308 (1984)
2. Bødker, S.: When second wave HCI meets third wave challenges. In: Proc. NordicCHI 2006, pp. 1–8. ACM Press (2006)
3. Brynjarsdottir, H., et al.: Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In: Proc. CHI 2012, pp. 947–956. ACM Press (2012)
4. Comber, R., Thieme, A.: Designing beyond habit: opening space for improved recycling and food waste behaviors through processes of persuasion, social influence and aversive affect. Personal and Ubiquitous Computing (online first)
5. 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)
6. Csikszentmihalyi, M.: Finding Flow: The Psychology Of Engagement With Everyday Life. Basic Books (1998)
7. Deterding, S., et al.: From game design elements to gamefulness: defining "gamification". In: Proc. MindTrek 2011, pp. 9–15. ACM Press (2011)
8. Deutsch, M., Gerard, H.G.: A study of normative and informational social influence upon individual judgment. Journal of Abnormal and Social Psychology 51, 629–636 (1955)
9. Dewey, J.: How we think. Dover Publications (1997)
10. DiSalvo, C., Sengers, P., Brynjarsdóttir, H.: Mapping the landscape of sustainable HCI. In: Proc. CHI 2010, pp. 1975–1984. ACM Press (1984)
11. Doherty, G., Colye, D., Sharry, J.: Engagement with online mental health interventions: An exploratory study of a treatment for depression. In: Proc. CHI 2012, pp. 1421–1430. ACM Press (2012)
12. 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)
13. Fogg, B.J.: A behavioral model for persuasive design. In: Proc. Persuasive 2006 (2006)

14. Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann (2003)
15. Foster, D., et al.: Wattsup?: motivating reductions in domestic energy consumption using social networks. In: Proc. NordiCHI 2010, pp. 178–187 (2010)
16. Fröhlich, J., et al.: UbiGreen: Investigating a mobile tool for tracking and supporting green transportation habits. In: Proc. CHI 2009, pp. 1043–1052. ACM Press (2009)
17. Fröhlich, J., Findlater, L., Landay, J.: The Design of Eco-Feedback Technology. In: Proc. CHI 2010, pp. 1999–2008. ACM Press (2008)
18. Gartland, A., Piasek, P.: Weigh your waste: A sustainable way to reduce waste. In: Ext. Abst. CHI 2009, pp. 2853–2858. ACM Press (2009)
19. Hartmann, T., Klimmt, C.: Gender and Computer Games: Exploring Females' Dislikes. Journal of Computer-Mediated Communication 11(4), 910–931 (2006)
20. Holmes, T.: Eco-visualization: Combining art and technology to reduce energy consumption. In: Proc. C&C, pp. 153–162 (2007)
21. Jamison-Powell, S., et al.: I can't get no sleep": discussing #insomnia on twitter. In: Proc. CHI 2012, pp. 1501–1510. ACM Press (2012)
22. Jianqiang, D.S., et al.: Farmer's Tale: A facebook game to promote volunteerism. In: Proc. CHI 2011, pp. 581–584. ACM Press (2011)
23. Kirman, B., et al.: Improving social game engagement on facebook through enhanced socio-contextual information. In: Proc. CHI 2010, pp. 1753–1756 (2010)
24. Lehmann, J., Lalmas, M., Yom-Tov, E., Dupret, G.: Models of user engagement. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (eds.) UMAP 2012. LNCS, vol. 7379, pp. 164–175. Springer, Heidelberg (2012)
25. Li, I., Dey, A.K., Forlizzi, J.A.: Stage-Based Model of Personal Informatics Systems. In: Proc. CHI 2010, pp. 557–566. ACM Press (2010)
26. Liu, Y., Alexandrova, T., Nakajima, T.: Gamifying intelligent environments. In: Proc. Ubi-MUI 2011, pp. 7–12. ACM Press (2011)
27. Mankoff, J.C., Matthews, D., Fussell, S.R., Johnson, M.: Leveraging social networks to motivate individuals to reduce their ecological footprints. In: Proc. of HICSS, pp. 1–10 (2007)
28. McCarthy, J., Wright, P., Wallace, J.: The experience of enchantment in human – computer interaction. Personal and Ubiquitous Computing 10, 369–378 (2006)
29. Peters, C., Castellano, G., de Freitas, S.: An exploration of user engagement in HCI. Proc. AFFINE 9(3) (2009)
30. Shiraishi, M., et al.: Using individual, social and economic persuasion techniques to reduce CO2 emissions in a family setting. In: Proc. Persuasive 2009 (2009)
31. Strengers, Y.: Designing eco-feedback systems for everyday life. In: Proc. CHI 2011, pp. 2135–2144 (2011)
32. Thieme, A., Comber, R., et al.: "We've bin watching you.": designing for reflection and social persuasion to promote sustainable lifestyles. In: Proc. CHI 2012, pp. 2337–2247. ACM Press (2012)
33. Thieme, A., et al.: Lovers' box: Designing for reflection within romantic relationships. International Journal of Human Computer Studies 69(5), 283–297 (2011)
34. Thom, J., Millen, D., DiMicco, J.: Removing gamification from an enterprise SNS. In: Proc. CSCW 2012, pp. 1067–1070. ACM Press (2012)
35. WRAP. Waste and resources action programme, `http://www.wrap.org.uk` (accessed December 29, 2010)
36. Zhong, B., Hardin, M., Sun, T.: Less effortful thinking leads to more social networking? The associations between the use of social network sites and personality traits. Computers in Human Behavior 27(3), 1265–1271 (2011)

OpinionBlocks: A Crowd-Powered, Self-improving Interactive Visual Analytic System for Understanding Opinion Text

Mengdie Hu[1], Huahai Yang[2], Michelle X. Zhou[2], Liang Gou[2], Yunyao Li[2], and Eben Haber[2]

[1] Georgia Institute of Technology
mengdie.hu@gatech.edu
[2] IBM Almaden Research Center
{hyang,mzhou,lgou,yunyaoli,lgou,ehaber}@us.ibm.com

Abstract. Millions of people rely on online opinions to make their decisions. To better help people glean insights from massive amounts of opinions, we present the design, implementation, and evaluation of OpinionBlocks, a novel interactive visual text analytic system. Our system offers two unique features. First, it automatically creates a fine-grained, aspect-based visual summary of opinions, which provides users with insights at multiple levels. Second, it solicits and supports user interactions to rectify text-analytic errors, which helps improve the overall system quality. Through two crowd-sourced studies on Amazon Mechanical Turk involving 101 users, OpinionBlocks demonstrates its effectiveness in helping users perform real-world opinion analysis tasks. Moreover, our studies show that the crowd is willing to correct analytic errors, and the corrections help improve user task completion time significantly.

Keywords: Text analytics, text visualization, self-improving, crowd-sourcing.

1 Introduction

Hundreds of millions of people voice their opinions online daily. Large portions of these opinions are product reviews about "experienced goods"—products or services of which characteristics are difficult to observe in advance but can be learned after purchase [21]. Not only do product reviews provide great value to individual consumers and influence their purchasing decisions [24], but they also impact the product or service strategies of businesses [32]. However, gaining insights becomes increasingly challenging for users as the number of reviews gets larger and larger [7, 12, 13, 31].

To help users wade through a large number of reviews, commercial sites often employ one of two approaches. One approach, used by sites such as Amazon.com, lets readers vote on the helpfulness of each review, and directs future readers to the most helpful reviews. The other approach, applied by sites such as Bing Shopping and Google Product Search, provides an overview of the most frequently mentioned product/service features, and the overall sentiment expressed in a collection of reviews. Users can then filter the reviews based on the identified features.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 116–134, 2013.

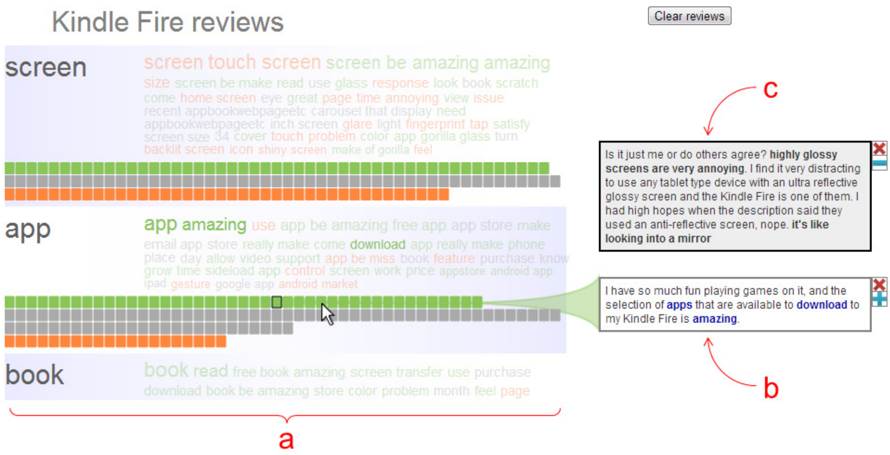


Fig. 1. The interface of OpinionBlocks: (a) a system-generated aspect-based summary; (b) a system-extracted review snippet; (c) the full text of a review.

While a few high-quality reviews or the aggregated sentiment may provide useful information, previous research shows that users often desire finer-grained understanding of the reviews [7, 31]. In particular, people process information in an attribute-driven manner in the absence of actual products (e.g., online shopping) [15]. In such cases, people examine the attributes of a product to evaluate whether the product fits their purchase goal (e.g., buying a camera for underwater adventures). In addition, the positive or negative sentiment expressed by the reviewers toward *each* attribute helps justify the suitability of the product [31].

To facilitate attribute-driven evaluation of products, a number of systems produce an aspect-based summary, including the extraction of sentiment toward each of the aspects [7, 13, 16, 31]. Among these systems, several recent ones use noun-adjective pairs to summarize the aspects of a product/service (noun) and the sentiment (adjective) toward each aspect [13, 31]. However, this approach has several limitations. First, they cannot handle implicit opinions. For example, they cannot extract aspects "*weight*" or "*size*" implied by the expression "*it is light and portable*" [31]. Second, they do not deal with conflicting opinions expressed by different reviewers. For example, one reviewer raves "*the screen is fantastic*", while the other complains "*positively claustrophobic in terms of screen usage*". In such cases, it is unclear whether multiple noun-adjective pairs would be displayed or one noun (e.g., "*screen*") would be associated with multiple adjectives (e.g., "*fantastic*" and "*claustrophobic*"). Third, the performance of these systems is limited by the imperfections in the underlying natural language processing (NLP) techniques. Because of the flaws in NLP (e.g., classifying "impeccable" as a negative sentiment), users may find certain summaries mystifying [31].

To improve the quality of aspect-based opinion summarization, researchers have developed sophisticated NLP techniques for aspect extraction and sentiment analysis (see Section 2.2). However, due to the challenging nature of the problems, even with a large amount of domain-specific training data, state-of-the-art NLP techniques can only achieve 50% to 85% accuracy for either of these tasks. The imperfections in text analytic results often lead to user frustrations and even distrust in the system [31].

To address the challenges mentioned above, we have developed a novel interactive visual analytic system, OpinionBlocks, to meet two design goals: (1) automated creation of an aspect-based, effective visual summary to support users' real-world opinion analysis tasks, and (2) support of user corrections of system text analytic errors to improve the system quality over time. On the one hand, meeting the first goal motivates users to correct system errors. On the other hand, achieving the second goal improves the system quality, which then better aids users in their tasks.

To achieve the first goal, OpinionBlocks employs advanced NLP technologies to automatically create and present users with a fine-grained, aspect-based visual summary of opinions. As shown in Figure 1, the created visual summary allows a user to gain insights into a collection of reviews at multiple levels:

1. Frequently mentioned aspects of a product/service, including those *explicitly* and *implicitly* expressed in the reviews (Figure 1a).
2. The description of each aspect in a form of key phrases, a set of associated review snippets, and the inferred sentiment of each key phrase and snippet (Figure 1b).
3. The full review containing extracted aspects (Figure 1c).

To achieve our second design goal, OpinionBlocks allows users to interact with the visual summary to amend analytic errors (Figure 2). It then aggregates user contributions to update and improve the visual summary for future users.

To demonstrate the effectiveness of OpinionBlocks in meeting both design goals, we conducted two crowdsourced studies on Amazon Mechanical Turk involving 101 users for the analysis of 18,000 reviews of the Amazon Kindle Fire (2.8 million words). Our results show that more than 70% users successfully accomplished non-trivial opinion analysis tasks using OpinionBlocks. These tasks involve answering questions beyond the capability of existing systems, such as "*What is the most common use of the product*?" and "*Which aspect received most conflicting reviews?*". Furthermore, our studies show that users are not only willing to use our system to correct text classification mistakes, their corrections also produce high quality results. The participants in our study successfully identified many mistakes and their aggregated corrections achieved 89% accuracy. Incorporating the crowd corrections, OpinionBlocks is also able to help users significantly improve their task completion time. As a result, OpinionBlocks offers two unique contributions:

- It supports real-world, opinion analysis tasks beyond that of existing visual opinion analysis systems.
- It leverages the power of the crowd to self-improve the quality of the text analytic results and compensate for the limitations in today's NLP technologies.

In the rest of the paper, we present the details of OpinionBlocks after an overview of related work. We then describe our two crowdsourced studies and their results. Finally we discuss limitations and implications of our work before concluding.

2 Related Work

Our work is related to four main areas of work across HCI and text analytics.

2.1 User Interfaces for Understanding Opinion Text

To better help users extract insights from a large number of online reviews, researchers have developed various interactive systems. For example, Faridani et al. created Opinion Space, an interactive tool that allows users to visualize and navigate collected opinions [9]. Yatani et al. developed Review Spotlight, which presents a word-cloud summary of online reviews in noun-adjective pairs [31]. Carenini and Rizoli built a multimedia interface that facilitates the comparison of different reviews [7]. More recently, Huang et al. presented RevMiner, an interactive system that summarizes reviews in noun-adjective pairs to be presented in a compact mobile phone interface [13]; and Rohrdantz et al. designed a visualization system that supports feature-based sentiment analysis of time-stamped review documents [26]. Similar to these efforts, our work aims at creating summaries of online opinions to help users in their decision-making processes. However, we go beyond existing systems to help users answer more complex analytical questions, such as identifying the aspects with the most conflicting reviews. While these systems are limited by the NLP techniques they employ, OpinionBlocks also leverages crowd input to compensate for its deficiencies in text analysis and improve its quality over time.

2.2 Opinion Mining and Sentiment Analysis

To help users digest massive amounts of online reviews, an active research area intersecting NLP and machine learning is known as opinion mining. State-of-the-art opinion mining technologies automatically extract aspects discussed in the reviews, and identify the sentiment expressed toward each aspect. Comprehensive reviews of the field can be found in [23, 17, 18], and these works point out limitations with aspect-based sentiment analysis: existing techniques are yet to go beyond parsing relatively simple sentence structures and modeling sentiment words (often domain specific) within sentences. And few can handle implied opinions well. Because of these difficulties, even the latest work published in this field can only achieve accuracy scores ranging from 50% to 85% for only aspect extraction or sentiment analysis alone, depending on the domain and training data [14, 17, 18, 19, 20, 29]. Moreover, achieving this level of accuracy often requires large amounts of training data (e.g., labeled sentences indicating aspect and sentiment expressions), which is often difficult and costly to obtain, especially when covering reviews for a diverse set of products [17]. While OpinionBlocks employs state-of-the-art opinion mining techniques to create an initial summary of reviews, it supports user interactions to rectify the imperfections in machine-generated summaries and improve the summary results over time. Furthermore, our implicitly crowd-sourced user inputs become valuable training data for the NLP community [28].

2.3 Interactive Machine Learning

Our work is also related to research efforts in interactive machine learning, where a machine learning process is augmented by human intelligence to improve the results. For example, Patel et al. presented a development environment that helps developers find and fix bugs in machine learning systems [25]. Amershi et al. developed systems that can iteratively learn the desired results based on end-user interaction behavior [3, 4]. Similar to these efforts, our work also aims to improve machine intelligence (i.e., text analysis results) through user interaction. However, while prior work focuses on leveraging individual users to improve machine learning, we focus on leveraging the wisdom of the crowd to improve analysis of unstructured text, which presents unique challenges as described later (e.g., reconciling crowd inputs).

2.4 Crowd-Powered Systems

Since OpinionBlocks is designed to leverage the crowd to identify and amend system imperfections in text analytics, it is related to an emerging research area on creating crowd-powered systems. This new class of software systems combines machine and human intelligence to solve problems that are extremely difficult or impossible for either approach alone. For example, Soylent guides the crowd on Amazon Mechanical Turk to rewrite and shorten text on demand [5]. n.fluent employs both machine translation and online crowd to help translate documents[1]. Carlier et al. combines content analysis and crowdsourcing to optimize the selection of video viewports [8]. While existing crowd-powered systems explicitly solicit a crowd's help (e.g., via Amazon Mechanical Turk) and use the results to help others, OpinionBlocks leverages its own users as the crowd implicitly, and motivates them to perform tasks that ultimately benefit both themselves and others (e.g., correctly identifying both positives and negatives of a product aspect).

3 OpinionBlocks

OpinionBlocks is a web-based system with three key components: a visual interface, a text analytic component, and a user feedback integration component. Below we describe each of the components, including our design rationales.

3.1 Interactive Visualization

The visual interface is designed to support two main user tasks: interacting with the generated visual summary and the original reviews, and correcting system errors in text analytics.

[1] https://www.ibm.com/developerworks/mydeveloperworks/blogs/c7f41400-4eb9-477c-b6fb-042466407259/?lang=en

Visual Features to Support User Decision Making

OpinionBlocks aims at aiding users in their information-driven decision-making processes. Based on previous research [31, 13, 15] and our own informal user studies (interviews with 10 colleagues who recently made a major purchase), we learned that a user's first step is to gain an overall impression of the important aspects of a product from available information. Our visual interface thus consists of two main parts. As shown in Figure 1, the left panel displays a visual summary of all the major aspects extracted from a set of reviews. The right panel is initially empty but shows relevant review snippets as a user interacts with the visual summary on the left.

A generated visual summary is made up of a set of aspect blocks (Figure 1a). From top to bottom, the aspects are ordered by their number of mentions in a review collection[2]. Each aspect block further consists of three parts: (1) the aspect name, (2) a text cloud of keywords and phrases describing the aspect, and (3) a set of colored squares, each of which represents a review snippet describing the aspect. Automatically extracted from a review document (see below), a *review snippet* includes a sentence that expresses opinions toward the aspect. Three colors are used to encode the sentiment expressed in a snippet: green (positive), gray (neutral), and orange (negative). The words and phrases in the text cloud are extracted from the snippets, and are colored based on the aggregated sentiment orientation of the relevant snippets. The colored squares are placed in different rows by their sentiment orientation, facilitating the comparisons of contrasting sentiments in each aspect (e.g., how many positive versus negative comments for the *Screen* aspect?) and across all the aspects (e.g., which aspect received most conflicting reviews?).

Our design is motivated by previous research and our own study that review readers tend to form and adjust their impression of opinions by looking for most discussed and most debated aspects, and they tend to verbalize their impression with short descriptive phrases [31]. Thus we designed the colored snippet boxes to support explicit comparison of comment frequency and polarity of sentiment under different aspects. And we help users highlight review snippets by keywords and phrases (Figure 2 Left).

Furthermore, readers often wish to see the concrete evidence behind the extracted aspects and sentiment in a summary [16, 26]. OpinionBlocks enables users to "drill-down" through clicking or hovering on the visual elements, allowing them to see snippets associated with blocks, keywords associated with snippets, snippets associated with keywords, or even the full context of the original reviews (Figure 2 Right).

---

[2] Unlike existing systems, which count how many reviews contain an aspect, we compute how many *sentences* refer to an aspect. We thus decided not to show the count to avoid potential confusion.

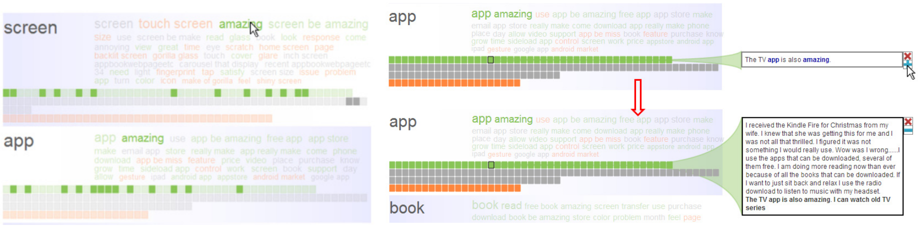


Fig. 2. Left: Hovering over the keyword "amazing" highlights all snippets containing the word. Right: Clicking on the "+" button next to a review snippet brings up the full review.

Interactive Features to Support User Feedback

As discussed earlier, one of our main design goals is to leverage the power of the crowd to identify and correct system errors, in particular, NLP errors that occurred in review analysis and summarization. Yatani et al. [31] suggests that showing the contextual text behind the phrases and sentiment classification helps compensate for the imperfect analytic results. However, we wish to take a step further and encourage the users to identify and correct text analytic errors. By correcting the errors, users not only obtain a more accurate visual summary for themselves, but also help future users of the system. We have identified four major types of system errors:

1. *Snippet omissions:* snippets that contain an opinion but were not extracted.
2. *Erroneous snippet extraction:* snippets without meaningful opinions
3. *Erroneous aspect:* snippets classified with the wrong aspect
4. *Erroneous sentiment:* snippets associated with the wrong sentiment

OpinionBlocks focuses on leveraging users to fix the last three types of errors, since identifying the first type of errors would require the users to be familiar with the entire review corpus. To rectify the errors, users can drag the colored square representing a misclassified snippet to the correct aspect or sentiment row (Figure 3 Left), or to somewhere out of the display area entirely if the snippet contains no meaningful opinion. Users can also click on an aspect name and change it to something more appropriate (Figure 3 Right).

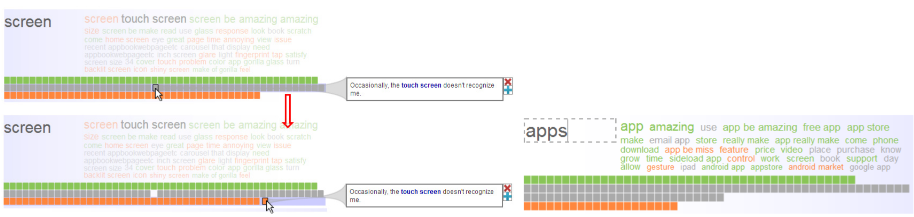


Fig. 3. Left: Moving a snippet misclassified as "neutral" to the "negative" row. Right: Changing the name of an aspect.

3.2 Opinion Mining and Sentiment Analysis

To generate the information used in our visualization, OpinionBlocks performs a four-step process: 1) review snippet extraction, 2) aspect extraction, 3) keyword extraction, and 4) sentiment analysis.

Review Snippet Extraction

From a collection of reviews of a product, OpinionBlocks extracts a set of review snippets that describe various aspects of the product. To extract a review snippet, OpinionBlocks first uses the OpenNLP parser [22] to obtain a parse tree for each sentence in a review. It then builds subject-verb-object (SVO) triples based on the parse tree. For each SVO triple, it checks whether the lemma of the verb matches a selective list of verbs (e.g., be, look, appear, etc.) from VerbNet [27], which are often associated with various aspects mentioned in a review. If there is a match, OpinionBlocks then keeps the subject of a SVO triple as an aspect candidate and the sentence containing the SVO triple as a review snippet. For example, given a sentence "*The display is made of Gorilla Glass, which is highly damage resistant*", the extracted SVO triple is: [*the display*, *make*, *Gorilla Glass*]. The sentence itself is a review snippet, and the subject "*the display*" then becomes an aspect candidate.

Note that we generate aspect candidates by considering only noun phrases that are also subjects of a restrictive subset of sentences in the review texts (by requiring their verbs to match a limited list). This approach is inherently resistant to noise introduced by common contextual information, such as prepositional phrases and discussions irrelevant to product aspects (e.g., detailed life experience like "*I tried out several different magazines*").

Aspect Extraction

Aspect extraction is to identify frequent *n*-grams from aspect candidates. Specifically, we first tokenize each aspect candidate and lemmatize its tokens with the Stanford Natural Language Processing Package [2]. Next, we extract all possible *n*-grams of size 3 from each candidate (or the candidate itself, if its length is shorter than 3), remove any stop word at the beginning or end of the n-grams, and calculate the frequency for each unique *n*-gram. Our preference of longer *n*-grams (e.g. tri-gram vs. bi-gram) is intentional: we observed that longer *n*-grams are typically more informative than shorter ones and thus are better at conveying concrete information to users. We then select and use the top-*K* (*K* is adjustable in our system) most frequent *n*-grams as a set of extracted aspects to summarize a collection of reviews.

We conducted several experiments to investigate whether our approach of aspect extraction can generate a consistent set of aspects given different sizes of the review collections. Here, we used the top-*K* aspects with the full review collection as the base line to investigate the performance of our approach with different sample ratios. Two metrics are employed here: Spearman's rank correlation coefficient (rho) [10] and coverage rate, where *rho* measures the correlation of two ranks of top-*K* aspects, and *coverage rate* measures the fraction of the top-*K* aspects from the full collection that also occur in the top-*K* aspects from the subset of the collection. The two metrics were computed using twenty sample ratios. We performed ten test runs for each sample ratio and averaged the two metrics over the ten runs.

Figure 4 shows our experiment results. On the left, all reported *rho* values are over 0.8, which indicates that the top-*K* aspects identified with the samples are positively correlated to the aspects identified with all reviews (all values are significant). We also find that even with a small sample ratio of 0.35, the top-10 aspects have the exactly same rank as those identified using the full collection. The performance for top-20 and top-30 aspects with our approach is also very promising. For coverage rate, with a sample ratio of 0.35, our approach yields very good coverage (> 0.95) for top-10 aspects and around 0.8 for top-20 and top-30 aspects. As a result, our aspect extraction generates consistent results over different sizes of review collections.

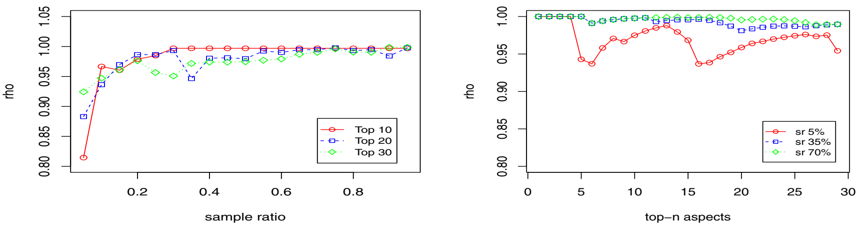


Fig. 4. Left: Spearman's rank correlation coefficient (rho) of top-K aspects with different sample ratios. Right: Coverage rate of top-K aspects with different sample ratios.

Keyword Extraction

To enrich an aspect-based summary, we also extract keywords from the relevant snippets for each aspect. We identify *n*-grams (unigrams, bigrams and trigrams) of all words and their frequencies from a collection of snippets related to an aspect. The *n*-grams with high frequency are used as keywords to describe each aspect. Because such a keyword relates to both an aspect and a snippet, we also use these keywords to index snippets. This way, keywords can be highlighted and easily spotted in a snippet when a user examines the snippet associated with an aspect (Figure 1b).

Sentiment Analysis

We use a simple lexicon-based approach [12] to infer the sentiment expressed in each snippet. This approach uses a public sentiment lexicon of around 6800 English words [1] to determine word sentiment orientation (positive or negative). We first tokenize and lemmatize a review snippet, and remove all stop words. The polarity of a snippet s is decided by its sentiment score $S(s)$, where $S(s)$ = |positive words in s| - |negative words in s|. If $S(s) > 0$, then s is considered positive; if $S(s) = 0$, then s is neutral; otherwise s is negative. If the verb of a SVO-triple contained in a snippet is associated with negation (e.g. "is not"), this simple method may not work. In such a case, we set the sentiment score of the snippet to 0 (neutral).

3.3 User Feedback Integration

As mentioned earlier, a user can interact with the system-generated visual summary and change parts of the summary, e.g., editing the displayed name of an extracted aspect and modifying the sentiment orientation of a snippet. When the user makes a correction, OpinionBlocks does two things: (1) updates its interface for *this user only* to reflect the user input, and (2) sends the user input to the back-end server and stores it in a database. Currently a system administrator decides when to incorporate user feedback to update the system interface for *all* its users.

To incorporate user feedback, OpinionBlocks first selects qualified user changes among all the inputs, and then uses them to update the system. Since users may make mistakes, move things around randomly, or even try to game the system, not all user feedback can be trusted. Similar to adopting crowd-sourced results [5], OpinionBlocks incorporates user feedback only when multiple users report the same error and propose the same solution. It checks the number of identical user-corrections made against a threshold. The threshold is now set by the system administrator and may be different for different user groups (e.g., trustworthy user population versus the general public). For our user studies with the Kindle Fire reviews, we used three as the threshold. That is, if three or more users made the same change, the change is then adopted. For example, the review snippet *"The touch screen has given me no problem so far"* was misclassified as negative by OpinionBlocks. Six participants in our study moved this snippet to the positive row. Thus, OpinionBlocks later marked it as positive.

In practice, user-submitted corrections likely contain conflicts. A very common conflict happens when multiple users identify the same error, but recommend different solutions. For example, the review snippet "*However, hardware volume control, bilateral speakers, and a more thoughtfully placed power button would have earned the Fire 5 stars from me*" was classified as positive by OpinionBlocks. While four participants changed it to neutral, other three moved it to negative. In such cases, OpinionBlocks currently takes the solution by the largest number of "votes", assuming that the number of "votes" passes the threshold described above. Consequently, the sentiment of this review snippet was changed to "neutral". Note that we do not require a "majority rule" here. Our rationale is that when an error is identified by many, it is better to correct it than to leave it in the system, even when there is no consensus on the solution. Adopting the most suggested solution that passes the threshold seems sensible.

4 User Studies

To validate the effectiveness of OpinionBlocks in meeting our two design goals mentioned in the introduction, we conducted user studies to answer two sets of questions:

1. How well does OpinionBlocks support real-world, opinion analysis tasks?
   a) How well can users find important aspects mentioned in the reviews along with their associated sentiment?
   b) How well can users find evidence behind reviewers' opinions?
   c) How well can users get to the detailed facts and discussions as needed?
2. How practical is it for OpinionBlocks to leverage the crowd to improve its quality?
   a) How accurately can users make amendments to correct system errors?
   b) How willing are users to make such contributions?
   c) How well do the amendments improve the system to benefit new users?

4.1 Study Design

To answer the questions mentioned above, we designed two identical studies and conducted them in sequence under two different experimental conditions. Both studies were used to answer the first set of questions and questions 2 (a-b) by steering the participants to identify and correct system analytic errors. In Study 2, however, the user corrections submitted in the first study were incorporated to answer question 2(c). We compared the user performance between the two studies to assess any improvements (e.g., task time) due to user corrections made in the first study. We used disjoint sets of subjects between the two studies, i.e., a between-subject experiment design, to avoid any learning effect.

Participants
Since OpinionBlocks is designed to help end users, we conducted both studies by recruiting participants from Amazon Mechanical Turk (called turkers from now on). After a pilot, we recruited 50 turkers for each study. Turker qualifications included being located in the United States, having done at least 50 approved Human Intelligence Tasks (HITS) on the site, and having over 98% approval rating for all HITS. Each approved task completion was paid $2.59 US dollars. Measures were taken to ensure that one turker could do the task only once.

Data Set
We used Kindle Fire reviews from Amazon.com as our primary data source. We selected this data set for two reasons: First, it is a large data set that can be used to assess user performance in real-world tasks. Second, it is in a domain that may appeal to a general audience. At the time when we conducted the studies, there were over 18,000 reviews on Kindle Fire, with more reviews added daily, indicating people's strong interest in the product. Overall, OpinionBlocks extracted 3034 aspects and 48,000 review snippets from the 18,000 reviews.

Tasks and Measures
Each turker was first directed to an online survey that contained a set of instructions and questions about the tasks. The survey started with a scenario: "*Suppose you want to buy a tablet. You have just heard about Kindle Fire. You'd like to learn more about it so you can make an informed decision.*" The turker was then given a brief tutorial in

a sequence of annotated screen shots of OpinionBlocks, explaining each interface element and function.

After the tutorial, the turker was given a link to launch the *live* OpinionBlocks tool in a separate browser window/tab. After OpinionBlocks was launched, the turker was then instructed to go to the next page of the survey to answer questions using the tool. There were a total of 27 questions in each survey, including fact-finding questions about the product (e.g., “*Which aspect of the product received the most conflicting reviews*”) and questions about the tool (e.g., “*How would you rate your experience using our tool to explore the reviews*”). A timer was started when the page of the survey containing the fact-finding tasks was loaded. The timer stopped if all the questions on the page were answered and the page was turned to the next one. The timed duration was used as a measure of completion time for fact-finding tasks.

4.2 Results

We received 50 completed surveys for our first study and 51 for the second one. After reviewing each response, we approved all of them. On average, each turker spent 35.5 minutes on our survey.

1(a) How Well Can Users Identify Important Aspects/Sentiments?

Suppose that users are potential customers in the market for a tablet. We designed two related questions to investigate this aspect. First, we asked them a yes/no question on whether they could make an informed decision on the tablet based on their use of OpinionBlocks. This question was to assess the users' overall confidence in their comprehension of important factors and their associated sentiment to influence their buying decisions. 81 out of the 101 turkers confirmed that the information is sufficient for them to make a decision on the product. One user also provided the rationale for his "Yes" answer: "*there are more green bars than orange*".

The second question asked the turkers to find the important aspects of the product. This question was to examine whether a user's understanding of the aspects was consistent with what the system provided. To do so, we counted the number of times that the users' responses contained at least one of the top-three aspects identified by the system: "*screen*", "*app*", and "*book*". 76 out of the 101 turkers' produced correct answers, indicating user-identified main aspects were consistent with that of the system.

In addition to these two questions, we also used a set of questions such as "*Which aspect has received the most conflicting reviews*?" to assess how well users can use OpinonBlocks to identify aspects with distinct characters (e.g., most positive, negative, and controversial). For these questions, for example, 66% of turkers in Study 1 successfully identified "*screen*" as the aspect that received most conflicting reviews, while 72% turkers did so in Study 2. Moreover, the turkers were able to cite both positive and negative sentiments to substantiate their findings (see more below). Considering that there were 3034 aspects extracted from 18,000 reviews, OpinionBlocks demonstrated its effectiveness in helping users identify salient aspects of the product.

1(b) How Well Can People Find Evidence to Substantiate an Opinion?
We designed three questions to ask the turkers about various review details (e.g., "*What products are the main competitors of the Kindle Fire*?"). For all of these fact-finding questions, users were required to excerpt one or two sentences from the reviews to support their answers. Two coders independently read all turkers' responses (3x101=303 responses from two studies) and marked the responses (Yes or No) based on whether the cited sentences correctly supported the answer. Krippendorff's alpha was computed to measure the inter-coder reliability, where alpha = 0.70, suggesting a good level of consistency between the two coders. We then computed the percentage of turkers that correctly found evidence to back up their answers (0.5 was used when the two coders diverged). Out of 303 responses, 274 were correct (90.4%). Clearly, OpinionBlocks was able to help users find specific evidence for opinions.

1(c) How Well Can People Get to Important Details?
As described above, we learned that users were able to cite relevant evidence to back their answers. However, we also wanted to measure how *accurate* their answers were. To do so, two coders independently read each answer to judge whether it was consistent with the answers suggested by the original data. The inter-coder reliability was measured at alpha = 0.82. The percentage of turkers that gave the correct answer was 95.2%, indicating that the majority of users were able to use OpinionBlocks to find desired details of the product when needed.

2(a) How Accurately Can People Make Amendments?
During the studies, each turker was asked to identify and correct at least ten text analytic errors in OpinionBlocks. Each turker was randomly assigned five aspects displayed in the visual summary to perform this task. From Study 1, we collected a total of 659 user-made changes. Many of the changes were made by multiple participants. After removing the duplicates, we obtained 378 distinct amendments. Among them, 47 corrected misclassification of snippets by aspect; 347 corrected misclassification of snippets by sentiment; and 16 corrected both at the same time. After applying our rules for integrating user feedback, 49 unique amendments were incorporated into OpinionBlocks for Study 2.

Two coders examined the 378 unique changes and coded each of them to assess the correctness of the changes. Due to the inherent semantic ambiguities in interpreting the snippets, the initial independent codings had relatively low inter-coder reliability with alpha=0.36 for both aspect and sentiment placement. This low agreement in perception of aspect-based sentiment is also observed by Brody et al. [6]. Meetings were held between the coders to discuss a more consistent way of coding the results. They identified two common cases of ambiguity and built a set of coding rules: (a) the interpretation of sentiment orientation should be anchored around the aspect first then the product. For example, one snippet stated "*after using the fire for a few weeks now, my ipad is gathering dust.*" If this snippet is under aspect "iPad", then it should be classified as negative, but if under "tablet", it then should be positive; (b) if a change makes sense or does not make it wrong, count it as correct. For example, the snippet "*Software Controls – I can see why the lack of external buttons would annoy some but for me it is not a problem*" can be interpreted as positive or neutral.

After applying these coding rules, we achieved good inter-coder reliability, 0.91 for aspect and 0.97 for sentiment respectively. The averages of the two coder's ratings were used in the accuracy calculation. For aspect placement, 22 of the 47 changes were coded as correct (46.8%); while 247 of the 347 sentiment changes were accurate (71.2%). These results suggest that people are more capable of fixing sentiment errors than aspect errors. Since the accuracy rates were not as high as we had hoped, we computed the accuracy rate for the 49 changes incorporated by OpinionBlocks, and found that these changes achieved an accuracy of 88.8%. This demonstrates the effectiveness of our user feedback integration rules (section 3.3), and suggests that OpinionBlocks can be improved by crowd-sourced input over the use of state-of-the-art machine learning techniques alone.

2(b) How Willing Are Users to Make Amendments?

We explicitly asked turkers about their willingness to make changes while using the system. From their answers, most users (95%) are willing to contribute.

We also asked the turkers to explain their main reasons for their answers. The reasons given by people who were willing to contribute fell into several categories:

About 50% of the turkers said that they would like to help improve the quality of the tool for its better use. For example, one said, "*I'd be willing to spare a few seconds to improve a tool that I would gladly use.*" Another commented: "*Those features are key to the tool's use" and "it can make the tool more useful and correct*".

About 15% cited the community and social benefits. The reasons include "*I thought this was a useful feature that made the tool more of a community-use tool rather than just an individual-use tool.*"; and "*I think it will go a long way in making users of this app feel like they're contributing in some way. It may even become a draw of sorts for the app.*"

About another 15% felt simply that it was fun and cool to correct things. They mentioned "*It's fun!*", "*It was interesting to correct the errors, because I found myself trying to figure out why each incorrect snippet had been improperly categorized.*"; "*It's cool that you can edit things.*"; "*I like organizing things. Especially when mis-rated reviews stick out like a sore thumb.*"

The majority of people who expressed their unwillingness to contribute (5% of participants) voiced their main concerns about the potential abuse of the system: "*If this was used by multiple people, it would end up being very abused.*"; "*My only concern here is people messing with the system to improve reviews of their own products or make competitors look bad.*"; and "*it's handy but should be checked by someone".*

Other unwilling participants just did not want to bother, or wanted to get paid*: "I'm not really interested in correcting mistakes.*"; and "*I can't see doing it out of the kindness of my heart. If it were on Mechanical Turk I could see doing it for a small amount of money.*"

Overall our results suggest that it is feasible to leverage the power of the crowd to help improve the system.

2(c) How Much Have User-Amendments Made the System Better?
As discussed earlier, the turkers made many changes, of which 49 most common ones were integrated by OpinionBlocks. The incorporated amendments achieved an accuracy of 89%, thus improving the quality of the visual summary.

To measure the impact of integrating the user edits from Study 1 on user tasks, we compared user performance in both studies. To do so, we performed statistical tests using the sequence number of the studies as the independent variable, and all the performance measures taken in the studies as the dependent variables. We found that the turkers' time for completing fact-finding tasks in Study 2 (M=768.1, SD=338.5) was significantly lower than that of Study 1 (M 916.6 seconds, SD =370.2), t_{98}=2.10, p=0.04. Turkers in two studies performed equally well in term of finding correct facts about the products and relevant evidences. In addition, turkers were equally satisfied with our system in both studies. On a 5-point Likert scale, both obtained a median 4 satisfaction ratings, with 5 being "very satisfied".

Overall, our results showed that it is practical to improve the system by leveraging the crowd to correct system errors, and the resulting improved system lets users perform tasks equally well, but significantly faster. One plausible reason for the improved task completion speed is that in the improved system, there is less misplaced unhelpful information, so users do not need to waste time reading.

5 Discussion

Based on our study results, we discuss the limitations and implications of our work.

5.1 Limitations in Text Analytics

OpinionBlocks has adopted several text mining approaches to analyze opinion text and glean useful insights. It also leverages the power of the crowd to help compensate for system mistakes and improve the overall analysis quality. Nonetheless, due to inherent difficulties in text mining, our current approach presents several limitations.

One difficulty is to decide which review snippets to include and how “big” each snippet should be. Currently, OpinionBlocks includes only text snippets following the sentence structure described in Section 3.2. This means it may miss out many useful sentences that do not conform to such a structure. Currently, each snippet contains only one sentence. This might be undesirable in situations where multiple adjacent sentences are used to express an opinion. The challenge is to balance the accuracy and recall when extracting review snippets, as well as balance the size of a snippet to provide sufficient information without overburdening the text analytic engine or the reader. To make the problem more difficult, striking such a balance may depend on factors particular to the data sets.

Another difficulty we have encountered is to determine which aspects to extract. Currently we extract aspects directly out of subject noun phrases, thus covering multiple categories. Besides aspects, such as “screen” and “app”, which describe the Kindle Fire, we also extracted “iPad” which is a major competitor of the Kindle Fire. Other extracted aspects, such as “wife”, “husband”, and “kid”, describe possible user groups of the Kindle Fire, and the aspect “problem” falls in a generic category applicable to any product. Depending on users and use cases, some might want to see only the aspects pertinent to the product, while others may want to learn more about the aspects of competing products (e.g., aspects of iPad in the context of Kindle). More work is needed to make aspect extraction more meaningful and extensible.

5.2 Common Ground versus Personalization

In the User Studies section, we show that opinions are often ambiguous and that different people may interpret them very differently. Building a “ground-truth” of opinion summary is non-trivial and is unlikely to satisfy every user. Allowing a certain degree of personalization may be desirable in support of individual users’ decision making. Currently, OpinionBlocks allows each user to make amendments that affect only that user’s private session. These changes are propagated more widely when the system administrator decides to do so, and only high-quality changes suggested by many users are adopted. Thus, the standard version of OpinionBlocks that every user starts with is quality controlled, even though users may make amendments to their own private sessions. Complications may arise when merging divergent sets of amendments from many users. This will certainly make a good future research topic.

5.3 Potential System Abuse

A few participants of our user studies expressed their concerns over potential abuse of a system like OpinionBlocks, including trolling or businesses manipulating the information for their own commercial gains through user amendments of opinion summaries. Currently, OpinionBlocks gives the system administrator a great deal of control over which amendments can be integrated into the system. The system administrator can decide to tighten or loosen the threshold for integration or filter out changes from certain users. While further research is required to figure out how to best monitor and moderate user behavior, one approach is to the leverage the crowd themselves. As shown in our user studies, the accuracy of aggregated user amendments is much higher than that of individual changes. This means aggregation of crowd input may help prevent or reduce malicious behavior. Currently our aggregation rules are very simple, future research is needed to develop more sophisticated rules, e.g., incorporating information such as the degree of difficulty of text analytic tasks and user reputation.

5.4 Fostering Healthy Online Review Communities

Our work bears a major implication on the research in online communities. Gilbert and Karahalios [11] pointed out two problems of current review sites: (1) large numbers of reviews are never read and in essence wasted; and (2) "pro" reviewers dominate the community and it's hard to hear the voice of "amateur" reviewers. They call on system designers to nudge community members toward community-wide goals. OpinionBlocks helps address both problems: It summarizes the reviews and helps users understand large collections of reviews. It also fosters a democratic environment for others to contribute. In short, we have taken the first step to create a platform to foster a healthier online community where users can potentially help the system and help one another.

5.5 Value to Text Analytics Research

It is also worth noting that our approach of marrying machine and human intelligence to text analytics produces invaluable assets for text analytics research. First, crowd feedback can be used as an indicator to identify "high-value" areas for users. As shown by our study results, users mostly made corrections to the sentiment classification but only a few on the aspect classification. This suggests that users may be more sensitive to certain types of errors than others. Moreover, user-submitted corrections can be used as a training corpus to help tune analytic algorithms.

6 Conclusion

We have presented OpinionBlocks, a novel visual analytic system that aids users to analyze large sets of opinion text. It is uniquely designed to combine state-of-art NLP technologies with crowdsourcing to aid users in their real-world opinion analysis tasks. It employs multiple NLP technologies to automatically generate a fine-grained, aspect-based visual summary of opinions. As demonstrated by our user studies involving 101 users on Amazon Mechanical Turk, the majority of participants not only were able to use OpinionBlocks to complete real-world opinion analysis tasks, but they also exhibited a surprisingly high degree of altruism and concerns for the well-being of online review communities. As users gain value from the system, they become willing contributors to help correct system analytic errors and improve the system. Moreover, the crowd-assisted system enhancement significantly improved task completion time. Based on these findings, combining visual analytics with crowd-sourced correction is thus shown both feasible and effective.

References

1. Sentiment lexicon, `http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar` (accessed: September 09, 2012)
2. Stanford log-linear part-of-speech tagger, `http://nlp.stanford.edu/software/tagger.shtml` (accessed: September 09, 2012)

3. Amershi, S., Fogarty, J., Weld, D.: Regroup: interactive machine learning for on-demand group creation in social networks. In: CHI, pp. 21–30 (2012)
4. Amershi, S., Lee, B., Kapoor, A., Mahajan, R., Christian, B.: Cuet: human-guided fast and accurate network alarm triage. In: CHI, pp. 157–166 (2011)
5. 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: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, pp. 313–322. ACM, New York (2010)
6. Brody, S., Elhadad, E.: An unsupervised aspect-sentiment model for online reviews. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT 2010 (2010)
7. Carenini, G., Rizoli, L.: A multimedia interface for facilitating comparisons of opinions. In: IUI, pp. 325–334 (2012)
8. Carlier, A., Ravindra, G., Charvillat, V., W., O.: Combining content-based analysis and crowdsourcing to improve user interaction with zoomable video. In: ACM MM, pp. 43–52 (2011)
9. Faridani, S., Bitton, E., Ryokai, K., Goldberg, K.: Opinion space: a scalable tool for browsing online comments. In: CHI 2010, pp. 1175–1184. ACM, New York (2010)
10. Fieller, E., Hartley, H., Pearson, E.: Tests for rank correlation coefficients. I. Biometrika 44(3/4), 470–481 (1957)
11. Gilbert, E., Karahalios, K.: Understanding deja reviewers. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010, pp. 225–228. ACM, New York (2010)
12. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: KDD, pp. 168–177 (2004)
13. Huang, J., Etzioni, O., Zettlemoyer, L., Clark, K., Lee, C.: Revminer: An extractive interface for navigating reviews on a smartphone. In: UIST (2012)
14. Jo, Y., Oh, A.: Aspect and sentiment unification model for online review analysis. In: Proceedings of ACM Conference on Web Search and Data Mining, WSDM 2011 (2011)
15. Lee, Y.E., Benbasat, I.: Interaction design for mobile product recommendation agents: Supporting users' decisions in retail stores. ACM Trans. Comput.-Hum. Interact. 17(4), 17:1–17:32 (2010)
16. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: WWW, pp. 342–351 (2005)
17. Liu, B., Zhang, L.: A Survey of Opinion Mining and Sentiment Analysis. In: Mining Text Data. Kluwer Academic Publishers (2012)
18. Liu, B.: Sentiment analysis and opinion mining. In: Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers (2012)
19. Moghaddam, S., Ester, M.: ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: Proceedings of the Annual ACM SIGIR International Conference on Research and Development in Information Retrieval, SIGIR 2011 (2011)
20. Mukherjee, A., Bing, L.: Aspect Extraction through Semi-Supervised Modeling. In: Proceedings of 50th Anunal Meeting of Association for Computational Linguistics, ACL 2012 (2012)
21. Nelson, P.: Information and consumer behavior. J. of Political Economy 78(2), 311–329 (1970)
22. OpenNLP, `http://opennlp.apache.org`
23. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval (2008)

24. Park, D., Lee, J., Han, J.: The effect of online consumer reviews on consumer purchasing intention: The moderating role of involvement. Intl. J. of E-Commerce 11(4), 125–148 (2007)
25. Patel, K.: Lowering the barrier to applying machine learning. In: Adjunct Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, pp. 355–358. ACM, New York (2010)
26. Rohrdantz, C., Hao, M.C., Dayal, U., Haug, L.-E., Keim, D.A.: Feature-based visual sentiment analysis of text document streams. ACM Trans. Intell. Syst. Technol. 3(2), 26:1–26:25 (2012)
27. Schuler, K.K.: VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon. PhD thesis, University of Pennsylvania (2006)
28. Snow, R., O'Connor, B., Jurafsky, D., Ng, A.: Cheap and fast—but is it good?: evaluating non-expertannotations for natural language task. In: Proc. EMNLP 2008, pp. 254–263 (2008)
29. Thet, T., Na, J., Khoo, C.: Aspect-based sentiment analysis of movie reviews on discussion boards. J. Inf. Sci. 36(6), 823–848 (2010)
30. Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., Qu, H.: Opinionseer: Interactive visualization of hotel customer feedback. IEEE Trans. Vis. Comput. Graph. 16(6), 1109–1118 (2010)
31. Yatani, K., Novati, M., Trusty, A., Truong, K.: Review Spotlight: a user interface for summarizing user-generated reviews using adjective-noun word pairs. In: CHI, pp. 1541–1550 (2011)
32. Zhu, F., Zhang, X.: Impract of online consumer reviews on sales: The moderating role of products and consumer. J. of Marketing 74, 133–148 (2010)

PolemicTweet: Video Annotation and Analysis through Tagged Tweets

Samuel Huron[1,2], Petra Isenberg[1], and Jean Daniel Fekete[1]

[1] INRIA-Team Aviz, Bât 660, Université Paris-Sud,
91405 Orsay Cedex, France
[2] IRI, 4 rue Aubry le Boucher,
75004 Paris, France
{Samuel.Huron,Petra.Isenberg,Jean-Daniel.Fekete}@inria.fr

Abstract. We present PolemicTweet a system with an encompassing, economic, and engaging approach to video tagging and analysis. Annotating and tagging videos manually is a boring and time-consuming process. Yet, in the last couple of years the audiences of events—such as academic conferences—have begun to produce unexploited metadata in the form of micropost activities. With PolemicTweet we explore the use of tagged microposts for both video annotation and browsing aid. PolemicTweet is a system 1) to crowd source conference video tagging with structured sentiment metadata, 2) to engage audiences in a tagging process, and 3) to visualize these annotations for browsing and analyzing a video. We describe the the system and its components as well as the results from a one-year live deployment in 27 different events.

Keywords: Backchannel, Video annotation, Crowdsourcing, Video analysis, Live tagging.

1 Introduction

We present a web-based structured video tagging, browsing, and analysis system called PolemicTweet as well as the results from a long-term evaluation of the system under live deployment. The system is unique in that it supports a wide variety of related activities ranging from video tagging to analysis and provides its different components to a variety of benefits for different kinds of audiences.

PolemicTweet (PT) was developed in response to the need for effortless but rich video tagging for event recordings and their subsequent browsing and analysis. A number of organizations in the world, including ours, regularly organize mid-sized conference-style events, with audiences of 25-300 attendees. Many of these events are recorded using videos for archival purposes or for sharing on the web. For example, a rapidly growing global audience regularly follows the TED-events series [1]. However, with current online video players it is difficult to get additional information on a recorded video. Video meta-data does not generally give good information scent about the events unfolding during the course of a video, such as the

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 135–152, 2013.

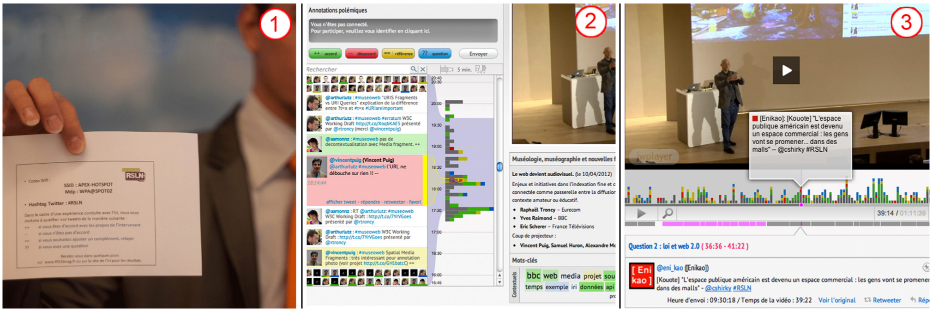


Fig. 1. The three stages of PolemicTweet: 1) printed information on connection & tags, 2) the live Twitter client & structured annotations, 3) video player showing tweet annotations

topic(s) covered, controversies that may have arisen, or reactions of the audience to a speaker shown in the video. This is because videos are complex to summarize, analyze, search and browse. Yet, we argue that providing additional information scent is highly valuable to external viewers. We thus, designed PT to capture structured annotations in a cost-effective manner and provide this information for perusal. The captured meta-data is meant to help both casual browsing and analysis of video content. PT also provides visual summarization of annotations captured during live events and includes additional features for search and browsing.

We specifically contribute a web-based solution to video annotation and analysis that is easy to deploy and use. It is engaging in that the annotation phase offers benefits to participants beyond their involvement in a tagging activity. PT is a well-rounded system of integrated components and has been evaluated in a long-term study involving the capture and annotation of video from 27 different events over one year.

Fig. 2. The Microsoft conference room in Paris where PolemicTweet was deployed; the use of laptops or other wifi-capable devices was prevalent and ensured active participation

2 Motivation

Our design of PT was motivated by the proliferation of microblogging activities at live events associated with our organization: the Institut de Recherche et d'Innovation (IRI) du Centre Pompidou. IRI is a research association focusing on the study of cultural practices enabled by digital technologies. A part of IRI's activities is to organize seminars on subjects such as philosophy, design, and digital cultures. A typical seminar is video recorded. A formal procedure of human annotation and tagging on the video footage was used to follow the event in order to make the videos more searchable and comprehensible. This video annotation task was generally assigned to interns (often archivist students) using a custom software called “Ligne de temps”. This process had two major disadvantages: without having attended the event the job was tedious and time-consuming for the student and costly for the organization.

After this internal annotation process, the video recordings were published online on a web platform displaying the same annotation interface as “Ligne de temps”. In addition to the first set of annotations, the attendees of the event were allowed to add further annotations. However, the participation remained low: the interface for annotation was complex and did not entice the public participation we had wished for. From these past experiences, we wanted to redesign this process with new tools that took advantage of the micropost data produced by the audiences of our events. The resulting system, PolemicTweet, was, thus, motivated by three goals:

1. Engage the audience to tweet for crowdsourced video annotation and tagging,
2. Provide a visual backchannel to incite people to tag their tweets with predefined tags and thus provide structured annotations,
3. Provide easily accessible tools in an encompassing system that span the whole process of annotation to video analysis.

In order to reduce the complexity of the annotation process we decided to make it more engaging and to open the process to a wider audience. As indicated in Goal 1 above, we opted for a crowdsourced solution. Crowdsourcing was a promising solution since the behavior of participants in live events has evolved due to two digital enablers: microblogging and portable devices. The speaker nowadays is no longer the only information channel during an event, a second information channel (backchannel) is now regularly used: the audience's tweets. This information channel is the data source we wanted to leverage for video annotation and analysis in our tool PT.

3 Related Work

To remember, analyze or study an event, video recordings and/or audio tapes are frequently used [2]. For this reason the study and design of video analysis software is a popular topic in information retrieval [3], computer human interaction [4–7], computer supported collaborative work [8–10], and visual analytics [11–14]. In this section, we discuss efforts most closely related to our goals and solutions in regards to our video analysis interface, backchannels for live-events, crowdsourced media annotation and tweet sentiment analysis.

3.1 Crowdsourcing Media Annotation and Motivation

Social network activities such as tweeting can be used for crowdsourcing activities. Shamma and Diakopoulos [15] showed that the social structure and the conversational content of tweets can provide insight into a media event's structure and semantic content: quantitatively through activity peaks and qualitatively through keyword mining. Diakoloulos et al. [16], after having collected tweets during the U.S. State of the Union presidential address in 2010, used them to annotate a video of the event. Based on this assessment and our own experience observing twitter use during conferences, we decided for PT to use Twitter for crowdsourcing video tagging.

Yet, despite the benefits of crowdsourcing for annotation, there are disadvantages to consider. Bernstein et al. [17] named the time-consuming nature of crowdsourcing (due to the recruitment of the crowd and the time to achieve a task). To avoid this problem, the authors designed a recruitment strategy and a retainer model (paying workers to wait and respond quickly when asked). Another identified problem is the motivation of workers. Wu et al. [18] encountered this problem in their video summarization technique via paid crowd workers and expressed the need to further investigate incentive mechanisms. This is one of the issues we addressed in PT.

A large number of pitfalls have also been identified on crowdsourcing task quality [19, 20]. Workers are often untrained and lack task context, particularly in specific domains such as science or the humanities. In PT, however, we do not use a crowdsourcing platform but a casual social network (Twitter) to achieve the task through a community of interest. This way we can expect annotators who are engaged and aware of the context and do not require specific training.

3.2 Backchannels for Live-Events

A digital backchannel can be understood as a thread of information that accompanies a live event. For more than ten years now, digital backchannels, have become a growing research area, in human computer interaction [21–23], computer supported collaborative work [21, 24, 25] and visual analytics [26, 27]. Backchannels have been studied and used in several contexts like conferences [21, 23], classrooms [28], and meetings [22]. A taxonomy of different backchannels was presented by Cogdill et al. [25]. The researchers present five types of backchannels: process-oriented, content-oriented, participation-enabling, tangential and independent backchannels. In PT we use a process-oriented public backchannel, to achieve a live tagging task. A digital backchannel such as ours is useful to provide awareness [28] to local and remote participants [22] of questions, comments [23], shared work, and references[29], and can encourage real-world discussions [22]. McCarthy et al. [21] studied backchannels in a use case scenario similar to ours. They studied digital backchannels in academic conferences using IRC and relate different types of use and problems, like the cost to spread the information for a backchannel connection. McNely [24] suggests that the problem has been partially solved by the increased availability of micropost web services like Twitter, their simplicity of use, and the large adoption rate. Sopan et al. [30] showed that a micropost backchannel during a conference, like in our case, permits to connect local and distant participants. In another academic context, Harry et al. [23] implemented and described the use of a backchannel to leverage participation in an auditorium to allow for the audience to vote on questions and give feedback. In PT we also use an explicit backchannel to crowdsource the process of a real-time distributed user task but—in contrast to the previous work—we use it to produce a document that summarizes this activity as an annotated videotape.

With the prevalence of social networks, portable devices, and wireless network connections, backchannels are no more an emerging social phenomenon but a real trend. When this increased use of social network and portable device meets large-scale events like the Olympics, the Arab spring, or national elections, visual analytic techniques can help to summarize and understand what has happened during the event. Dörk et al. [26], for example, introduced the visual backchannel, a timeline-based visualization that presents an overview of an event through social stream aggregation. Marcus et al. [27] presented an algorithm for event detection in a streams to produce annotation on a timeline-based visualization of the social activity surrounding an event. Diakopoulos et al.[16], synchronized a recorded TV show with related tweets and provided two timelines, one for sentiment analysis over time, the other for volume of tweets.

Similar to these three previous systems ([7, 26, 27]) PT provides a summary visualization for video-recorded events but builds on this idea by using its backchannel as a specific crowdsourcing tool for tagged annotation.

3.3 Tweet Content Analysis

Transcripts, tweets, and other temporal data streams are often used to annotate video. In the case of temporal data composed of text, it is common to use named-entity recognition, sentiment analysis, and natural language processing to recognize and tag events of interest. When we started developing PT, sentiment classification methods on tweets were not as effective as on classical natural language corpora due to the limited length of tweets and the common use of shortened and non-standard English words [31]. Now, the combination of different techniques from machine learning [29], semantic rule based approaches [32], and graph-based optimization [33], have improved sentiment classification of tweets significantly, and can achieve 85.6% accuracy [32]. In this last work, the authors considered three types of sentiments: negative, positive, and neutral. The reported accuracy rate was achieved on an English tweet corpus of the most popular queries on Twitter (Obama, Google, iPad, Lakers, Lady Gaga), and on a clearly targeted subject. We did not take advantage of this approach as it only works for English tweets, it is complex to set up, and has only been tested on a topic-limited tweet corpus. Instead, for PT [34], we needed to find a solution that would work in multiple and mixed languages, was simple to deploy, and would work on a variety of specialized corpora like philosophy, aesthetics, or design.

Diakopoulos and Shamma [15] used another approach for sentiment classification: they used Amazon Mechanical Turk to perform hand-annotated sentiment classification on tweets. Turkers were compensated $0.05 per ten tweets analyzed. Turkers were asked to tag four types of sentiments: negative, positive, mixed, and "other." The corpus of tweets was in English and about politics. Others [16] described the use of machine-learning algorithms to perform the same analysis with lower accuracy and cost but higher speed. Crowdsourcing sentiment classification on tweets with its higher success rate and relatively low overall cost is more and more common. For instance, it is one of the products of CrowdFlower [35] a well-known crowdsourcing platform. However, this approach cannot be used to tag tweets in real-time.

Whether or not to use crowdsourcing as a tool depends on one's annotation goal [36] since there is a speed vs. quality tradeoff to consider. For PT, we wanted to have the best of both worlds: classify tweets with a low cost (both in computer processing and in money) and a high precision rate, even if the recall rate depended on the adoption of our tool. We also wanted to have this classification in real time.

3.4 Video Analysis Support

Mackay's EVA [5], was one of the earliest systems on video annotation, tagging, and analysis. Mackay synchronized records of metadata with her video records, such as the movements of a mouse on a screen. Nowadays recorded metadata is often taken from real-time web social services like chat or microblogging. PT relates to this trend but asks Twitter users to tag their own tweets for structured annotation.

Another solution to aid in the process of video analysis has been to support multiple analysts in parallel. Cockburn and Dale [8], for example, designed and developed CEVA a synchronous collaborative video annotation system that focuses on supporting parallel analysis. Parallel analysis is based on the idea that multiple video analysts (five in their prototype) share their analyses in real-time for distributing the workload. The authors argue that this synchronicity property offers two potential benefits: 1) synergy of group participation and 2) distributing the analysis workload. We were particularly interested in supporting the synchronicity property for the PT tagging phases to similarly take advantage of the synergy of the audience during event recording. The Videolyzer tool [7] also includes an asynchronous formal semantic tagging tool to increase information quality by allowing users to collaboratively organize their comments on the video. Their tool differs from ours in that we did not want to rely on a complex tagging interface to more easily spur user tagging in real-time during the event. Nevertheless, Videolyzer inspired the social use case scenario of PT.

4 System Design

PT is made of four interconnected components: 1) the definition of four tags to annotate tweets and the backchannel interface to read and write tagged tweets during the conference event, 2) a social protocol to set up and run an event live, and 3) a web video player synchronized to a tweet visualization to navigate and replay conference video recordings, 4) a website to provide a fluid browsing between all PT components. We describe the components in that order.

4.1 Backchannel to Crowd Source Tag Annotation

To provide an enriched visualization of video annotations we wanted to collect tweets structured by specific annotation tags. We chose to provide a limited number of tags with a simple syntax and a specific color code (in brackets): agreement with the speaker (++,■), disagreement with the speaker (--,■), questions raised by or proposed to the speaker (??,■), and references (==,■), e.g. quotes or URLs to related content. Our choice of tags was informed by observation of the types of tweets typically used in the conference-style events that we target with PT. For choosing our set of tags, we had to balance expressive power and simplicity. To provide a memorable set of tags, we decided to use the simplest possible tagging format. At the same time, our interface needed to provide enough incentive for using the tags. The design rationales for choosing the tag syntax were:

- *Simplicity of use and memorability*: to facilitate its use and adoption;
- *Brevity:* to cope with the 140 character limitation of Twitter and for fast typing;
- *Client and language neutrality:* to be usable from any twitter client with any kind of text input method, usable in any language;
- *Ease of parsing:* for automatic processing tools;
- *Expressiveness and univocality*: to allow clear statement of intent and for machines to interpret it unambiguously.

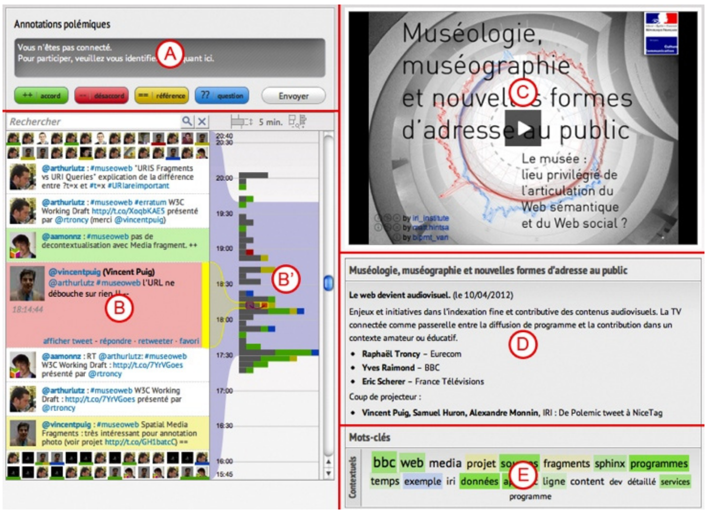


Fig. 3. The PolemicTweet backchannel interface

The four tags were introduced to the audience before the start of an event. Fig. 3 shows the real-time backchannel interface we designed to allow an audience to monitor the tagged tweet activity in real time. It consists of six different components (Fig. 3): A) an enhanced Twitter client with a text field, four buttons to easily add tags to tweets, and a send button, B) a list of tweets relative to the scrollbar's focus, B') a vertical bar chart showing tweet activity over time, colored blocks represent tweets containing a tag using the tag's representative color, C) a video player showing the live video stream, D) abstract and program of the event, E) dynamic tag-cloud extract from the tweets in focus, color-coded according to the polemic tag colors.

For organizers, setting-up this interface is easily done in two steps on the PT web server: 1) cloning the folder containing the default interface, and 2) changing the settings file to specify event title, hash tag, timestamp of the beginning, expected duration, and abstract.

4.2 Engaging the Crowd: Social Protocol for Synchronous Annotation

One very unique feature of PT is a social protocol we developed in order to help event organizers make best use of PT. The main goal of our protocol is to inform the audience about the capture and future use of their tweets and to introduce the four specific tags. The protocol suggested event organizers to follow three phases:
Before the talk: Instructing the crowd. We designed a "connection package" to inform the audience. It consists of a flyer (Fig. 1) given to attendees during the registration process, containing information about the network access policy, the PT tags, the website URL, and the Twitter hashtag for the event. Additionally, we asked organizers to make an announcement prior to the conference.

During the talk: Crowd source sentiment analysis and video annotation. Attendees send tweets with the PT tags, a program records all the tweets relative to the twitter event's hashtag. The visual backchannel website is set up to provide real-time visual feedback of tweets and to give an incentive for participation and easy access to people who do not use Twitter (Fig. 5). Organizers can also make use of this interface to get informed about questions to ask at the end of a talk.

After the talk: Publishing. Organizers synchronize the recorded data with the timestamp of the video streaming server and the tweets' timestamps. Then they publish the video on the web (copying and modifying a configuration file) with the MetadataPlayer, that we discuss in the next section.

Synchronization of the video and tweets could be complex due to the lack of reliable video timestamp. In the case of Diakopoulos and Shamma [15], synchronization was simple because the video was provided by an official TV channel and had a well-known timestamp. In the academic context, an event can be recorded by different means and not be broadcasted live. Most of the time—if a video is not streamed—the video timestamp depends on settings on the recording hardware and is, thus, not reliable. We, thus, deploy several strategies for synchronizing video recording and tweet activity: we video-record a clock before the event, have a special tweet at the beginning of the event, and re-synchronizes it with a custom script. Of course for resynchronization of tweets to a videotape it is important to use the originally captured video footage and not an already cut and edited version.

4.3 Video Analysis Web Video Player

The PT video player (Fig. 4) is designed to play a conference video while showing the tagged and untagged tweets sent with the conference hashtag during the time of video recording. Our design goal was: 1) to provide a visualization that gives an overview of the tags and activity spikes, 2) to design a compact player that can be easily embedded, just like the YouTube player, and 3) to augment the player's video navigation techniques to get more benefit from the annotations.

We designed all the time-related components as graphical horizontal projections on the time axis. All components have the same width and horizontal scale to allow for a vertical alignment between them.

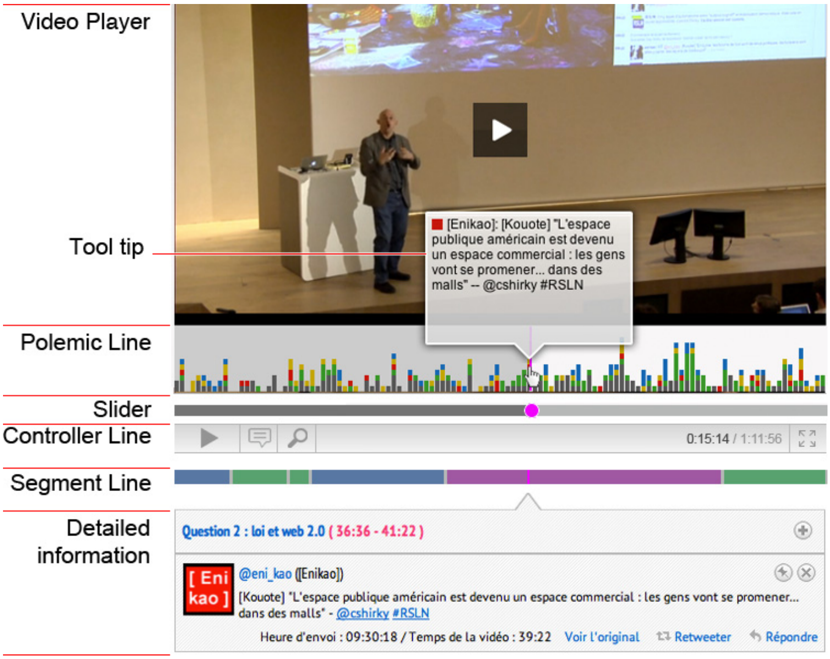


Fig. 4. Each component of the MetadataPlayer used in PT

Navigation on the video is performed using the standard video player components and controls: a controller line with the buttons to play, pause, change the audio level, and a seek bar. Two novel components show context and allow for contextualized interactions. The segment line allows quick access to chapters of the video—such as topics of a presentation or name of presenters for a panel. The data for this line can be manually or automatically extracted from the video. The polemic line visualizes the tweet activity during the event, positioned on the time axis; it provides contextual feedback on the tweet activity, a direct access to video segments highlighted by tweets, and direct access to the tweet contents through hovering. The polemic line has the following characteristics:

- *Polemic line visualization:* The visualization is a bar chart composed of colored tweet squares. Each square represents a tweet colored according to the tweet's tag or gray if the tweet does not contain any tag. The square has a fixed size of 5 pixels to facilitate its selection without taking too much screen real-estate. Depending on the length of the video, each 5px slot represents a certain time range. At each time-slot the recorded tweet squares are sorted top-to-bottom by type (green, red, blue, yellow, and gray) so that the most expressive tweets are at the top of the bars.

- *Interaction:* Moving the mouse over a square shows a tooltip with the tweet's text, author name, and tag color. Clicking on it seeks the video to the time of the tweet, and makes meta-information appear in the details information component (Fig. 5). The position of each square provides information about when the tweet was emitted but also about the contextual activity level at the moment.
- *Search:* A search text-field (Fig. 5.) in the controller bar allows finding text in the tweets. When a search string is entered, the visualization is updated to highlight each represented tweet containing the string. As shown in Fig. 5 this feedback shows the distribution of tweets containing the specified string (in purple).

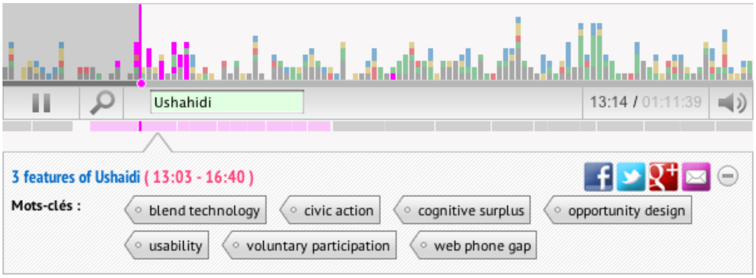


Fig. 5. Search for "ushahidi" in the polemic line". The tweets referring to the section of the talk related to this query are highlighted in pink.

4.4 Website

To provide a fluid browsing experience between all the PT components we embed them inside a public website composed of two main components: a home page presenting the past, present and future events and an archive page to find older events. On the pages containing past events, the MetadataPlayer shows the annotated videos of these events. The backchannel interface is only available during a live event on the "present events" part of the page.

5 Assessing the Impact of Polemic Tweet

In order to better understand how PT supports our goals of cheap, useful, and engaging video annotation, we studied the tool using mixed-methods —both qualitative and quantitative approaches under long-term deployment. We successfully deployed PT in 27 real conferences, seminars, and events of different locations, topics, and sizes of audience over the last year. For these 27 events we recorded 46 tracks (video and tweets), each track being one speaker or one session.

20 events took place at the Pompidou Center in our conference room, 7 events took place at various places in the city of Paris. Nine events had between 25-50 attendees, 13 events 51-100 and five events 101-290. The topics of the events varied greatly, including academic topics such as science, technology, philosophy, and design.

The duration of 24 events was limited to one day while three events took place over several days. The recorded tracks from each event ranged from 30 to 395 minutes. Overall, we received and collected 9,088 tweets from 1,012 unique accounts. Over the last year, the web site (player, backchannel, but also other additional pages like home and archives) received about 157,000 page views from over 15,000 unique visitors with an average visit duration of 3 minutes which is comparable to National video websites with substantially more content and public cover like Francetvod.fr [37] with a 3.2 minute average visit duration during February 2012.

Except for the video recording, sound capture, and the communication support for the tags (flyers and poster boards), the components of PT were deployed on the web, which greatly simplified the process for the organizers. The recorded videos are all published on the open site http://www.polemictweet.com.

5.1 Data Collection

To study the use of PT we gathered three types of data: all tweets using the various conference's hashtags, logs from the usage of the PT MetadaPlayer and the backchannel interface, and two questionnaires sent to various users of PT, either as designers, organizers, speakers, or audience participants.

We used Google Analytics to log the website pages and a custom logging system to collect low-level traces of user interactions on the video player that Google Analytics could not provide. On the backchannel interface, we recorded usage logs only via Google Analytics to give us information about where users came from (direct access, twitter, devices, others) and where the users were physically located (inside or outside the conference room). This last information does not take into account attendees not using the provided free Wifi connection (e.g. using their mobile's own 3G network) and possible errors of domain detection. Despite a few tablet users most connected with personal computers. We also recorded the web client's signature for each tweet to understand where tweets were sent from.

We sent two web surveys, the first consisted mainly of close ended questions sent by tweet to 140 randomly chosen attendees of one of the recorded events; the second questionnaire was sent to understand how PT faired in the organizer's point of view. This questionnaire consisted of open-ended question sent by mail to five members of the event organization team in our institute.

For the first questionnaire we received 47 responses, 27 of which completed all parts of the survey. 70 % (19) of the respondents attended at least one of the conferences. 96 % (26) had already used twitter at conferences, 3 % just for reading. This shows a clear positive feedback towards the goals of PT. For the second questionnaire we received four completed responses.

5.2 Analysis

In this section we report on the analysis of our data according to three main questions: 1) Is the system sufficiently engaging to collect data for crowd sourced video tagging? 2) Does the visual backchannel incite people to tag their tweets and thus provide annotation structure? 3) Is the video web player with annotations useful?

Is the System Sufficiently Engaging?

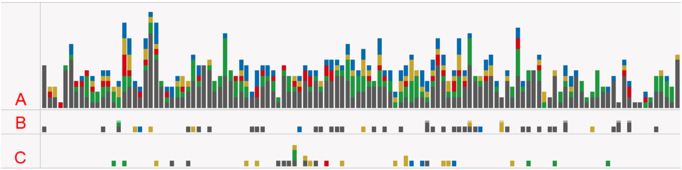


Fig. 6. Polemic lines of A) “Les Stratégies Virtuelles des Musé: l'Heure de Vérité”, B) “Mashup Remix”, C) “Du vinyl au numérique,” a 2h event with different density of annotation, number of peaks, and tagging activity.

Through our live deployment of PT, we found that engagement in the crowdsourced video tagging was highly diverse. We recorded from 0.20 to 7.94 tweets per minute (mean at 1.47, median at 0.89), with 0.12 to 2.13 tags per minute (mean 0.47, median 0.32). Averaged across all events, tweets were tagged using our syntax 40% of the time. Considering tagging a tweet as an additional cost, we conclude that the system provided enough tagging incentive on average.

Otherwise, we found that engagement in the annotation process was highly varied and depended on several factors including the type of audience, the content of the event, the number of twitter users, and whether the PT protocol had been respected or not. We discuss these in more detail now:

Type of Audience. Our audiences were varied in terms of size, culture, equipment, incentive to achieve a task, and distribution of local and distant attendees. All of these factors played a role in the participation and engagement of attendees in the tweeting and tagging task. In our collected data, we had a group of events in which the number of recorded twitter accounts in comparison to the number of attendees was very high (64%–187% on 7 events with 825 tweets on average, (e.g. Fig. 6. A). Groups with values higher than 100% represent those in which more twitter accounts were recorded than participants in the events. Here remote attendees also used the conference hashtag to participate in the backchannel. The recorded events can be categorized according to two participation types: medium participation (20%–51% on twelve events with 209 tweets on average) and low participation (1%–18%, on seven events with 61 tweets on average, Fig. 6. C).

Type of Speaker and Twitter Reaction. Some attendees reported through our informal interview that writing a tweet and tagging it took more attention in-situ than when following the event online. We had originally hypothesized that a captivating speaker would probably lead to less Twitter activity, but from our observations organizing and attending our events we could not confirm this hypothesis. On some events we observed a high level of tweet activity during a captivating talk and others with a captivating speaker had hardly any tweet activity. On the other hand we also noticed that in cases of less captivating talks, a peak of tweet activity arose because the audience was no longer focused on the speaker. Overall, regardless of the quality of the speaker the audiences of our events were polite and only used negative tags 14% of the time.

Type of Content. We found that the type of content of an event and its audience could be highly related factors when it comes to annotation activity. For Instance, one of the events' topic was the websites of the Pompidou Center and the Tate Gallery. In accordance to the topic, the audience consisted of people active on the web and social networks. During this event we saw a very high participation (Fig. 6. A) in terms of tweets and tags: 6.06 tweets per minute and 2.13 tags per minute. On the other hand, we were also surprised to see low participation for certain events where the content would have suggested high Twitter usage. For example the "Muséologie 2.0" had the topic: museum digitization, and preservation policies. Here, the level of participation was extremely low: 0.20 tweets per minute and 0.12 tags per minute.

Observance of the PT Protocol. We found that the protocol was generally well applied by the event organizers. Yet, in some cases we observed that deviance from the protocol could impact user activity. In particular we found problems related to hashtag selection, instruction diffusion, and connectivity.For the event "L'Open Data, et nous, et nous, et nous ?", organizers chose to use an existing hashtag to boost interaction between audience and remote Twitter users. They used #opendata, a common hashtag, which had been in use daily before the event by this community and others. The effect was that the tweet number per minute exploded to 7.94 but the number of tags per minute stayed really low at 0.88 as the tags were not known outside of the event. The backchannel was crowded by other tweets relative to the subject but not to the conference. This was a source of frustration for some attendees and made our system less useful in real-time and as a video annotation tool. Choosing a specific hashthag is important if video annotation is a goal, even if the event should stay connected to existing Twitter communities. Another issue regarding the observance of the PT protocol was that the instruction distribution to the audience was always different. Sometimes flyers were put on a seat, sometimes handed out with oral instructions, and sometimes placed somewhere on a table at the entrance. Despite these differences, most of the time attendees found the information and tweeted. Yet on some events we observed that flyers were not provided and just an announcement of instructions was made. This dramatically impacted the audience participation. For example this resulted in only 0.27 tweets per minute for the event "Du vinyl au numérique" (Fig. 6. C). The last and most obvious factor that highly impacted the audience participation was network connectivity. For instance, in the event "Mashup and Remix" (Fig. 6. B) mobile networks (3G and others) were hidden because the conference room was usually a cinema. A wifi network was provided but connectivity information was just given orally and not on the flyer. Thus, despite an audience of 70 people, only five sent any tweets (0.85 per minute).

Does the Visual Backchannel Incite People to Add Tags?

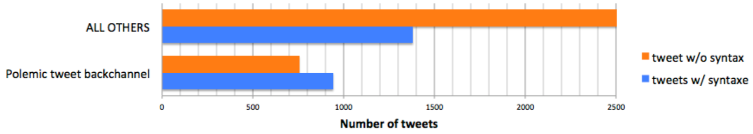


Fig. 7. Tweet with and without PT tags split by backchannel vs. other clients

Of the 27 responses to our first questionnaire, 89 % (24) reported to remember PT tags from attending the event. While our general goal in making the tags known was reached, we were interested to learn more about people's tagging practices and in particular in regards to our backchannel interface.

We observed that tweets sent by the PT backchannel interface had a higher rate of tag adoption (55% of tweets were tagged) than for the top ten other clients used (20% of tweet were tagged) (Fig. 7.). A Welch's t-test was conducted to compare percentage of tagged tweets emitted through the backchannel and by all others clients. There was a significant difference in the score for the backchannel (M=70%, SD=26%) compared to all other clients (M=24%, SD=11): $t(23)=6.74$, $p<.001$. Attendees sent significantly more tagged tweets through the backchannel than through all other clients, suggesting that the backchannel interface did engage people to participate in tagged tweeting. Remote attendees spent more time on the backchannel interface than local attendees. Of those attendees who spent more than 5 minutes on the interface, 62% were remotely connected through outside the conference wifi. From informal feedback from attendees we hypothesize that local attendees did not need some backchannel interface features like video streaming and preferred their own Twitter client but that these features were useful for remote attendees.

We observed that, on average, attendees who used our tags on more than one tweet per event tweeted significantly more than those who did not tweet using our tags (average of 5.3 tweets per person without tags, 15.5 with tags, Welch's t-test, $p < .01$). This shows a strong correlation between tweeting activity and their use of the tags. Although we cannot conclude on causality at this point, the two alternative explanations are either that 1) attendees using the syntax were tweeting more — meaning they were more engaged, or that 2) attendees who tweeted frequently had no difficulty adopting our tags. Informal feedback leads us to think that both situations happened.

What Is the Impact of Presenting Tweets and Tags on a Video Web Player?

To measure if the tweet visualization component was useful, we logged user interaction on the video player described in Section 4. To identify impact on the user activity on six events, we conducted a Welch's t-test to compare if there is a correlation between where the video was played and the density of tweets present at these positions. According to these measures, viewers seemed to specifically seek out spikes of Twitter activity on five of the six videos: three trails followed the twitter activity significantly ($p < .01$), two showed trends ($.01 < p < .05$) and one did not follow ($p > .1$). All trails that were correlated to the tweets had on average more than one tweet by bar in the chart. According to the attendees survey sent after the end of the event 40 % (11) reported to have used the Polemic Line visualization. 90 % (10) of these respondents found that the interface provided the following useful information:

- Item summary of twitter activity: 90 % (10).
- To view a part of video with:
  - Many tweets: 81 % (9)
  - Positive opinions (++): 36 % (4)
  - Negative opinions (--): 54 % (6)
  - Questions (??): 45 % (5)
  - References (==): 36 % (4)

81 % (9) respondents agreed or strongly agreed that the visualization helped them to browse the video and 100 % (11) would recommend PT to a friend.

We received additional feedback from open questions in our surveys and report the most interesting here. One particularly prominent comment was the notion of hotspot and points of interest, also resonating in the answers above. Several comments were similar to this one: "PT is useful to browse videos faster and find a hotspot (polemic, debate, synthesis, and minutes)". Another frequent comment related to video overview. Five users wrote PT was useful for "viewing the mood (of the audience)", and that this overview helped to "construct an opinion", and to put the talk in perspective to the audience's opinions and references". Others commented on the usefulness of the player after an event to "re-find some information and share it with people who could not attend." We also received a comment from one of the organizers that the player was useful to "prepare the chaptering, for taking notes of a speaker change, and references (to others resources)." On the same point, probably one of the most enthusiastic organizers reported that during the event he used the system for note taking: "it replaced my pencil and paper for note taking."

6 Discussion

Evaluating PT as a deployed system over a long period of time was challenging because we had little control over our audience and settings. Yet, our analysis and feedback of PT has been predominantly positive on both usability and usefulness as shown in the results from questionnaires and informal feedback.

According to our analysis, the three main goals of PT were achieved: 1) the system engaged the audience to provide data for crowdsourced video tagging, 2) the visual backchannel incited people to tag their tweets and thus provided annotation structure, and 3) the media player augmented with the annotated tweet visualization was considered a significant improvement over traditional video players.

We found that PT was successful beyond its originally intended purpose of video tagging. In particular remote participants were able to get real-time visual feedback about audience sentiment and could be involved in the event as a commentator and annotator and not just a passive listener. The PT environment and backchannel interface provide a friendly tool for augmenting the public debate around conferences; this is something we consider valuable for improving the communication between important societal issues such as citizens and science or politics in the digital era.

Beyond that, the PT outcomes indicate that we can crowdsource an annotation task using social networks (and not only with dedicated crowdsourcing platforms) depending on the content and community. Last, the use of the backchannel interface can aid by providing an incentive to achieve a task, like tagging during a collaborative synchronous activity. In summary we found a considerable amount of excitement of participants and event organizers. PolemicTweet can be considered as a first step into exploring real-time information visualization of crowdsourced tasks. Our design is applicable to a variety of different scenarios and events such as classroom presentations, synchronous web seminars, public debates, and social TV or even to popular video websites to leverage the crowd for annotating the videos to provide a richer user experience. Despite this, some of our choices should be re-assessed:

- *Possible optimization.* We collected a large number of annotations through tweets but not all were tagged. With our pre-defined tags we could achieve a better precision than sentiment analysis algorithms, but for some events had only low recall. More work is needed to research incentives for the audience to tag tweets.
- *Shared attention.* Asking the audience to tweet during an event and moreover to tag their tweets comes with challenges. First, it requires additional work to remember the tags and add them. The danger is that attendees may be losing focus of their main activity (listening to the conference) and even of their secondary activity such as writing a tweet. Can we further simplify the interface to spare some attention?
- *Tags Property.* We received a lot of feedback from Twitter users on our tags. Some of them criticized their restrictiveness; others felt that tags should be based on existing usage for expressing sentiment. We have several reasons not to rely on existing tags or practices: tags like #fail, #happy, as well as Smileys, are not universal, are ambiguous and are longer than our syntax.
- *Scalability, User interface design and system.* PT comes with some scalability problems. Due to the homothetic representation of tweets, the interface is limited by temporal density. Our design works well for events in which Twitter activity is between 0.5-5 tweets per minute. A graphical scalability problem existed with two of our recorded events: the standard representation took too much screen real estate. To solve this problem, we used a classical aggregated stacked area chart.

7 Conclusion

In this paper we present PolemicTweet a system and protocol that allows reducing video annotation complexity and cost for events of approx. 25-300 attendees, such as academic conferences. We crowdsourced the annotation process through Twitter and made the annotation process more engaging and open for a wider audience. The system's initial goal was to provide an open and engaging method for event organizers to tag and annotate the videos they recorded during their events, but also to supply an accessible tool to play the video over the web and show the annotations.

PolemicTweet is a deployed system, used regularly now at least once a month. This system was deployed for more than one year and co-evolved during several cycles with its users (audience and organizers). This is the reason for its success as well as for the difficulty of evaluating it according to its design goals: they have evolved.

We assessed the success of PolemicTweet and the factors influencing adoption, participation, and use in a long-term deployment. We recorded a wide range of different events and despite the differences in audience, topic content, and event location, the system showed its robustness and effectiveness: 1) to produce useful structured annotation and tagging in most cases, 2) to provide a method and tool to engage the audience in a live tagging activity, 3) to support web-based video browsing activity while providing useful landmarks with a simple yet powerful navigation tool.

References

1. TED: Ideas worth spreading, `http://www.ted.com/` (retrieved January 01, 2013)
2. Allen, C.: The use of video in organizational studies. SIGCHI Bull. 21, 115–117 (1989)
3. Hauptmann, A.G.: Lessons for the future from a decade of informedia video analysis research. In: Leow, W.-K., Lew, M., Chua, T.-S., Ma, W.-Y., Chaisorn, L., Bakker, E.M. (eds.) CIVR 2005. LNCS, vol. 3568, pp. 1–10. Springer, Heidelberg (2005)
4. Roschelle, J., Goldman, S.: VideoNoter: A productivity tool for video data analysis. Journal of Behavior Research Methods 23, 219–224 (1991)
5. Mackay, W.E.: EVA: An experimental video annotator for symbolic analysis of video data. SIGCHI Bull. 21, 68–71 (1989)
6. Abowd, G.D., Gauger, M., Lachenmann, A.: The Family Video Archive: An annotation and browsing environment for home movies. In: Proc. Workshop on Multimedia Information Retrieval, pp. 1–8. ACM (2003)
7. Diakopoulos, N., Goldenberg, S., Essa, I.: Videolyzer: Quality analysis of online informational video for bloggers and journalists. In: Proc. CHI, pp. 799–808. ACM (2009)
8. Cockburn, A., Dale, T.: CEVA: A tool for collaborative video analysis. In: Proc. GROUP, pp. 47–55. ACM (1997)
9. Schroeter, R., Hunter, J., Kosovic, D.: Vannotea: A collaborative video indexing, annotation and discussion system for broadband networks. In: Proc. of. K-CAP, pp. 1–8. ACM (2011)
10. Engström, A., Perry, M., Juhlin, O.: Amateur vision and recreational orientation: Creating live video together. In: Proc. CSCW, pp. 651–660. ACM (2012)
11. Daniel, G., Chen, M.: Video visualization. In: Proc. VIS, pp. 409–416. IEEE (2003)
12. Ruppert, T., Kohlhammer, J.: A radial visualization tool for depicting hierarchically structured video content. In: Proc. VAST, pp. 251–252. IEEE (2010)
13. Tonomura, Y., Akutsu, A., Otsuji, K., Sadakata, T.: VideoMAP and VideoSpaceIcon: tools for anatomizing video content. In: Proc. INTERACT & CHI, pp. 131–136. ACM (1993)
14. Parry, M.L., Legg, P.A., Chung, D.H.S., Griffiths, I.W., Chen, M.: Hierarchical event selection for video storyboards with a case study on snooker video visualization. IEEE TVCG 17, 1747–1756 (2011)
15. Diakopoulos, N.A., Shamma, D.A.: Characterizing debate performance via aggregated twitter sentiment. In: Proc. of CHI, pp. 1195–1198. ACM (2010)
16. Diakopoulos, N., Naaman, M., Kivran-Swaine, F.: Diamonds in the rough: Social media visual analytics for journalistic inquiry. In: Proc. of Visual Analytics Science and Technology (VAST), pp. 115–122. IEEE (2010)

17. Bernstein, M.S., Brandt, J., Miller, R.C., Karger, D.R.: Crowds in two seconds: Enabling realtime crowd-powered interfaces. In: Proc. UIST, pp. 33–42. ACM (2011)
18. Wu, S.-Y.Y., Thawonmas, R., Chen, K.-T.T.: Video summarization via crowdsourcing. In: Extended Abstracts CHI, pp. 1531–1536. ACM (2011)
19. Snow, R., Connor, B.O., Jurafsky, D., Ng, A.Y., Labs, D., St, C.: Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. In: Proc. EMNLP, pp. 254–263. ACL (2008)
20. Ipeirotis, P.G., Provost, F., Wang, J.: Quality management on Amazon Mechanical Turk. In: Proc. SIGKDD Workshop on Human Computation, p. 1403. ACM (2010)
21. McCarthy, J.F., Boyd, D., Churchill, E.F., Griswold, W.G., Lawley, E., Zaner, M.: Digital backchannels in shared physical spaces: attention, intention and contention. In: Proc. CSCW, pp. 550–553. ACM (2004)
22. Rekimoto, J., Ayatsuka, Y., Uoi, H., Arai, T.: Adding another communication channel to reality: An experience with a chat-augmented conference. In: Proc. CHI. ACM (1998)
23. Harry, D., Green, J., Donath, J.: Backchan.nl: integrating backchannels in physical space. In: Proc. of CHI, pp. 1361–1370. ACM (2009)
24. McNely, B.: Backchannel persistence and collaborative meaning-making. In: Proc. SIGDOC, pp. 297–304. ACM (2009)
25. Cogdill, S., Fanderclai, T.L., Kilborn, J., Williams, M.G.: Backchannel: Whispering in digital conversation. In: Proc. HICSS, p. 8 (2001)
26. Dörk, M., Gruen, D., Williamson, C., Carpendale, S.: A Visual Backchannel for Large-Scale Events. IEEE TVCG 16, 1129–1138 (2010)
27. 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, pp. 227–236. ACM (2011)
28. Bry, F., Gehlen-Baum, V., Pohl, A.: Promoting awareness and participation in large class lectures: The digital backchannel backstage. In: Proc. IADIS, pp. 27–34. Emerald Group Publishing Limited (2011)
29. McCarthy, J.F., Boyd, D.: Digital backchannels in shared physical spaces: experiences at an academic conference. In: Extended Abstracts CHI, pp. 1641–1644. ACM (2005)
30. Sopan, A., Rey, P., Butler, B., Shneiderman, B.: Monitoring Academic Conferences: Real-time Visualization and Retrospective Analysis of Backchannel Conversations. In: Proc. SOCINFO (2012)
31. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Journal of Foundations and Trends in Information Retrieval - FTIR, 135 (2008)
32. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proc. ACL HLT, pp. 151–160 (2011)
33. Speriosu, M., Sudan, N., Upadhyay, S., Baldridge, J.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proc. HLT, pp. 53–63. ACM (2011)
34. Huron, S., Fekete, J.-D.: User-Supplied Sentiments in Tweets. In: 2nd Workshop on Interactive Visual Text Analysis. IEEE (2012)
35. Crowdflower: Senti crowdflower, `https://senti.crowdflower.com/senti` (retrieved January 01, 2013)
36. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: Proc. CHI, pp. 1403–1412. ACM (2011)
37. Médiamétrie-eStat: La fréquentation des sites Internet français Février (2012), `http://www.mediametrie-estat.com/la-frequentation-des-sites-internet-francais-en-fevrier-2012/` (retrieved January 01, 2013)

Eyes Only: Navigating Hypertext with Gaze

Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber

Department of Computer Science, University of Auckland,
38 Princes Street, Auckland 1010, New Zealand
{moiz,christof,gerald}@cs.auckland.ac.nz

Abstract. Eye gaze tracking is an obvious candidate for a future input device, perhaps even for everyday computing. The hard problems with gaze-controlled interfaces are inaccuracy and inadvertent clicking. We attempt to mitigate these problems in the context of a gaze-controlled web browser. Four click alternatives (Dwell, Single Confirm, Multiple Confirm and Radial Confirm) were implemented along with a fifth mouse-controlled version for comparison. Two alternatives make use of additional buttons that confirm a selection made previously by dwell, hence improving accuracy. Our results indicate that the Multiple Confirm alternative performed best among the gaze-based alternatives; it makes use of multiple confirmation buttons when letting the user choose between different options. When compared to the mouse, the clicking times were worse but the accuracy was indistinguishable. User feedback also indicates that, although mouse was considered best, Multiple Confirm was not perceived as slow and generated excitement. This indicates that the Multiple Confirm click alternative has potential as an interaction method for gaze interfaces.

Keywords: Eye gaze tracking, dwell, navigation.

1 Introduction

An eye gaze tracker makes it possible to monitor the user's point of gaze, i.e. where the user is looking on the screen. It is considered a promising component of future natural user interfaces. A gaze tracker as an input device is easy to learn and use [1], and it is the fastest pointing device in some metrics [2]. Also, users necessarily look at the task-relevant objects [3], [4], [5] regardless of the input device used, so intuitively a synergy can be expected if gaze is used as an input device.

Pointing with gaze tracking is relatively straightforward, but using it for performing actions (or clicks) is challenging. Inaccuracy and inadvertent clicking are the main drawbacks of using a gaze tracker as an everyday input device. The inaccuracy can be caused by gaze tracker error, or it can stem from fundamental limitations such as involuntary eye movements (including jitter and drifts) and the lack of pixel-pointing precision [6]. Inadvertent clicking, also known as the *Midas Touch* problem, occurs because eyes are sensory (input) organs and this creates a mismatch when using them as a means of communicating intention to the computer.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 153–169, 2013.

Techniques devised to perform actions (or mimic mouse clicks) are called *click alternatives*. There are many such alternatives being researched for encoding an action with gaze. The most obvious and natural one is called *dwell* or fixation, which triggers an action when the gaze dwells on (i.e. fixates) the same area for a defined time interval. There is a reported preference for simple and natural gaze interaction techniques, requiring minimal deliberate eye movements [7]. Conscious gazing or blinking as a control mechanism is unnatural and tiring. Consequently, complicated selection techniques are only feasible for limited accessibility uses. Since we consider gaze interaction for all users, the click alternatives being investigated in this study will be based on dwell. The users will not be required to learn or perform anything other than just looking at particular user interface components.

The suitability of gaze tracking as an alternative to the mouse may depend on the application domain. Therefore we decided to narrow the investigation down to a limited but highly relevant use case: gaze tracking for navigating hypertext, i.e. clicking text hyperlinks (or hypertext). We used Wikipedia for our study as it is one of the most visited websites, and it is close to the original vision of hypertext. We consider it therefore representative for hypertext navigation. The specific question being investigated is how inaccuracies and inadvertent clicks can be mitigated when navigating hypertext. The investigated gaze tracking techniques were also compared with the mouse.

Four click alternatives, referred to as Dwell, Single Confirm, Multiple Confirm and Radial Confirm, were developed. All of them use a different way of clicking with gaze, but all are based on dwell, as explained before. After a pilot study, two of the click alternatives (Dwell and Radial Confirm) were discarded from further experimentation as they did not seem to be usable enough. The remaining alternatives were fine-tuned and a full experiment was performed.

The experiment yielded the following main finding: Multiple Confirm, although slower than the mouse, is a feasible click alternative using the gaze and should be explored further. We conclude that eye gaze tracking has the potential to be used as an efficient input method, as long as certain design parameters are considered.

Section 2 summarizes related research. Section 3 describes the experiment design, detailing the click alternatives and tasks performed. Section 4 provides an overview of the findings of the pilot study and the changes made for the full experiment. The results are presented in Section 5, followed by a discussion of the results (Section 6) and limitations (Section 7). The paper ends with a conclusion, which summarizes the findings and identifies further research in this area.

2 Related Work

The first reported study of how the gaze behaves in a reading task dates back to 1878, and was probably the first to analyze the gaze as a combination of saccades and fixations [8]. Considerable effort has been invested in finding efficient click alternatives. The most straightforward one is considered to be dwelling or fixating on a clickable area. Formally, this is described by a two-state machine, as shown in Fig. 1. This click alternative primarily suffers from inadvertent clicking, a problem that can be mitigated if it is possible to place the content or labels outside the clickable areas [9].

Blinks, winks and different muscle sensors have also been experimented with, to minimize inadvertent clicking, especially for limited-accessibility applications [10]. Other alternatives are based on recognizing eye gestures [11] or anti-saccades [12] to indicate a click, or using a physical button for clicking [2]. Different techniques for selecting menu options with gaze have also been evaluated (e.g. [13], [14]).

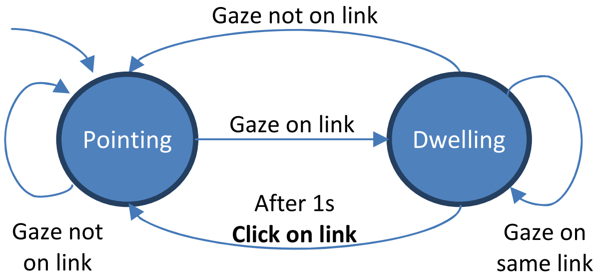


Fig. 1. State machine for the Dwell click alternative

Research has also focused on compensating for the inaccuracy and imprecision of gaze tracking. MAGIC is one such example in which the user relies on explicit commands using other input devices, e.g. the mouse, while also benefiting from the speed of pointing with gaze [15]. This approach moves the mouse pointer quickly to the area being gazed at, but relies on the mouse for finer adjustments and clicking. Different zooming [16], [17] and fish-eye methodologies [18] have been evaluated for the same purpose. The EyePoint [19] is another possible solution to cater for lack of accuracy and precision in gaze tracking. This solution involves magnification (of the area being gazed at) on the press of a keyboard button. In the magnified view, the key can be released while dwelling at the object of interest, resulting in an action on that object (e.g. left or right click). However, these techniques either block or distort the screen content, and loss of contextual information can be inconvenient and problematic, especially for tasks involving visual search [20].

Some recent work has been carried out on gaze-aware or gaze-controlled web browsers. The Text 2.0 Framework [21] allows setting and using gaze handlers the same way mouse and keyboard event handlers are used in HTML and JavaScript. The primary objective of this framework is to detect and assist in comprehension difficulties while reading. The IntelliGaze Desktop 2.0 [22] extension allows web browsing with the gaze by magnifying the area being gazed at if there are multiple clickable options (hyperlinks, buttons or fields) in the gaze area. Dwell is used to click on a link or field in a rectangular magnified view, which always appears in the center of the window. A 'close' button accompanies the magnified view, which can be gazed at to close the magnified view if no click was intended. More controls are dynamically provided in the right margin of the window.

Our aim is to extend this work by evaluating if a more natural click alternative is feasible that supports everyday computing for a general population.

3 Experiment Design

3.1 Variables

The experiment was performed using a within-subjects design with “click alternative” as the independent variable, using a nominal scale. The dependent variables measured are “time to click,” “number of incorrect clicks” and “number of failed clicks.” Furthermore, user satisfaction was measured with a questionnaire.

3.2 Click Alternatives Design

In the following we describe each of the investigated click alternatives.

Dwell. Fixating or dwelling on any hyperlink for one second would click that particular link. Visual feedback, using a darkening outline around the hyperlink being gazed at, is provided to indicate the time progression. A state diagram for this click alternative is shown in Fig. 1.

Single Confirm. When dwelling on a hyperlink, a confirm button is activated in the right margin. The user has to look at this button to click the hyperlink. A screenshot of the implementation is shown in Fig. 2 and the state machine in Fig. 3. This design tries to control inadvertent clicking by providing a confirmation step. The confirm button has a darkening border to show time progression during dwell.

The initial design had the button appearing right next to the line with the hypertext being looked at, in order to minimize gaze travel. A light grey line extended from the button to the edge of the hyperlink to disambiguate among multiple links present in the same line. This was changed after the pilot, as outlined in Section 4.

Fig. 2. Single Confirm: User navigating to “UNESCO World Heritage Site”

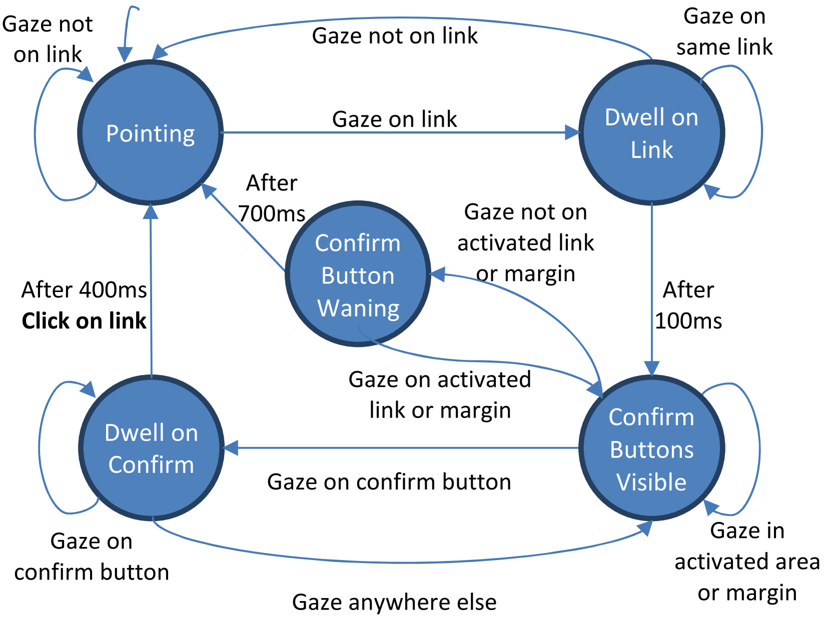


Fig. 3. State machine for Single Confirm click alternative

Multiple Confirm. On dwelling, the confirm buttons of all hyperlinks, in a specified radius around the reported gaze, were activated in the margin. A screenshot of the implementation is shown in Fig. 4 and the state machine in Fig. 5. While avoiding inadvertent clicks, we also expected this alternative to compensate for the inaccuracy

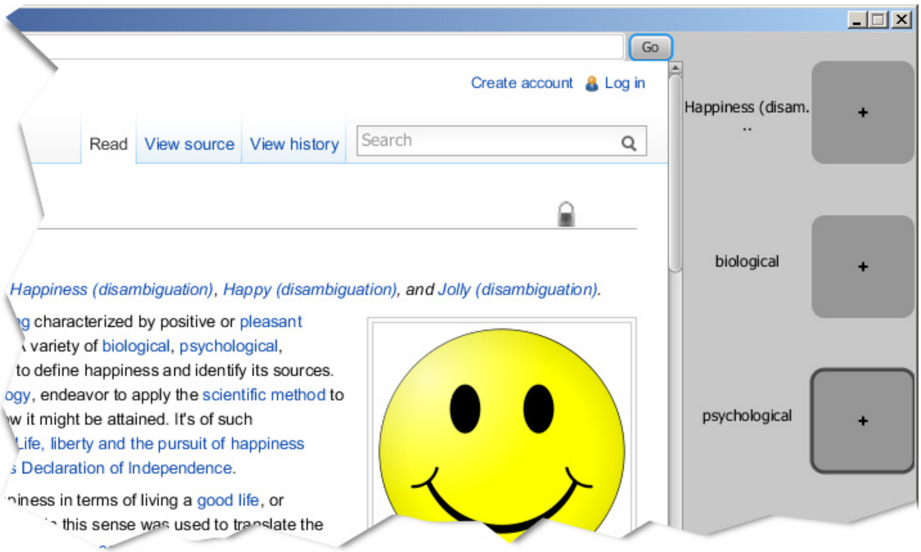


Fig. 4. Multiple Confirm: User navigating to “psychological”

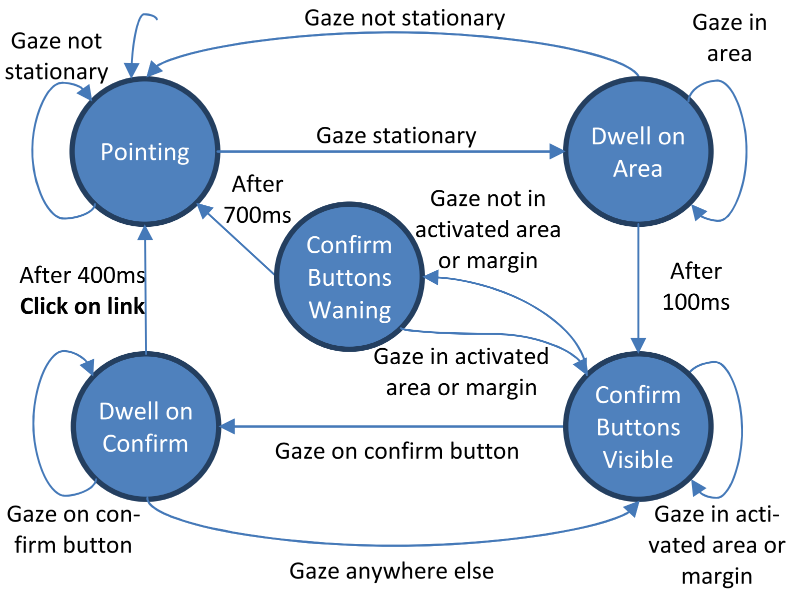


Fig. 5. State machine for Multiple Confirm click alternative

and imprecision of gaze trackers, similar to magnification, but in a more natural way. Some changes were made to the user interface after the pilot (outlined in Section 4).

Radial Confirm. This click alternative is based on the same idea as the Multiple Confirm alternative (and therefore has the same state machine). Additionally, it was designed to minimize the gaze travel from the hypertext to the confirm buttons. A screenshot of the implementation is shown in Fig. 6. Instead of appearing in the right margin, the confirm buttons are drawn around the reported gaze point on the web page. Each confirm button is placed in the direction of the corresponding hyperlink

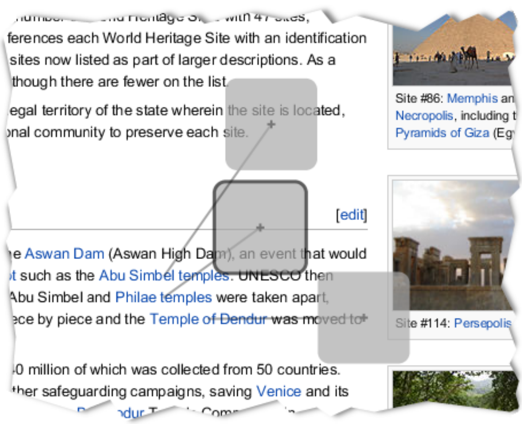


Fig. 6. Radial Confirm: User navigating to “Philae temples”

from the gaze point. In case the gaze point is near an edge of the browser and drawing a confirm button is not possible in one direction, the button gets drawn in a different direction where space is available. The confirm buttons in this click alternative are connected to the respective hypertext with light grey lines for disambiguation.

Mouse. Standard mouse interaction was used as a control condition, while performing the tasks in the same prototype web browser.

3.3 Setup

The web browser prototypes were developed in Java. The stereo infrared cameras of a non-invasive remote gaze tracker were mounted below a 15 inch LCD screen running at a resolution of 1280 x 1024 pixels. The gaze direction was determined by the gaze tracker software using the pupil-center-corneal-reflection (PCCR) method. A fully adjustable chair with headrest was used to help maintain the participant's head position, as the maximum freedom of head movement allowed by this gaze tracker was only 2.7 cubic inches. The LCD screen and the gaze tracker cameras were mounted on a movable arm (complete setup shown in Fig. 7). A travel pillow was also used by the participants to help keep their head rested in one position without getting tired. The lab was illuminated with fluorescent lights, and sunlight from windows was blocked during the experiments.

Fig. 7. Gaze tracking setup

The participants were first given an overview of the equipment and the experimental tasks. After responding to the demographics questionnaire, participants were comfortably seated and the chair's height and distance from the gaze tracker were adjusted. The LCD screen and cameras were then adjusted on the movable arm, if needed. The gaze tracker was calibrated, taking 15 to 20 seconds, before starting the tasks for each gaze click alternative.

3.4 Task Description

The participants performed the same two navigation tasks using each click alternative (as shown in Fig. 8). Before starting a task, the contents section of the involved Wikipedia pages was collapsed using its ‘hide’ option and any banner appearing on the top (e.g. for fund raising) was also closed, to maximize the space used for the actual page content. Since Wikipedia content is being updated continuously, specific versions of the pages were used to ensure uniformity. Participants were allowed to familiarize themselves with a gaze click alternative for a few minutes before using it in the tasks.

For each navigation task, a start page was shown and the participants were told to click four hypertext links one after the other. The first navigation task involved clicking four hyperlinks with comparatively few links in the vicinity, while the second navigation task had a higher hyperlink density in the involved pages. The same two navigation tasks were used for training before performing the actual trials.

Participants were allowed to request assistance if they faced difficulty in clicking any hyperlinks. In that case, the experimenter would click the link with the mouse. If an incorrect link was clicked by the participant, the experimenter would use the mouse to go back to the previous page. No scrolling was required for the tasks.

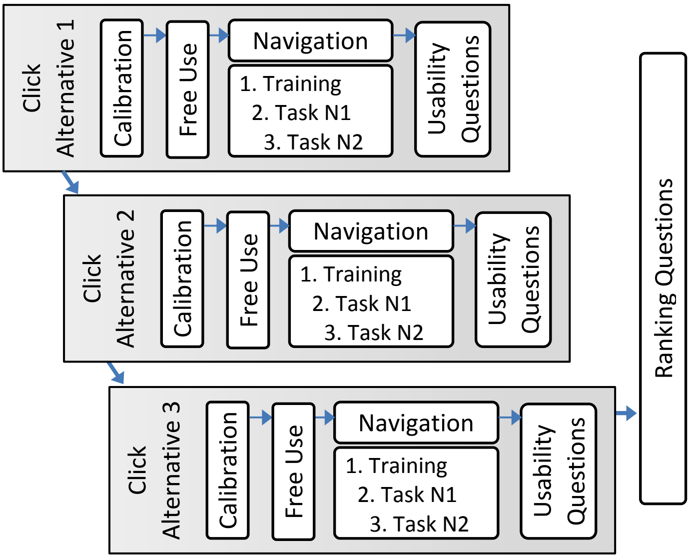


Fig. 8. Experimental procedure for each participant

3.5 Data Collection

The order of click alternatives was permuted to mitigate order bias and training effects. As the experiment (after the pilot) had three alternatives, six permutations were used. Therefore, a multiple-of-six number of participants were recruited.

Events were logged for each web page along with timestamps in CSV log files for each participant's use of each click alternative. The types of events logged included "Gaze on Link," "Button Drawn" and "Click." The data from these log files was analyzed to find the time required for each click, the incorrect links clicked and the mouse clicks done by the experimenter (when assistance was requested by the participant). Six usability questions, derived from the System Usability Scale [23], were answered on a 5-point Likert-scale after using each click alternative. The responses were scored for each participant and averaged to get the usability score of each alternative, with a range from -12 to +12. After completing the experiment, the participants ranked the click alternatives. These ranks were scored (two points for being ranked first, one for being ranked second and zero for being ranked third) and summed up to form a total ranking score for each click alternative. Comments were also solicited in the questionnaire to help explain the rankings.

4 Pilot Study Results

Five participants took part in the pilot study, one of which had difficulties using the eye tracker and could not complete the experiment. All participants were university students aged 19 to 23. The pilot exposed important shortcomings which are documented in this section.

Out of the five click alternatives, Dwell turned out to be very problematic. The first two participants had so much difficulty that it was decided to not evaluate it further. Users found it extremely difficult to hold the gaze on a link for the required 1 second. In case multiple links were present close to each other, an incorrect link would get highlighted or clicked more often than the correct one.

We believed the Radial Confirm alternative to be very promising, but it turned out to have the following problems. The buttons were drawn depending on the reported gaze location. This resulted in a different and unpredictable button arrangement every time the buttons were drawn. If the text had four or more hyperlinks in an area, it was difficult to find the right button to click. While looking for the right button, there was a possibility of the buttons disappearing and more buttons being drawn. If the gaze was near the border of the web page, especially in any of the four corners, the buttons ended up appearing in a non-intuitive pattern. This added to the difficulty of finding the right button associated with each hyperlink. Moreover, the buttons and connecting lines occluded the web page content. In short, this design had many usability issues. According to the participants:

- "Radial Confirm is easier to select links as you don't need to look away to select them. However, it is impossible to read other text."

- "Radial Confirm was bad as while reading the buttons were easily clicked by mistake."

The Dwell and Radial Confirm were discarded from further experimentation. On the other hand, Single and Multiple Confirm buttons were equally preferred in the pilot. The following are some comments from the participants:

- "Single Confirm is easier to use as there are less chances of making mistakes with it. Multiple Confirm was easiest to use but did have difficulties looking at the link as well as looking for the words shown up on the side."
- "Single Confirm was useful, however, having several words on the same line made clicking the button very hard."
- "Single Confirm is better than Multiple Confirm as you don't need to look away if the right link has been selected. However, this can be replaced by something that doesn't cover any text (e.g. underline). In Multiple Confirm, need to look away constantly to check if the right link has been selected and also takes time to do so. So it was distracting when trying to read through text."

Many refinements were made in the Single and Multiple Confirm alternatives after conducting the pilot. An underlying objective of these refinements was to minimize confounding factors by making the Single and Multiple Confirm designs more similar to each other. For both alternatives, text was placed outside the buttons to avoid inadvertent clicking and a crosshair was added in the center of buttons to provide a visual anchor, in accordance with earlier findings [9]. More text was made visible in button labels since some links were not disambiguated easily by only the first five letters. These text labels were placed between the web page and buttons, so that the gaze would not have to travel over the buttons to read the labels. As a result, the connecting line was removed from the Single Confirm alternative as the text label would already associate the button with a hyperlink. Highlighting of hypertext was also considered but not implemented in order to minimize the design space (e.g. color and type of highlighting) and visual distractions.

The size of the right margin was increased accordingly. This increase in size has the disadvantage of using up more screen real estate, which can be problematic on mobile devices. However, the buttons can be moved off-screen for such devices. The users would then just look at the edge of the device next to the text label to click that link (most gaze trackers can still track the gaze when slightly outside of the screen).

Since the predictability of the button location helps in reducing inadvertent clicks, the Single Confirm button was made to only appear vertically centered in the margin. The Multiple Confirm buttons would always start from the top of the margin. Based on observations, the threshold values were adjusted as well. The first threshold to activate the confirm buttons was reduced to 100ms, while the threshold to click a confirm button was reduced to 400ms. The threshold to deactivate or remove the confirm buttons was increased slightly to 700ms.

5 Main Study Results

A total of 19 volunteers performed the experiment, out of which 18 (13 men and 5 women) were successful. The age of the participants varied from 20 to 39, including one professional worker, one post-doctoral researcher, two lecturers and 14 students. The participants reported reading English text for between 10 and 80 (averaging 40) hours weekly and using computers for between 2 and 14 (averaging 7.8) hours daily.

5.1 Ranking

In the click alternative ranking, eight participants ranked Mouse as the best, five participants ranked Multiple Confirm as the best, while one participant ranked Single Confirm as the best (Fig. 9). The Mouse is the clear (Condorcet) winner; Multiple Confirm is the clear second. Some comments of participants who did not rank Mouse as the best are copied below, with minor grammar corrections:

- "Multiple Confirm is much better than Single Confirm and it's more interesting to use than regular mouse click."
- "As I sometime feel RSI, for me this alternative (Multiple Confirm) is good. Though a more mature solution would be great."
- "Multiple Confirm is probably the better one as it gives more options while clicking. Mouse, however, requires using hands which is more physical as compared to eye-tracking single and multiple buttons alternatives."

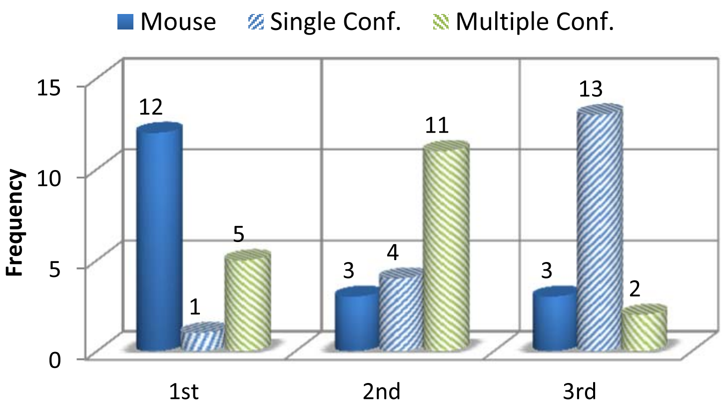


Fig. 9. Histogram of click alternative rankings

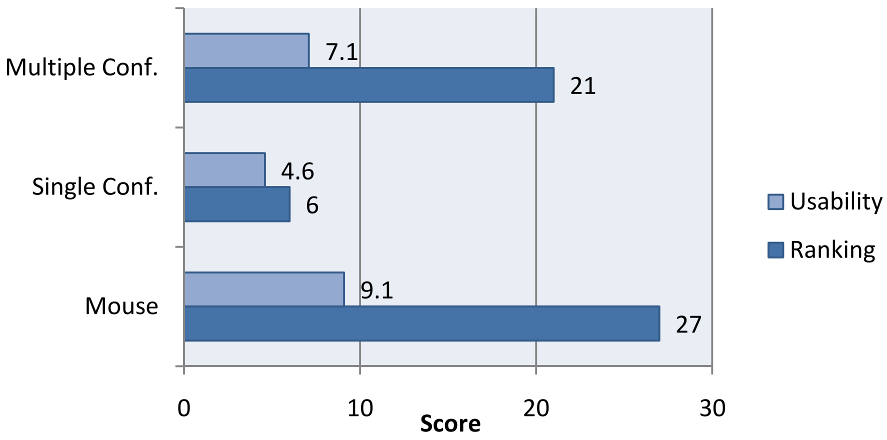


Fig. 10. Total usability and ranking scores of each click alternative

Table 1. Performance of all click alternatives

	Mouse	*Single Confirm*	*Multiple Confirm*
Average time to click a link (sec)	2.1	10.3	7.4
Number of incorrect links clicked	0	4	0
Number of times clicked with mouse	n/a	26	1

Out of the 12 participants who ranked Mouse as best, eight indicated that they preferred the mouse over the gaze-tracking alternatives as they had been using mouse for many years and are so well-trained with it. Some comments of such participants are copied below:

- "I only faced occasional issues in using Multiple Confirm buttons. It is unfair to compare Mouse with eye gaze tracking as I am well-trained with the mouse but used eye gaze tracking just once. My response might be different if I had used gaze tracking for longer."
- "I found the Multiple Confirm to be better than Single as it was easier to choose between links close by using multiple buttons. Mouse is only the best because it is what I am most used to."
- "If I use the mouse too much, I might just end up having CTS in the future. I found Multiple Confirm buttons just right. It is a promising click alternative. I think there's something that can be done about the head movements."

5.2 Usability

The usability scores (Fig. 10) confirm the user satisfaction measured by the ranking, i.e. Multiple Confirm is a bit worse than Mouse, and Single Confirm is the worst.

As evident from the comments, two participants mentioned Repetitive Strain Injury (RSI) or Carpal Tunnel Syndrome (CTS) as a concern with using the mouse. The equipment's constraints on head movement, which was also mentioned in the responses, likely had a negative effect on the usability of the gaze alternatives.

5.3 Performance

Looking at the data from the navigation tasks, Single Confirm does not seem feasible as it was the slowest as well as the most inaccurate (Table 1). While using Single Confirm, 16 participants required clicking with the mouse at least once, as they could not get the required hyperlink activated in the margin. Four users ended up on an incorrect page by mistake. They thought that the button appearing in the margin was for the correct hyperlink and did not read the button label to validate this. The more difficult navigation task of the two had more problems due to the higher hyperlink density. Single Confirm did not take a lot longer than Multiple Confirm, but it did require many 'mouse clicks' and suffered from inadvertent clicking. Both Multiple Confirm and Mouse were fully accurate and did not result in any incorrect click. There was only one instance with Multiple Confirm where the participant asked for assistance and the link was clicked with the mouse. Pairwise testing of the average time to click a link indicates that the differences between all click alternatives are significant at the 1% level, using the Wilcoxon Signed-Rank test.

The following are some observations from the experiment:

1. Some participants were impressed and excited to use eye gaze tracking with comments like "that's pretty cool," "wow" and "that's so quick". This effect was most prominent when Multiple Confirm was being tested after the Mouse but before the Single Confirm alternative.
2. Single Confirm seemed to be easier to learn and easier to use than Multiple Confirm as it did not require choosing from multiple buttons (less cognitive load). But due to inherent inaccuracy, it was the most irritating and error-prone.
3. Multiple Confirm was easiest to learn and use when it was the last click alternative tested. This learning effect indicates that users can get better at using this click alternative with more practice. This effect was not noted with Single Confirm, probably because it was already relatively easy to learn.

6 Discussion

The time taken to click during the navigation tasks shows that Multiple Confirm was slower than the Mouse, but the users did not report that as a shortcoming. One user did mention that the speed of the computer was slow, which was evident by the time taken to load webpages (about 5 to 7 seconds depending on page size) in all click alternatives.

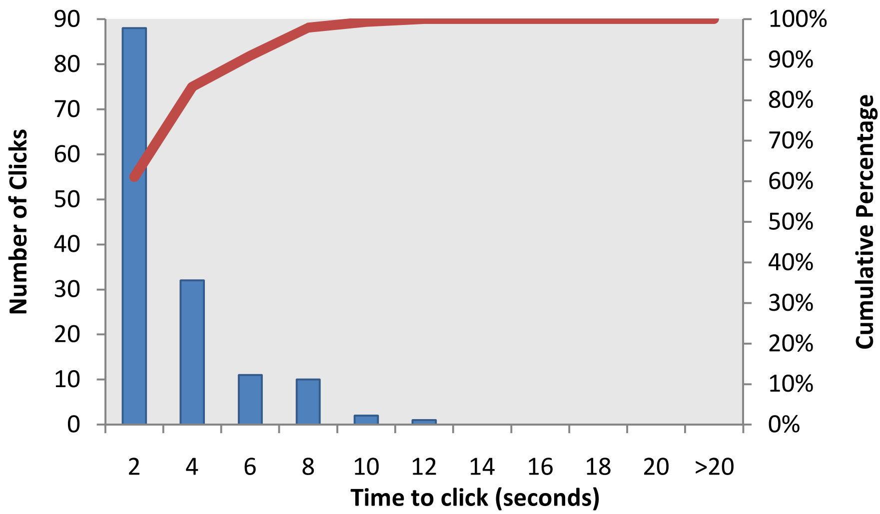


Fig. 11. Frequency distribution of time to click using Mouse

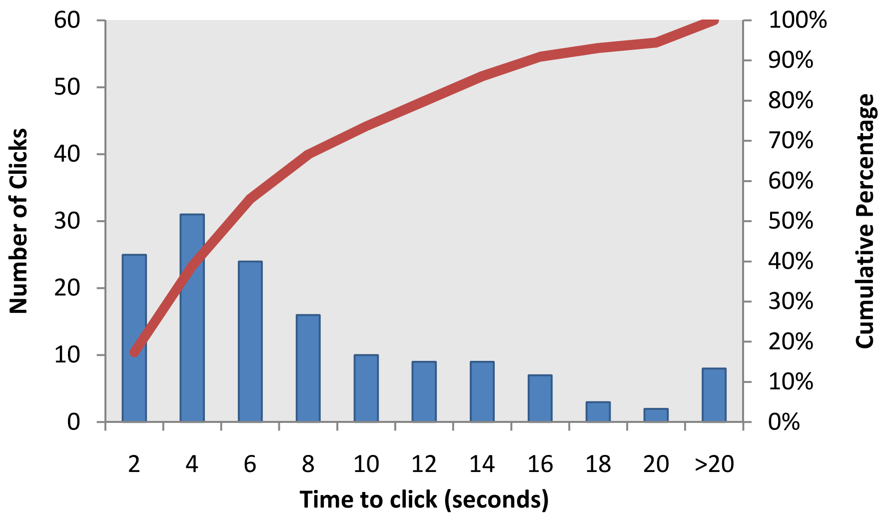


Fig. 12. Frequency distribution of time to click using Multiple Confirm

Single Confirm is problematic and not feasible due to the inherent gaze tracker inaccuracy as well as eye jitter. It was difficult for users to activate the confirm button by looking at a particular link. The problems are similar to those of the "Dwell" click alternative, in which users had difficulty holding the gaze on the correct hyperlink.

Multiple Confirm, as expected, was not as fast as the mouse. But contrary to expectations, not all participants found the mouse to be the best. Even some of those who judged it best believed that with some more practice they would likely change their ranking. The Multiple Confirm buttons were able to solve the two problems associated with gaze tracking, namely inaccuracy and inadvertent clicking. It should be noted that conceptually the Multiple Confirm buttons are performing the same function as magnification techniques, without any visual magnification. The small size of hyperlinks makes it difficult to click with gaze tracking due to the inherent lack of precision, equipment error and involuntary eye movements. Multiple Confirm buttons seem to be able to compensate for these difficulties.

In order to better understand the difference in the average times to click for Mouse and Multiple Confirm, further analysis was performed by looking at the frequency distributions and cumulative percentages of the times to click (Figs. 11 & 12). The analysis shows that most of the links were clicked within 2 seconds with the mouse and within 6 seconds using the Multiple Confirm buttons. Although not as good as the performance of the mouse, Multiple Confirm worked reasonably well. With further research and better equipment (e.g. allowing for more head movement), we expect that the performance of Multiple Confirm can be improved further.

Multiple Confirm involves two steps for each click: first, looking at the hyperlink; then, activating the correct confirm button. Consequently, we expected Multiple Confirm to be slower than the mouse. However, the time difference is bigger than we expected. It can be explained by considering the following additional factors:

- Evaluating the labels next to the confirm buttons to find the right button took some time.
- Sometimes the desired button was not among the buttons shown in the margin, and the user had to look back at the hyperlink to activate the confirm buttons again.
- Activating the confirm buttons again delayed the clicking time further, due to the time thresholds for removing the previously activated buttons and activating the buttons again.
- Occasionally, users had difficulties triggering a confirm button by dwelling on it, especially when the head was moved slightly outside the optimal tracking volume. Some users quickly moved their heads a bit to get the accuracy back, and in some cases this did not help or even decreased the accuracy.
- During page loading, some participants already moved the mouse near the expected location of the next hyperlink, before the page was loaded. This somewhat obfuscated the speed advantage of the gaze over the mouse.

7 Limitations

One limitation was the time it took for the web browser to load and display web pages after a click was performed. These page load times varied from 3 to 10 seconds, depending on the amount of text and the number of images on the web page. Since the log file had only recorded the times between the clicks including the page load times, the page load times were subtracted from the click times later for analysis. The load time for each page was fairly constant, so this did not affect the precision of our measurements much. However, as mentioned before, the page load times favored the mouse as participants had time to move the mouse to the expected areas of the screen before the page was fully loaded.

The other major limitation was the restriction on head movement. Not only was this restriction unrealistic, it also added to the inaccuracy of the gaze click alternatives. The navigation task had to be restarted for a few participants as their gaze was not being tracked anymore after they had moved their heads more than the equipment allowed.

The findings cannot be generalized to all users as most participants had similar demographics. Additionally, only two participants had used an eye tracker before, in another experiment. Finally, it is not unreasonable to assume that there was a bias in favor of the gaze tracking alternatives due to the novelty factor.

8 Conclusion

An important finding is that eye gaze tracking has the potential to be used as an input method for web browsing. The inherent inaccuracy due to eye jitter and equipment error, lack of precision and inadvertent clicking can be compensated for by using the Multiple Confirm click alternative. The confirm buttons should have visual anchors, and button labels (text) should be outside the buttons, preferably between the web page and the buttons. Such a user interface was easy to learn and use, and was preferred by some users over the mouse despite using it only for a few minutes.

The biggest challenge was the limited freedom of head movement afforded by the equipment used. Even slight head movement could cause the gaze error to increase and affect the performance. As newer equipment allows more freedom of movement, the usability of the Multiple Confirm click alternative is expected to improve.

Further research is necessary to evaluate if hyperlinks can be highlighted in a non-distracting manner as visual feedback to efficiently identify the hyperlinks that have activated confirm buttons. Underlining the hypertext seems to be a feasible option. The order and location of confirm buttons might also be improved, e.g. to have the multiple confirm buttons centered vertically. There might also be a potential of using 'grab and hold' [24] to compensate for eye jitter and equipment error when clicking a confirm button.

There is also room for improvement in other design parameters. The inactivation, or removal, of confirm buttons (currently after 700ms of looking at other text) could be optimized. The time thresholds to activate and click the confirm buttons (currently 400ms) and the boundary radius of the gaze area from which buttons are activated (currently 30 pixels) may be fine-tuned further. The way confirm buttons are currently drawn in the margin is a potential distraction, which could be minimized in many ways, e.g. by keeping all the buttons visible, possibly with reduced contrast. The labels would only appear and the button would be made fully visible when a confirm button is activated. The labels may also be faded in and out, instead of appearing and disappearing abruptly.

Acknowledgement. We would like to acknowledge the assistance of Yoonseok Hong and Gen Li in designing the prototypes and performing the pilot study, and the voluntary participation of all users without which this study would not have been possible.

References

1. Stampe, D.M., Reingold, E.M.: Selection by Looking: A Novel Computer Interface and its Application to Psychological Research. Studies in Visual Information Processing 6, 467–478 (1995)
2. Ware, C., Mikaelian, H.H.: An Evaluation of an Eye Tracker as a Device for Computer Input, 183–188 (1987)
3. Hayhoe, M., Ballard, D.: Eye Movements in Natural Behavior. Trends Cogn. Sci. (Regul. ed.) 9, 188 (2005)
4. Just, M.A., Carpenter, P.A.: Eye Fixations and Cognitive Processes. Cognit. Psychol. 8, 441 (1976)
5. Land, M.F.: Eye Movements and the Control of Actions in Everyday Life. Prog. Retin. Eye Res. 25, 296 (2006)
6. MacKenzie, I.S.: An Eye on Input: Research Challenges in using the Eye for Computer Input Control, 11–12 (2010)
7. Jacob, R.J.K.: The use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is what You Get. ACM Trans. Inf. Syst. 9, 152–169 (1991)
8. Javal, E.: Essai Sur La Physiologie De La Lecture. Annales d'ocullistique 80, 61–73 (1878)
9. Penkar, A.M., Lutteroth, C., Weber, G.: Designing for the Eye: Design Parameters for Dwell in Gaze Interaction, 479–488 (2012)
10. Majaranta, P., Räihä, K.: Twenty Years of Eye Typing: Systems and Design Issues, 15–22 (2002)
11. Møllenbach, E., Lillholm, M., Gail, A., et al.: Single Gaze Gestures, 177–180 (2010)
12. Huckauf, A., Goettel, T., Heinbockel, M., et al.: What You Don't Look at is what You Get: Anti-Saccades can Reduce the Midas Touch-Problem, 170–170 (2005)
13. Kammerer, Y., Scheiter, K., Beinhauer, W.: Looking My Way through the Menu: The Impact of Menu Design and Multimodal Input on Gaze-Based Menu Selection, 213–220 (2008)
14. Urbina, M.H., Lorenz, M., Huckauf, A.: Pies with EYEs: The Limits of Hierarchical Pie Menus in Gaze Control, 93–96 (2010)
15. Zhai, S., Morimoto, C., Ihde, S.: Manual and Gaze Input Cascaded (MAGIC) Pointing, 246–253 (1999)
16. Lankford, C.: Effective Eye-Gaze Input into Windows, 23–27 (2000)
17. Skovsgaard, H., Mateo, J.C., Flach, J.M., et al.: Small-Target Selection with Gaze Alone, 145–148 (2010)
18. Ashmore, M., Duchowski, A.T., Shoemaker, G.: Efficient Eye Pointing with a Fisheye Lens, 203–210 (2005)
19. Kumar, M., Paepcke, A., Winograd, T.: EyePoint: Practical Pointing and Selection using Gaze and Keyboard, 421–430 (2007)
20. Greene, H.H., Rayner, K.: Eye Movements and Familiarity Effects in Visual Search. Vision Res. 41, 3763–3773 (2001)
21. Biedert, R., Buscher, G., Schwarz, S., et al.: The Text 2.0 Framework: Writing Web-Based Gaze-Controlled Realtime Applications Quickly and Easily, 114-117 (2010)
22. Aleatech: IntelliGaze, Desktop 2.0 - Web-Browsing with Gaze Control (2011), `http://youtu.be/ks_KDEbDSuw`
23. Brooke, J.: SUS-A Quick and Dirty Usability Scale. Usability Evaluation in Industry (1996)
24. Miniotas, D., Špakov, O.: An Algorithm to Counteract Eye Jitter in Gaze-Controlled Interfaces. Trial, 1 (2004)

Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch

Jayson Turner[1], Jason Alexander[1], Andreas Bulling[2], Dominik Schmidt[3], and Hans Gellersen[1]

[1] School of Computing and Communications, InfoLab21, Lancaster University, Lancaster, LA1 4WA, United Kingdom
[2] Max Planck Institute for Informatics, Perceptual User Interfaces Group, Campus E1 4, 66123 Saabrücken, Germany
[3] The Human Computer Interaction Lab, Hasso Plattner Institute, Prof-Dr-Helmert Str. 2-3, 14482 Potsdam, Germany

{j.turner,j.alexander}@lancaster.ac.uk, andreas.bulling@acm.org, dominik.schmidt@hpi.uni-potsdam.de, hwg@comp.lancs.ac.uk

Abstract. Previous work has validated the eyes and mobile input as a viable approach for pointing at, and selecting out of reach objects. This work presents Eye Pull, Eye Push, a novel interaction concept for content transfer between public and personal devices using gaze and touch. We present three techniques that enable this interaction: Eye Cut & Paste, Eye Drag & Drop, and Eye Summon & Cast. We outline and discuss several scenarios in which these techniques can be used. In a user study we found that participants responded well to the visual feedback provided by Eye Drag & Drop during object movement. In contrast, we found that although Eye Summon & Cast significantly improved performance, participants had difficulty coordinating their hands and eyes during interaction.

Keywords: Eye-Based Interaction, Mobile, Cross-Device, Content Transfer, Interaction Techniques.

1 Introduction

We are surrounded by out-of-reach digital information. Our private TVs and public shared displays often present URLs, physical addresses, phone numbers, route descriptions, and other information that we wish to 'pull' to our personal devices. Equally, we often wish to add personal content to notices, discussions, presentations and collections on shared screens. Yet we lack fluid mechanisms for moving content between public and personal displays.

We present Eye Pull, Eye Push, a novel interaction concept that allows for the acquisition (pulling) and publication (pushing) of content between personal and remote devices. Using a combination of gaze and touch it is possible to define techniques that enable this interaction style. Gaze is a natural modality choice for selecting objects that catch our visual attention, while touch actions can be performed on personal

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 170–186, 2013.

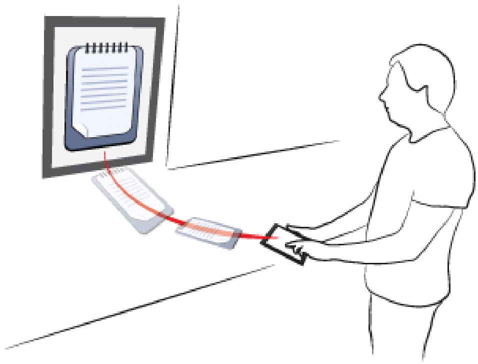


Fig. 1. Eye Pull, Eye Push: users pull and push objects between remote screens and their personal devices with a combination of gaze and touch. In this scenario, the user selects a form on a public service terminal simply by looking it, retrieves it to their touch device with a swipe, fills it in, and returns it with a swipe while looking up at the terminal.

devices without visual attention. Related work has employed handheld device input, combined with gaze interaction, to assist panning and zooming [22] and target acquisition [20] on desktop displays. Our work is distinct in demonstrating the gaze-supported transfer of objects across devices.

Figure 1 illustrates our vision: a user selects an object on a public display and, while still visually fixating on it, swipes on their handheld personal device to pull the object down for editing. Once editing is complete, the user re-fixates on the remote target, and returns the object with a further touch gesture. This style of interaction would benefit many contexts of use: group collaboration, classrooms, and public community displays [11, 17, 7]; our homes for lazy interaction between the TV screen and mobile devices; public terminals that we may find too exposed or too grimy for direct data entry; and anywhere that digital objects exist, that users would like to edit but cannot reach.

This paper makes a two-fold contribution. First, we introduce Eye Pull, Eye Push, a concept for multimodal cross-device content transfer. We define the required input attributes for such interaction and explore application scenarios where it makes a compelling impact.

Second, we define three novel techniques for the transfer of objects between remote screens and personal touch devices, each combines gaze and touch: *Eye Cut & Paste* (ECP): Objects are cut and pasted using gaze and touch tap events. *Eye Drag & Drop* (EDD): Objects are moved using gaze and touch hold/release events. *Eye Summon & Cast* (ESC): Objects are pulled using gaze and a swipe down action, and pushed using gaze and a swipe up action. All three techniques were implemented using a portable eye tracker extended for wider field of view [24]. We evaluated these techniques in a user study to understand their strengths and weaknesses in performance and usability. The results demonstrate that users are able to transfer content efficiently using our techniques, thus validating our approach. ECP and EDD performed similarly, with EDD being preferred due to the continuous visual feedback provided by drag-and-drop. ESC was the fastest of the techniques but was rejected by users due to the more complex hand-eye coordination required.

2 Related Work

2.1 Cross-Device Information Transfer

The case for moving objects easily between handheld devices and larger screens has been made widely, for group work settings [11, 17] as well as serendipitous encounters with public displays [7, 1].

Several works have focused on pushing and pulling content using touch-surfaces as proxies to public displays. Touch Projector [5] demonstrated improvements to work by Tani et al. that enabled the control of remote machinery through live video feeds while maintaining spatial context [23]. Boring's work made use of a phone camera feed to project touches on to public displays to manipulate and move objects. Similarly, Bubble Radar showed how users could interact at a distance using the representation of a public display on a tablet PC [1].

Bragdon et al. developed Code-space, a set of techniques focused on interactions between mobile and situated devices in developer meetings [6]. Their work utilised situated depth cameras and inertial sensors embedded in mobile devices to enable intuitive pointing and information transfer for collaboration. Earlier work by Rekimoto et al. entitled Pick-and-Drop [18] has shown how physical objects can be used to transfer content from one display to another. In this case a pen was used to represent a faux storage device that could pick and drop content.

Several techniques have explored obtaining content at a distance within a single large display. Baudisch et al. investigated different techniques for dragging and dropping objects [3]. Their Drag-and-Pop and Drag-and-Pick techniques used proxies of distant icons to effectively bring them closer to a user. Drop-and-Drag by Doeweling et al. [9] was a technique similar to traditional drag and drop technique that allowed for interaction to be suspended mid-transfer, thus allowing the user to perform fine-grained navigation before dropping an object. The above techniques were all found to be faster than traditional drag and drop for sufficiently distant targets. Finally, Schmidt et al. [19] described a range of interactions made available by combining a mobile phone with a multi-touch surface. Their techniques allow for fluid content transfer, personalisation of the surface and access-control over publicly visible elements.

2.2 Gaze Pointing

Early work on eye-based interaction showed that the eyes could be used as input in desktop environments. However an issue coined by Jacob et al. known as the *Midas Touch Problem* causes unwanted interactions when trying to explicitly issue commands [12]. Dwell-time overcomes this problem by allowing a user to fixate on a control for a set delay before activation occurs. Studies by Jacob et al showed that the delay incurred by dwell-time could be overridden, by using manual input to activate controls interaction can be sped up.

Prior to Jacob et al., Ware et al. [25] examined three picking techniques that used gaze combined with dwell, a virtual button and a hardware button for selection. Their experiments found that confirmation via a hardware button was fastest. It was also found that users would attempt to synchronise their eye movement with hardware button presses, causing occasional selection errors as the eyes move away before selection is confirmed. A fully developed alternative to mouse input using gaze and keyboard commands was demonstrated by Kumar et al. [13].

Further studies have evaluated gaze as an assistive modality for manual input. Zhai et al. [26] developed MAGIC pointing. In their paper they designed two techniques that combined gaze with mouse input: *liberal*, the mouse cursor is warped to objects being looked at, the final selection is performed by the mouse, *conservative*, the mouse cursor is only warped after the user moves the mouse. Their experiment found that users subjectively felt they could interact faster with MAGIC techniques. Their liberal technique was faster than manual input and their conservative technique was slower. Drewes et al. [10] followed up on this experiment by combining gaze with a touch enabled mouse. They found that warping the cursor based on when the mouse was touched, as opposed to moved, reduced the need for mouse repositioning, thus improving the overall speed. Bieg et al. [4] showed however that MAGIC pointing offered no performance boost over mouse only input when used on large displays.

2.3 Multi-modal Gaze Interaction with Public Displays

Gaze-based and gaze-supported interactions with public displays are concepts already explored in the literature. Mardenbegi et al. demonstrated the use of head gestures in combination with gaze to interact with applications on a public display [15]. This work followed the same principles as in the previously described work on pointing. Gaze is used to point, and an additional modality is used to issue commands.

Stellmach et al. evaluated techniques in several works that combine gaze and mobile input, i.e., inertial sensing and touch [22, 20]. Their work developed techniques to navigate large image collections on public displays. Techniques used gaze for pointing while touch and accelerometer values were used to pan and zoom through images [22]. Users perceived increased effort and complexity when panning and zooming. This was considered as acceptable however, as it allowed for simultaneous interactions not usually possible with gaze alone. In a later work they combined gaze with touch commands. This work defined five techniques for the remote selection of varying sized targets in a desktop setting [20]. Their findings gave rise to one technique in particular, MAGIC Tab, which allowed users to tab through a series of objects within close proximity to a users gaze, thus overcoming eye tracking accuracy issues. In further work Stellmach et al. evaluated techniques that utilise a combination of eye and head directed pointing with touch interaction for selection and manipulation of distant objects [21]. Their results highlighted that further improvements are required to allow for more precise distant cursor control with large displays.

2.4 Summary

The literature demonstrates success both using multimodal eye-based interactions for remote target acquisition and using touch-based proxies for distant content interaction. Our work joins these areas by using gaze and touch to pull and push objects between public and close proximity devices.

3 Eye Pull, Eye Push

Here we describe the concept of Eye Pull, Eye Push. 'Pulling' refers to moving content from a public context to a personal one. 'Pushing' refers to the opposite of this, moving from personal to public. The overall concept presents an interaction style whereby these tasks can be completed using a combination of gaze and touch.

Below we outline three techniques designed to pull and push objects. We define the stages of interaction required to transfer content between personal and public displays, and explain how each of our techniques provides the required input attributes for each stage.

3.1 Input and Interaction Flow

The transfer of an object between a public display and a personal device can be broken down in to four main steps: object location, confirmation of selection, destination location, and confirmation of drop. Each of these requires two attributes to be fulfilled: *Locate* (the location of the object or target) and *Confirm* (an action to confirm the location). The three techniques we propose combine gaze and touch actions in different ways. They are able to execute the outlined main steps and fulfil their attributes.

Each technique uses one of three touch commands: *Tap*, *Hold/Release* and *Swipe* and each is performed with a single finger. Tap combines two touch events, touch down and touch up, performed in quick succession. Hold/Release also combines touch down and touch up but they are used in considerably slower succession to confirm actions. Swipe combines touch down, touch moved and touch up, each must be performed in quick succession for the gesture to be recognised. The mappings of touch and gaze for each technique are shown in Table 1.

Table 1. Mapping of gaze and touch input to locate objects and confirm actions. Eye Summon & Cast is split in to two rows for clarity. Eye Summon and Eye Cast each involve a single swipe gesture (down or up) combined with gaze.

	Object Selection		**Destination Selection**	
	Locate	**Confirm**	**Locate**	**Confirm**
Eye Cut & Paste	Gaze	Tap	Gaze	Tap
Eye Drag & Drop	Gaze	Hold	Gaze	Release
Eye Summon	Gaze	Swipe	Swipe	Swipe
Eye Cast	Swipe	Swipe	Gaze	Swipe

3.2 Transfer Techniques

Eye Cut and Paste. The first of our techniques is Eye Cut & Paste; it adopts the familiar Cut & Paste semantic of desktop interaction. The steps of this technique are shown in Figure 2: To pull content, the user looks at an object, they then tap on their tablet to select and cut the object from view. A 'paste' is then performed by looking at the target device and a second tap inserts the object at the gaze location. To push content from a personal display, the same steps can be used, i.e., look at an object on the tablet, tap to select, look at the public display and tap again to drop.

Alternate semantics are possible for this technique, for example, once an object is cut, many copies can be pasted to a destination.

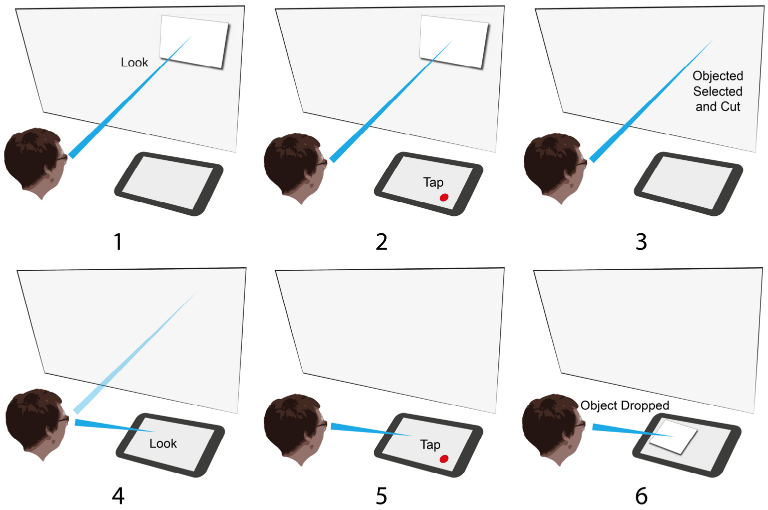


Fig. 2. Eye Cut & Paste: 1) Look at object, 2) Tap on tablet, 3) Object is selected and cut from view, 4) Look at tablet, 5) Second tap on tablet, 6) Object is dropped

Eye Drag and Drop. Our second technique, Eye Drag & Drop is likewise inspired by its desktop equivalent. Figure 3 shows, to pull content, an object is located by gaze and selected by a hold gesture. The object follows a users gaze for as long as they maintain holding with touch. As a user's gaze trajectory intersects the personal device, the object appears on the display. Once touch is released, the object is dropped. Similarly to Eye Cut & Paste, the steps of this technique can also be used to push content, i.e., look at an object on the tablet, hold touch, look at the public display and release touch.

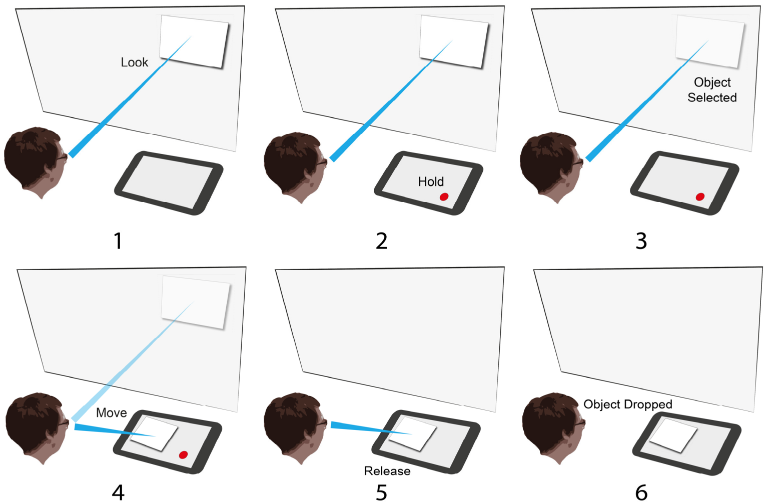


Fig. 3. Eye Drag & Drop: 1) Look at object, 2) Hold touch on tablet, 3) Object is selected and can be visibly moved, 4) Look at tablet, 5) Release touch from tablet, 6) Object is dropped

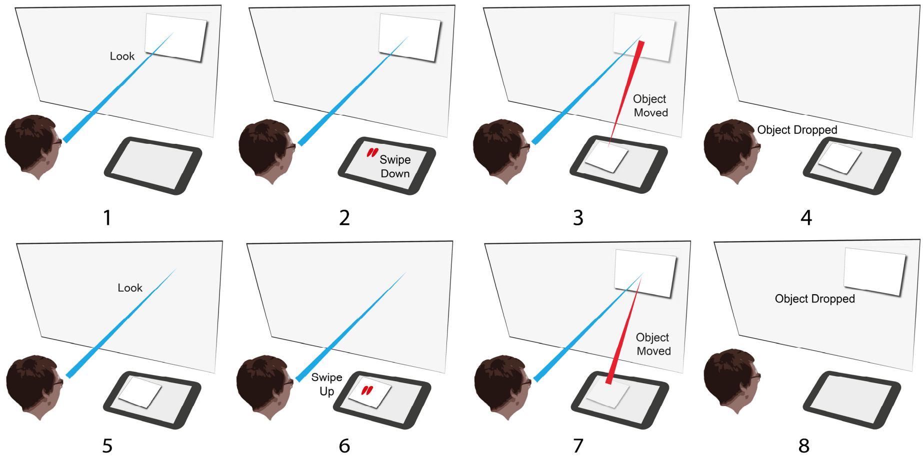


Fig. 4. Eye Summon & Cast. To summon: 1) Look at object. 2) Swipe down on tablet, 3) Object is moved to swipe location, 4) Object is dropped. To cast: 5) Look at destination, 6) Swipe up on object, 7) Object is moved to location of gaze, 8) Object is dropped

Eye Summon and Cast. Our final technique Eye Summon & Cast is based on a combination of gaze with a swipe gesture (see Figure 4). Unlike our other techniques, Eye Summon & Cast uses two differing methods (summon and cast) to pull and push content. An object on the remote screen can be located by gaze, and then summoned with a swipe down on the touch device. The swipe serves to confirm the object selection and simultaneously identifies the destination position on the target touch device. A cast is performed similarly: gaze now selects the destination, and a swipe up identifies the object to be transferred and implicitly confirms selection and drop.

Different semantics are possible for implicit identification, e.g., selecting the most recently 'pulled' object to be pushed back.

4 Application Scenarios

In the section we describe six application scenarios that demonstrate the versatility of Eye Pull, Eye Push. Each of our three techniques has been designed to complete the tasks, pull and push. As the flow of interaction differs between techniques, each can also be used for specialised tasks. Here we consider how each technique could be used in real-world scenarios to pull and/or push content. Table 2 outlines techniques, tasks and connected examples.

Table 2. Example application scenarios for each technique/task combination. Note for Eye Drag & Drop that the examples involve both tasks.

	Pull	**Push**
Eye Cut & Paste	Mid-Transfer Interaction	Duplicating for many users
Eye Drag & Drop	Sharing Read-only Content	Digital Form Filling
Eye Summon & Cast	On-the-go: Acquiring many objects	Sharing Content

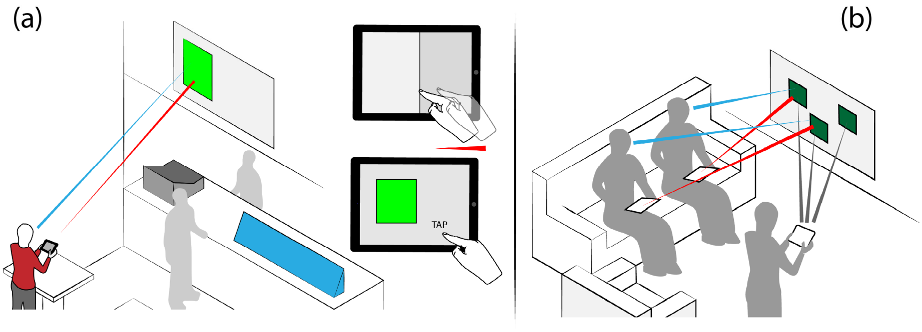


Fig. 5. (a) Mid-Transfer Interaction: A user pulls a flyer, they then switch to a suitable application before tapping to drop it. (b) Duplicating For Many Users: A user pushes three copies of an image using Eye Cut & Paste, two friends now have copies they can pull to keep

Mid-transfer Interaction. Eye Cut & Paste is analogous to desktop cut and paste. The advantages of this technique can be leveraged when pulling content. Traditional cut and paste allows for objects to be selected and temporarily stored on the clipboard. This allows for two further interactions, first it frees the user to perform other (usually navigation) tasks and second it allows for the duplication of content.

As an example, shown in Figure 5a: A user is typing up a document on a tablet pc in a café, a display above the café counter advertises weekly events. The user looks up at the display and notices a digital flyer about a music night at the café. To acquire a copy of this flyer, while still looking, the user taps on their personal device, the content is then held on the clipboard. Now, the user navigates to their calendar application, looking, taps to paste in the flyer and sets a reminder. Next the user switches to their social networking application and pastes in a second copy of the flyer to share with their friends.

Compared to our other techniques, Eye Cut & Paste is specialised to scenarios such as this, where interaction is required mid-transfer to allow content to be used for different purposes.

Duplicating for Many Users. As shown in the previous example, Eye Cut & Paste can be used for the duplication of content that has been ‘cut’. The following example demonstrates how this can be leveraged when pushing content.

A user has cut and pasted a single photograph to a television to show to two other users. The users all like the picture and so want to obtain their own copies. The user performs the paste stage of the technique twice more to create additional copies on the television for the friends to pull to their own devices (see Figure 5b).

Digital Form Filling. Eye Drag & Drop is suited to tasks where changing context is part of the natural flow of interaction, where transfer is performed in a slow and continuous manner.

Figure 6a shows an art gallery, where paintings are displayed along a wall. Next to each art piece is a digital comments display containing the thoughts of gallery patrons and empty comment cards. To leave a comment, a user looks at an empty comment card and pulls it to their tablet. This is performed following the steps of Eye Drag & Drop. The user then fills in the card with their thoughts. The card is then pushed back to the comments display by the same method.

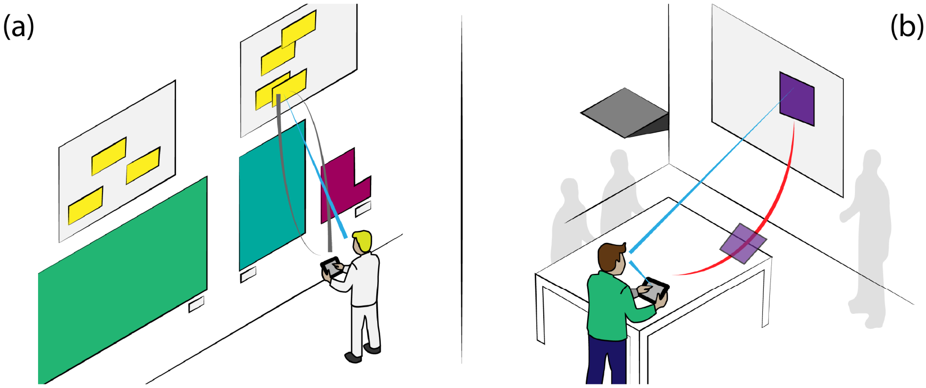


Fig. 6. (a) Digital Form Filling: A user shares their thoughts about artwork on a virtual comments board by pulling, completing and pushing a comment card. (b) Sharing Read-only Content: A user pushes and pulls an image for temporary viewing in a meeting.

As Eye Drag & Drop provides continuous visual feedback, content can be seen to visibly move as it follows a user's gaze. This allows the interaction to become analogous to physical tasks such as filling in and posting comment cards.

Sharing Read-Only Content. Users do not always want others to be able to obtain the content they share. Figure 6b shows how Eye Drag & Drop can be used to share in a read-only manner by maintaining control over content as it is displayed.

This technique allows a user to switch back and forth between large and personal display contexts in a steady and continuous manner: A user is in a meeting; they want to show a relevant image temporarily on a projected display without disturbing the current content. First they look at an image on their personal device and perform a touch hold, this attaches the image to the location of their gaze. The user then looks up at the larger display to show the picture, as they maintain holding their touch, the object does not drop. The user then reverts their eyes back to their personal device, removing the image from the large display. They then release their touch to drop the object.

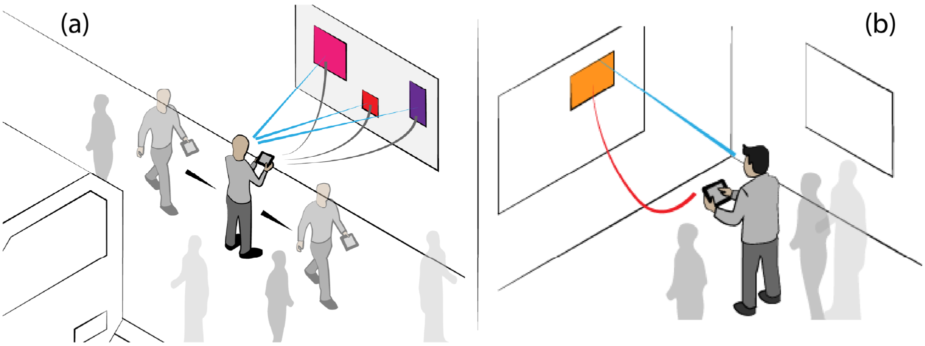


Fig. 7. (a) On-the-go: A user acquires many objects in quick succession at train a station. (b) Sharing Content: A user shares content to a display for viewing by a group of people.

On-the-go: Acquiring Many Objects. Eye Cut & Paste and Eye Drag & Drop require the user to change context between a large and personal display as they transfer an object. These two techniques are best suited to settings where the user's relative movement and schedule are not limited.

Eye Summon & Cast requires the eyes to identify a distant object, drop location is then defined by touch. This mechanism allows for the user to acquire an object without changing their visual context. This allows for the quick acquisition of many objects in sequence while on the go. Figure 7a demonstrates an example: a user has arrived in a busy train station and on the platform is a local information display. The display contains a wealth of tourist centric information on the local area. The user spots a train departure table, a local taxi number and a local map. The user swipes on their mobile device to grab each item in sequence as they pass by the display without having to change context.

Sharing Content. When browsing media on a personal device, users often want to share their experience with a large group on a bigger display. Figure 7b shows how Eye Summon & Cast can be used to allow for fluid interaction in this scenario: while browsing content, the user holds their finger on an image they wish to share. Now, looking at a larger display, the user can swipe upwards on their personal device to transfer the image for viewing. This interaction allows for a simple and natural method of choosing a public display, in particular in environments where more than one large display may exist.

5 User Study

In a user study we aimed to compare our three techniques to evaluate usability to understand which was better suited to each task and to users. We analysed performance and usability measures that were recorded as users pull and push a single object between a large display and a mounted tablet device.

5.1 Participants and Apparatus

We recruited 12 paid participants (11 male, 1 female, aged 22 to 41 (M = 25.4 S.D. = 5.1)), all had normal or corrected vision, and one was "colour-blind" but was able to distinguish the colours used in the experiment. Participants stood 150 cm from a 50" plasma display (whose base was 1 m from the floor). A tablet was mounted on a tripod at waist height. This decision was made to ensure eye-tracking accuracy remained constant throughout trials. This prevented parallax error that is inherent in monocular eye-tracking.

Participants wore a custom eye tracker that was calibrated with each participant at the beginning of the study. The eye tracker is based on SMI's iView X HED system but utilises an additional scene camera to detect personal device screens at close proximity using brightness thresholding with contour detection (see Figure 8) [24]. Contours were minimised to four points representing the rectangular surface of each screen. Gaze was then mapped to this rectangle using a perspective transformation to convert scene camera coordinates to on-screen coordinates. Although the system did

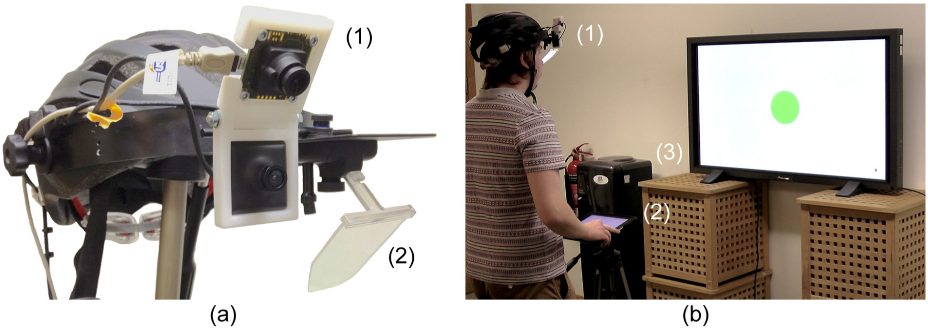


Fig. 8. System Setup: (a) Dual scene camera eye tracking system, (1) Additional scene camera (2) Eye camera. (b)(1) The system setup with head-mounted eye-tracker, (2) Touch tablet mounted on tripod (3) and Plasma TV

not use the commercial software provided, the system was accurate to within 1.5 degrees of visual angle, we found this to be sufficient accuracy for the target sizes used in this study. To compensate for parallax error, the system was calibrated twice, once for the public display and once for the tablet. The system switched between calibrations depending on which screen was in view.

5.2 Experimental Design and Procedure

The study followed a within-subjects repeated-measures design with two independent variables, *technique*, with three levels (1) Eye Cut & Paste (ECP), (2) Eye Drag & Drop (EDD), (3) Eye Summon & Cast (ESC) and *task*, with two levels (1) Pull, (2) Push. The dependent variables were task completion time and error rate. Users were asked to pull and push single objects between displays, this equated to one trial of the experiment.

For each technique participants performed 30 trials: one guided training, five practice, and 24 recorded trials. To begin a trial, participants fixated at a 175 px green circle on the public display and were asked to tap on the tablet. A red target would then appear on the public display. Targets had varying origins but were all located equidistant from the centre of the start point. This was to minimise anticipation when locating the next object. Participants pulled and dropped the object at arbitrary locations on the tablet. Upon dropping the object, its colour changed after a 5 sec delay to blue, prompting the participant to begin the push stage of the task. When pushing, the object had to be dropped within a target area double the size of the object (350 px in diameter) and in the same position from which it had been originally pulled. This was to ensure participants could complete the experiment without introducing a time penalty. All participants used the three techniques (order counterbalanced using a Latin square) and performed all trials with one technique before moving to the next. After completing all tasks with a particular technique, participants provided subjective feedback, including questions from the NASA Task Load Index (NASA-TLX). A final questionnaire gathered preference, task suitability, and general feedback.

All touch and gaze events, task completion times and errors were automatically logged. An error was logged under conditions where, selection failed on the first attempt, an object was dropped out of bounds of a target or an object was dropped out of bounds of a display.

6 Results

6.1 Task Completion Time

Participants completed a total of 864 (24 trials x 3 techniques x 12 participnats) trials. Figure 9 shows the mean completion times for each task. We compared these values in a 2 x 3 (task x technique) two-way repeated-measures ANOVA with Greenhouse Geisser correction. An interaction effect was found ($F_{1.721,18.933}$=5.178, p=.020). Further tests using a one-way repeated measures ANOVA with Greenhouse Geisser correction showed a significant difference for the pull task in completion time between the three techniques ($F_{1.992, 21.137}$=33.812, p<.0005). Further paired t-tests (Bonferroni corrected, new p-value=0.0083) showed that ESC was significantly faster than EDD (p<.0005) and ECP (p<.0005). ECP and EDD were not found to be significantly different (p=1.000).

For the push task, a significant difference was found across completion time ($F_{1.704, 18.749}$=19.235, p<.0005). Further post-hoc paired t-tests (Bonferroni corrected, new p-value=0.0083) showed that ESC was significantly faster than EDD (p<.001) and ECP (p<.001). A significant difference was not found between ECP and EDD when pushing objects.

No significant differences were found between tasks for each technique.

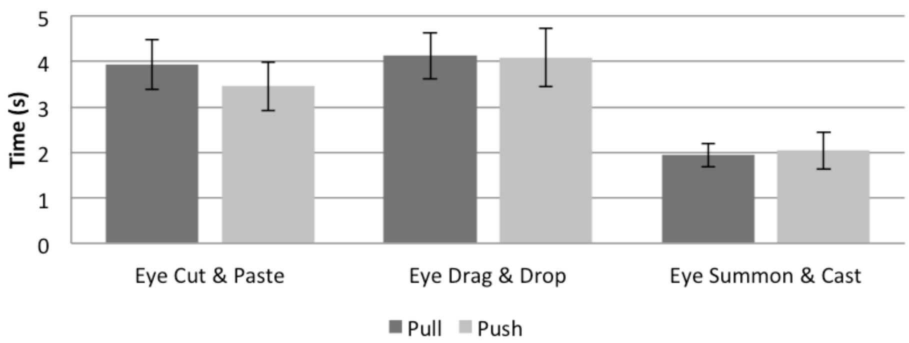


Fig. 9. Mean task completion time in seconds with 95% confidence intervals (CI)

6.2 Error Rate

The mean error rates for each technique are shown in Figure 10. In a 2 x 3 (task x technique) two-way repeated measures ANOVA with Greenhouse Geisser correction, we found no significant interaction or main effects. The means are calculated from 288 trials per technique per task. ECP showed a mean error rate of 1.58 for pulling and 1.66 for pushing. EDD had a higher mean error for pushing than pulling, (2.25 and 0.83 respectively). ESC had a slightly higher mean error rate for pushing (0.83) than pulling (1.25) also.

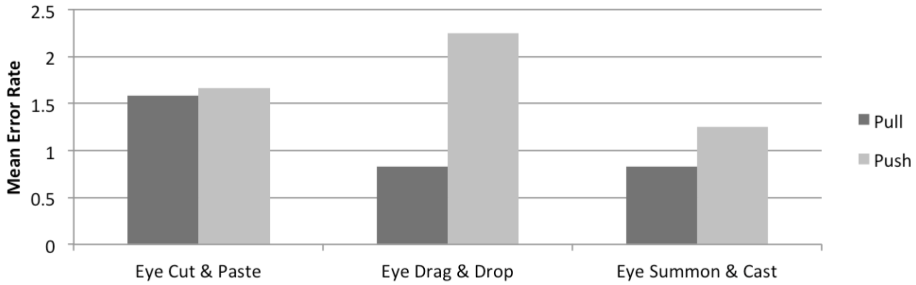


Fig. 10. Mean error rates, confidence levels omitted for clarity

6.3 Performance Perception

We recorded participant responses on a 7-point likert scale to questions regarding perceived speed, accuracy, ease of learning, suitability to task and preference. Friedman tests showed no significant differences in perceived speed for pulling or pushing objects, overall speed, accuracy or ease of learning. Participants were asked questions for each technique relating to their suitability and preference for the two tasks. No one technique was significantly suited to or preferred for pulling objects. For pushing objects ($X^2(2)=9.500$, $P<=.009$) ESC was significantly less preferred than EDD ($Z=-2.756$, $P<.006$) with no significant difference between other techniques and EDD.

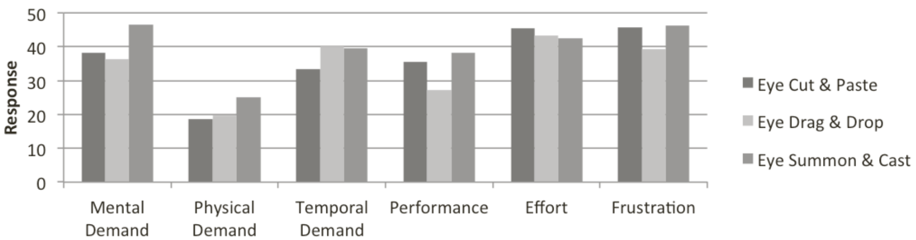


Fig. 11. NASA Task Load Index, scale 0-100, confidence levels omitted for clarity. Key: (ECP) Eye Cut & Paste, (EDD) Eye Drag & Drop, (ESC) Eye Summon & Cast.

Mean responses from NASA-TLX worksheets on a scale of 1-100 are documented in Figure 11. There were no significant differences for any factor. Overall, no one technique was significantly preferred.

6.4 Subjective Feedback

Participants provided subjective comments on the techniques they had just used. Participants commented on the perceived slowness of ECP: *It felt slow because it felt like*

I had to do twice as many actions and it required *a lot of tapping*. Several participants noted the techniques' similarity to its desktop counterpart saying *It's similar to copy and paste*. In comparison to ESC one participant said *I preferred that I didn't have to switch between selection techniques, I was always using my eyes* referring to varying swipe events used in ESC.

Participants perceived EDD to offer more control, stating *I felt I had more control moving objects* and that the *continuous feel of contact with the object* was *something that the other techniques lacked*. The sense of control also affected perceived speed and accuracy, *it felt slow, but it was definitely much more accurate because I could see the object in place before dropping it*. Similarly to ECP one participant found EDD similar to current desktop techniques saying, *It's just like moving windows around in an operating system*.

ESC was found to be difficult for participants: [it was] much harder than other techniques and I didn't know where to look. One participant found during the push task that it was frustrating that I had to look down to find the object, just out of peripheral vision. Finally the variations of swipe to perform summoning and casting were found to be confusing with participants saying, I didn't really like ESC because it had the addition of swiping in either direction.

7 Discussion

7.1 Results

Overall ESC was found to be the fastest but least preferred technique. Participants disliked ESC for two main reasons: (1) Confusion, the touch command used changed between swipe down and swipe up, this lead to confusion about which to use for each task. (2) Coordination, participants stated that they found it difficult to coordinate their hands and eyes. This result highlights an issue where eye-based input needs to correlate more naturally with a users need to use their eyes, to observe other actions they perform. This issue is specific to the requirements of pushing with ESC. The user must swipe up on a tablet-located object viewed in peripheral vision while simultaneously being required to fixate on a large display. A possible solution for this in further work would remove the need for simultaneous initial selection and targeting, and instead allow these to be performed in sequence, i.e., hold finger on tablet object to select, then look at large display, and finally perform a swipe up to transfer the object.

Participants responded well to EDD. In comparison to ECP, participants felt that being able to see the object moving gave them a greater sense of control. Although the system used a gaze cursor to provide continuous feedback to the user, it is clear that in EDD, this feedback is more obvious and familiar to users thus provoking a positive response.

As demonstrated in the example scenarios we outlined in section 4, it is possible to incorporate additional semantics in to our techniques. These can improve usability in more complicated scenarios. Users reported ECP felt slow due to the amount of tap commands required. To improve perceived speed, the paste behaviour can be leveraged in this technique to duplicate a selected object, thus reducing the need for context switching and quicker perception of transfer.

Furthermore, issues outlined above with ESC can be resolved by introducing an 'implicit object identification' semantic. In this case, the most recently pulled object would be pushed automatically, thereby removing the need to redirect visual attention to the touch modality.

7.2 Feasibility and Limitations

Eye Pull, Eye Push is dependant on the deployment of eye tracking as a pervasive technology. To realise such a vision there are several requirements and limitations: (1) Embedded or head-worn eye-tracking: it is imperative that users are always visible to the system. Current technology supports both, remote eye-tracking, where systems are embedded or situated below displays, and head-worn eye-tracking where users wear a personal eye-tracker. These are currently in the form of goggles but envisioned to become as small as standard glasses. (2) Calibration: current head-worn and remote eye-trackers require calibration before use. Calibration takes time and must be performed pre-interaction. More modern systems only require calibration that lasts less that 30 seconds but issues can still arise when interacting with displays at varying distances, this is due to a lack of robust parallax compensation. (3) Connection: users require a method to pair with displays as they interact. Do users implicitly pair with each display they look at? Are user's eye-tracking data globally broadcast for use? Or would authentication be required? To create seamless interaction, there would need to be a balance between privacy and functionality so that users are not inhibited by repeated authentication.

8 Conclusion

In this paper we presented a novel interaction concept Eye Pull, Eye Push, gaze-supported cross-device content transfer. In our design we considered transfer between public and personal devices and how gaze and touch can be combined to create interaction techniques for this task. We outlined the following techniques: Eye Cut & Paste, Eye Drag & Drop, and Eye Summon & Cast. We presented and discussed several usage scenarios for these techniques.

Our three techniques were evaluated in a user study. Users were able to complete the basic tasks of pull and push, and responded most positively to our Eye Drag & Drop technique. The results of our user study showed that Eye Summon & Cast outperformed Eye Cut & Paste and Eye Drag & Drop in terms of speed but was least preferred by users due to its hand-eye coordination requirements. Eye Cut & Paste and Eye Drag & Drop performed similarly in terms of speed although Eye Drag & Drop was preferred due the more apparent continuous visual feedback it provided. In our discussion we outlined how additional semantics can be applied to each technique to extend functionality in differing scenarios. Furthermore we discussed the feasibility and limitations of Eye Pull, Eye Push in the real world.

In future work we aim to explore this design space further, to gain a full understanding of factors within it and the implications they have on this style of interaction, i.e., users proximity to content, display sizes and varying content-types.

References

1. Aliakseyeu, D., Nacenta, M.A., Subramanian, S., Gutwin, C.: Bubble radar: efficient pen-based interaction. In: Proc. AVI 2006. ACM (2006)
2. Ballagas, R., Rohs, M., Sheridan, J., Borchers, J.: Byod: Bring your own device. In: UbiComp 2004 Workshop on Ubiquitous Display Environments, Nottingham, UK (September 2004)
3. 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. Proceedings of Interact, 57–64 (2003)
4. Bieg, H.J., Chuang, L.L., Fleming, R.W., Reiterer, H., Bülthoff, H.H.: Eye and pointer coordination in search and selection tasks. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA 2010, pp. 89–92. ACM, New York (2010)
5. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: Proc. CHI 2010. ACM (2010)
6. Bragdon, A., DeLine, R., Hinckley, K., Morris, M.R.: Code space: touch + air gesture hybrid interactions for supporting developer meetings. In: Proc. ITS 2011 (2011)
7. Carter, S., Churchill, E., Denoue, L., Helfman, J., Nelson, L.: Digital graffiti: public annotation of multimedia content. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1207–1210. ACM (2004)
8. Dickie, C., Hart, J., Vertegaal, R., Eiser, A.: Lookpoint: an evaluation of eye input for hands-free switching of input devices between multiple computers. In: Proc. OZCHI 2006, pp. 119–126. ACM (2006)
9. Doeweling, S., Glaubitt, U.: Drop-and-drag: easier drag & drop on large touchscreen displays. In: Proc. NordiCHI 2010, pp. 158–167. ACM (2010)
10. Drewes, H., Schmidt, A.: The magic touch: Combining magic-pointing with a touch-sensitive mouse. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 415–428. Springer, Heidelberg (2009)
11. Greenberg, S., Boyle, M., Laberge, J.: PDAs and shared public displays: Making personal information public, and public information personal. Pers. and Ubiq. Comp. 3, 54–64 (1999)
12. Jacob, R.J.K.: What you look at is what you get: eye movement-based interaction techniques. In: Proc. CHI 1990. ACM, New York (1990)
13. Kumar, M., Paepcke, A., Winograd, T.: Eyepoint: practical pointing and selection using gaze and keyboard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 421–430. ACM, New York (2007)
14. Lankford, C.: Effective eye-gaze input into windows. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA 2000, pp. 23–27. ACM, New York (2000)
15. Mardanbegi, D., Hansen, D.W., Pederson, T.: Eye-based head gestures. In: Proc. ETRA 2012. ACM (2012)
16. Mardanbegi, D., Hansen, D.W.: Mobile gaze-based screen interaction in 3D environments. In: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications, NGCA 2011, pp. 2:1–2:4. ACM, New York (2011)
17. Myers, B.A.: Using handhelds and PCs together. Comm. ACM 44, 34–41 (2001)
18. Rekimoto, J.: Pick-and-Drop: A direct manipulation technique for multiple computer environments. In: Proc. UIST, pp. 31–39 (1997)

19. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: Proceedings of the Designing Interactive Systems Conference, DIS 2012, pp. 318–327. ACM, New York (2012)
20. Stellmach, S., Dachselt, R.: Look & touch: Gaze-supported target acquisition. In: Proc. CHI 2012. ACM (2012)
21. Stellmach, S., Dachselt, R.: Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In: Proc. CHI 2013. ACM (2013)
22. Stellmach, S., Stober, S., Nürnberger, A., Dachselt, R.: Designing gaze-supported multimodal interactions for the exploration of large image collections. In: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications, NGCA 2011, pp. 1:1–1:8. ACM, New York
23. Tani, M., Yamaashi, K., Tanikoshi, K., Futakawa, M., Tanifuji, S.: Object-oriented video: Interaction with real-world objects through live video. In: CHI 1992, pp. 593–598 (1992)
24. Turner, J., Bulling, A., Gellersen, H.: Extending the visual field of a head-mounted eye tracker for pervasive eye-based interaction. In: Proceedings of the 2012 Symposium on Eye-Tracking Research and Applications, ETRA 2012. ACM Press (2012)
25. Ware, C., Mikaelian, H.: An evaluation of an eye tracker as a device for computer input. In: Proc. CHI 1987, pp. 183–188. ACM (1986)
26. Zhai, S., Morimoto, C., Ihde, S.: Manual and gaze input cascaded (magic) pointing. In: Proc. CHI 1999, pp. 246–253. ACM, New York (1999)

Tools for a Gaze-Controlled Drawing Application – Comparing Gaze Gestures against Dwell Buttons

Henna Heikkilä

Tampere Unit for Computer-Human Interaction
School of Information Sciences
University of Tampere, Finland
henna.heikkila@sis.uta.fi

Abstract. We designed and implemented a gaze-controlled drawing application that utilizes modifiable and movable shapes. Moving and resizing tools were implemented with gaze gestures. Our gaze gestures are simple one-segment gestures that end outside the screen. Also, we use the closure of the eyes to stop actions in the drawing application. We carried out an experiment to compare gaze gestures with a dwell-based implementation of the tools. Results showed that, in terms of performance, gaze gestures were an equally good input method as dwell buttons. Furthermore, more than 40% of the participants gave better ratings for gaze gestures than for the dwell-based implementation, and under 20% preferred dwell over gestures. Our study shows that gaze gestures can be a feasible alternative for dwell-based interaction when they are designed properly and implemented in the appropriate application area.

Keywords: gaze interaction, eye tracking, drawing with gaze, gaze gestures.

1 Introduction

Eye trackers and gaze-controlled applications enable many disabled users to join the information society independently. Gaze-controlled applications are controlled via eye gaze through an eye tracker. Before use, the eye tracker is calibrated to the user's eyes. Then, during use, the eye tracker follows the user's gaze point and delivers the data to applications. The applications use the data to determine the user's point of interest or intentions. Often, the applications have dwell buttons that the user clicks to control the application. A dwell button is "clicked" when the user's gaze point has remained on the button for a predetermined time (usually 200–500 milliseconds).

Over 30 years, eye tracking research has concentrated mostly on communication. In the last decade, the focus has shifted towards leisure applications, such as games and online communities. Many researchers work to enable disabled users to use applications similar to those that able-bodied users already use. Among these are writing applications, Internet browsers, drawing applications, and games of various types. We have concentrated on drawing applications, and our goal is to implement a drawing application that is easy to use and enables the user to correct their drawing mistakes easily.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 187–201, 2013.

Even when fixating on a target, the eye does not stay still. The small natural movements occurring during fixation are called microsaccades. Microsaccades usually stay within one degree of movement, which translates to remaining within one square centimeter on the screen when the user is sitting at arm's length from the monitor. The jitter caused by the microsaccades makes it difficult to hit small objects. Also, small calibration errors are common, especially after one has been using the tracker for a while. To avoid the problems caused by microsaccades and calibration errors, the controls, such as buttons and menus, need to be larger in gaze-controlled applications than normally. Larger controls take room from the screen, particularly in drawing applications, where most of the screen space is needed for the drawing area.

In gaze-controlled applications, the user uses the eyes to study the feedback the application is giving and to control the application. If these two actions are not well separated from each other, the "Midas Touch problem" may arise: wherever the user looks, a command is issued [11]. This problem is often rectified by extending the dwell time used for determining whether a command has been issued or not.

At first, buttons clicked through dwell time and blinking were the most commonly used input method for issuing commands. A little over 10 years ago, the first gesture-like input methods for gaze-controlled applications were introduced, see [3]. The concept of gaze gestures was first put forward five years ago [1], and it has evolved since. Fundamentally, gaze gestures are predetermined gaze paths, or patterns of eye movements, that are interpreted as commands to the application. Gaze gestures can be short or long, simple or more complex, location-bound or location-independent.

Next, we give an overview of research on gaze-controlled drawing applications and on gaze gestures that are relevant to our research. We then present our gaze-controlled drawing application, called EyeSketch, and show how gaze gestures are used in its tools. Towards the end of the piece, we present the study wherein we compared dwell time to gaze gestures, and we discuss the results. We conclude by considering the work done so far and discussing some future directions for our drawing application.

2 Related Research

To our knowledge, five drawing applications controlled via eye gaze have been presented previously. Four of them utilize only gaze, and the other combines gaze with voice commands.

The first eye drawing application, Eye Painting (also known as EaglePaint), was presented 16 years ago by Gips and Olivieri [2]. Eye Painting was one of the applications for their EagleEyes, an EOG-based eye tracking technology. In Eye Painting, the user was able to draw colored lines on the screen by moving the head and eyes.

Eye Painting utilizes so-called free-eye drawing [16], in which the line of drawing appears wherever the user looks and the person drawing has no way of lifting the pen from the drawing canvas. Thus, every shape is connected to the next by the line. A related problem with free-eye drawing is the lack of separation between drawing and looking around. When gaze is used to control a technology, usually the same channel is used for input and for examining the output. If users want to look around and examine the drawing, they probably want to pause the drawing process for the time being.

To solve the aforementioned problems with free-eye drawing, Hornof et al. [6, 7, 8] created an application called EyeDraw. In EyeDraw, looking and drawing are separated by two 500-millisecond dwell-time spans; drawing of a shape starts only when the gaze point has stayed relatively still for a second [7]. To end the drawing of the shape, the user needs to dwell for a second at the end point. Instead of free-eye drawing, EyeDraw utilizes shapes. In the first version, the user was able to draw only lines and ellipses, but the shape collection later grew to include rectangles and predefined stamps too. EyeArt [14] resembles EyeDraw in many respects, but it has a wider range of drawing tools, including seven different shapes, a text tool, and an eraser tool. In addition, the user can adjust the border thickness of a soon-to-be-drawn shape and fill drawn shapes with color by using the paint can tool.

Van der Kamp and Sundstedt [17] presented a drawing application that combines gaze and voice input. A voice command is used in place of dwelling to control drawing and to access the tools and their properties. These authors claim that their solution solves two problems that plagued the previous applications. First, they wanted to remove the need to dwell, since, they said, the use of dwell frustrates users. Second, they hid the tool menus to free screen space for the drawing canvas and to prevent unintended selections from the menus during drawing or looking. In their solution, tool menus appear only through a voice command.

The three drawing applications discussed above utilize separation of looking and drawing. They share the problem that the position or size of the shape drawn cannot be adjusted, which means that the user needs to draw the shape in precisely the right place and in exactly the right size or else undo/erase it and start again. Yeo and Chiu [18] introduced a third technique when designing their gaze-estimation model that constitutes an attempt to separate looking (or thinking and searching) from drawing through examination of gaze patterns. Their model assumes that when the gaze points are close together, the user wants to draw, and that when the gaze points are mostly far from each other, the user is thinking or searching for something. When the gaze points cluster within an area, that area is determined to be the "area of interest." If a fixation longer than 500 milliseconds falls within the area of interest, the centroid of the gaze points is calculated and drawing begins at that point.

In our drawing application, gaze gestures are used to move and resize shapes. Our gaze gestures utilize simple, one-segment gestures and off-screen space. Next, we present the three gaze-gesture implementations that are relevant for our study. Møllenbach et al. [15] created simple, single-segment gaze gestures. In their one-segment gaze gestures, the gesture was made across the screen: it started from what they called the gesture area and ended in another gesture area, on the opposite side of the screen. The assortment of these Single Gaze Gestures, as the authors call them, is small, but they can be used for simple tasks, such as top-level navigation of applications or controlling one's environment.

Isokoski [9] used off-screen targets in his eye writing application. He used five off-screen targets attached to the monitor frame. The user's gaze was tracked with a head-mounted SMI EyeLink tracker, which was able to track the gaze beyond the screen area when the user was seated 100 centimeters away from the monitor. A short dwell time, 100 milliseconds, was used as the threshold for determining whether the user's gaze actually stopped over an off-screen target or just wandered over it.

Another application using off-screen space is Snap Clutch, by Istance et al. [10]. In Snap Clutch, the user can switch mode in the application by looking outside the screen space. The authors claim that quick glances of this sort are a fast and effortless way to control their application.

In our application, we use the closure of the eyes also. The closure of both eyes for a longer time is a rarely used input method, although it is easier to recognize as intentional than are the more frequently used blinks and winks (closure of one or both eyes for only a short time). Especially in the case of blinks, it is difficult to determine which are intentional and which involuntary – reflexive actions that occur when the eyes are getting dry, as the eyes often do more readily when one is looking at a computer screen. As far as we know, only Hemmert et al. [4, 5] have used the closure of one and both eyes to control applications. By closing both eyes, the user was able to activate text-to-speech functionality in a writing application, and closing just one eye allowed users to switch between modes in a first-person shooter game and to filter other than the most recently used icons from the desktop.

3 EyeSketch: A Gaze-Controlled Drawing Application

Our motivation in the design of our drawing application is twofold. First, we wanted to create a drawing application with which the user can produce pleasing pictures without unintentional gaps between shapes or accidental overlapping of shapes. Second, we wanted to create a new kind of drawing application. Of the preexisting drawing applications, none used objects that can be modified, yet able-bodied users have their choice of many such applications. Moreover, we believe that modifiable objects can solve the positioning problem in addition.

3.1 The Application

We chose an approach wherein the shapes, or objects, drawn can be moved or resized, and in which their other properties can be modified after these are drawn.

For the first version of our application, we implemented basic drawing tools, tools for modifying the shapes drawn, and tools for saving and opening pictures drawn with the application. Tool buttons were placed around the drawing canvas and implemented as dwell buttons. We used 80 × 80 pixels as the size for a tool button. These buttons are selected when the gaze point has stayed on the button for 400 milliseconds.

Our basic drawing tools include tools for drawing rectangles, ellipses, and lines. Before and after drawing of the shapes, their fill color, border color, and border thickness can be changed. To aid in creation of the drawing, a grid is implemented behind the drawing canvas. The user can choose whether to display the grid or not.

For later modification of a shape, we have a *Select* tool. When a shape is selected, its color and line thickness can be changed, and it can be removed with the *Delete* tool.

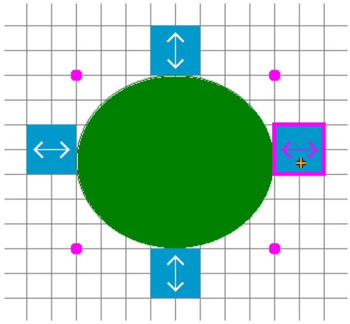


Fig. 1. The resizing handles for the *Resize* tool appear around the selected shape

The *Move*, *Nudge*, and *Resize* tools are implemented with gaze gestures. With the *Nudge* tool, the shape moves one grid square in the direction of the gaze gesture made. With the *Move* tool, the shape starts to move towards the gesture's direction and stops when the user closes the eyes or when the moving shape hits the edge of the drawing canvas. The *Resize* tool causes resizing handles to appear at the sides of the selected shape (see Fig. 1). The size of each handle is 50×50 pixels. A handle is selected by dwell first (100 milliseconds), followed by a gesture; depending on the gesture direction (inwards or outwards), the shape shrinks or grows from the side on which the handle is attached.

We integrated the COGAIN ETU Driver[1] into the drawing application to deliver the eye tracking data from the eye tracker to our drawing application. The ETU Driver supports several makes of eye tracker. Therefore, the application can be used with multiple eye trackers, since the ETU Driver makes sure that the gaze data will be delivered to the application in the same form regardless of the eye tracker used.

3.2 Gaze Gestures

We chose gaze gestures for this use since they are less vulnerable to calibration errors and the jitter in the eyes. Our gaze gestures are simple one-segment gestures that end outside the screen. We also use the closure of both eyes to stop a moving shape.

The gaze gesture used with the *Move* and *Nudge* tools always starts on top of a shape already drawn, proceeds in one of the eight directions (toward a side or corner of the screen), and ends outside the screen (see Fig. 2). We named the eight directions after the cardinal and ordinal directions, with north being toward the top of the screen, northeast toward the upper right-hand corner of the screen, east toward the right side of the screen, etc.

With the *Resize* tool, the gaze gesture starts on top of a resize handle attached to the side of the selected shape; proceeds left/right or up/down, depending on the side on which the handle is attached; and ends outside the screen area.

To be able to start the gaze gesture, the gaze has to stay on the shape or resize handle for 100 milliseconds for it to become selected. The user can start the gesture when the gaze cursor changes its color from black to orange. As the user makes the gaze gesture, three gaze points must fall into the same segment (see Fig. 3) in the direction of the gesture before exiting the screen area. Since the 60 Hz eye trackers take a gaze-point sample once every 16th millisecond, the move from the shape to outside the screen must take at least 64 milliseconds. If it takes more than 1,500 milliseconds, the gesture process stops. When the gaze has remained outside the screen

[1] The COGAIN ETU Driver (i.e., Eye-Tracking Universal Driver) can be downloaded from `http://www.sis.uta.fi/~csolsp/downloads.php`.

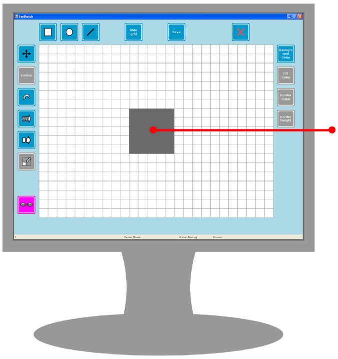


Fig. 2. The gaze gesture starts on top of a drawn shape (or a resizing handle) and ends outside the screen.

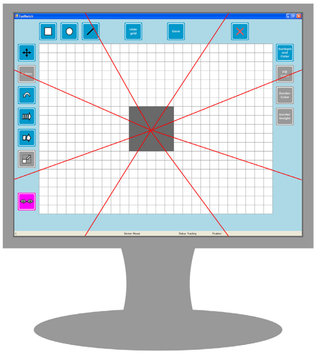


Fig. 3. At least three gaze points must fall into the same gesture segment before the gesture can be completed.

area for 100 milliseconds, the gesture-recognition process ends and the command is issued. A feedback sound is played to the user when they can return the gaze to the screen without canceling the action.

In use of the *Move* tool, the gaze gesture makes the shape move in the direction of the gesture. Closing both eyes for 300 milliseconds stops the movement. While the eyes are closed, the shape keeps moving until the threshold time has been reached. Once that time has elapsed, the shape returns to where it was when the eyes were closed. A feedback sound is played to the user when the eyes may be opened.

4 Gaze Gestures vs. Dwell – An Experiment

To find out whether our gaze gestures would be a feasible input method for the *Move*, *Nudge*, and *Resize* tools, we designed an experiment in which the gaze gestures and often-used dwell buttons were compared.

4.1 Participants

Twelve participants, seven male and five female, volunteered for the tests. Their ages ranged from 18 to 38 years (mean: 24.3 years). Only one of the participants wore eyeglasses during the test. None of the participants had prior experience of eye tracking. Nine participants were familiar with the concept of gestures, and five of them had tried gestures in some form; they reported having used them with cell phones (hand/finger gestures) and with video-game consoles (bodily gestures).

4.2 Apparatus

We used a Tobii T60 eye tracker with a sampling rate of 60 Hz to track the participant's gaze. The resolution of the screen was set to 1280 × 1024 pixels (17-inch LCD screen with a width of 338 mm and height of 272 mm).

The test application was a light version of our drawing application: only the *Move*, *Nudge*, *Resize*, *Undo*, and *Look around* tools were available. In each task, the object to be moved or resized was already drawn in the drawing area. The target size and position were indicated through a similar object with a thick red border (see Fig. 4).

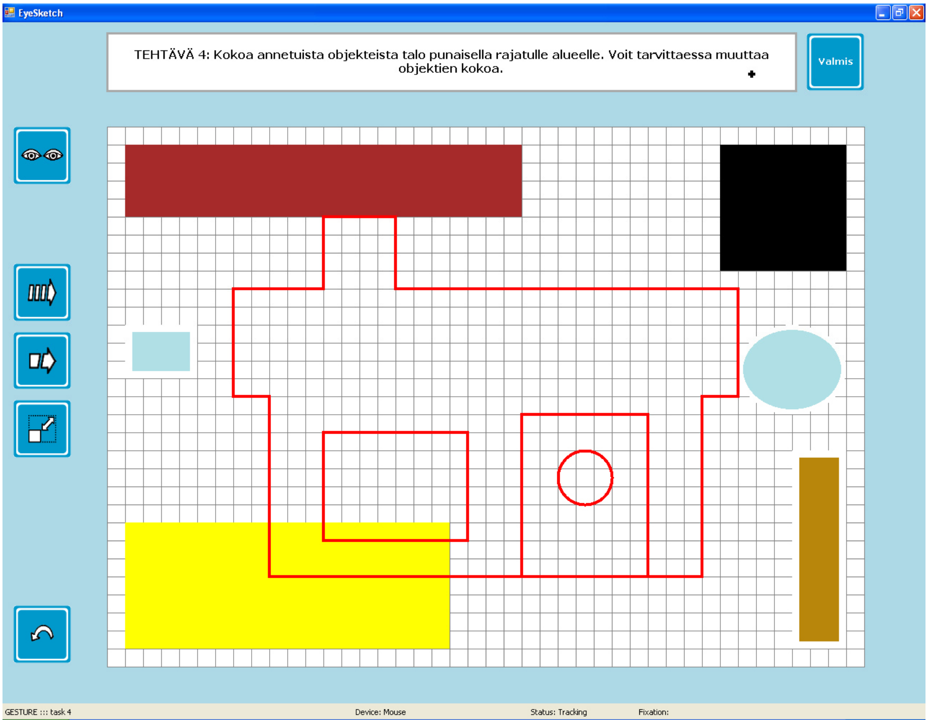


Fig. 4. Test application showing Task 4: Use drawn objects to build a house in the target area. Resize the objects when necessary.

4.3 Dwell Implementation

The dwell implementation differs from our gaze-gesture implementation described above only in terms of the implementation of the *Move*, *Nudge*, and *Resize* tools.

With the *Move* and *Nudge* tools, eight dwell buttons, with arrows showing the direction appear, around the selected shape (as shown in Fig. 5). The participant needs to keep the gaze on the button for 400 milliseconds for it to be clicked. The *Move* tool makes the shape start to move in the given direction when the arrow button is clicked. The movement is stopped in the same way as in the gesture implementation: by closing of both eyes. In the implementation of the *Nudge* tool, the shape moves one grid step in the given direction and stops automatically.

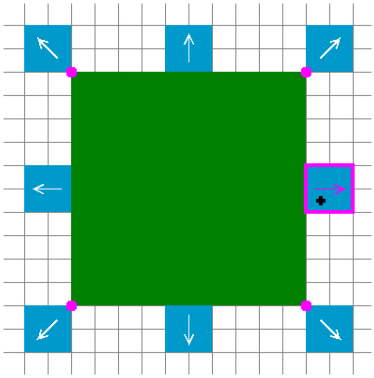


Fig. 5. The dwell buttons for the *Move* tool appear around the selected shape in the dwell implementation

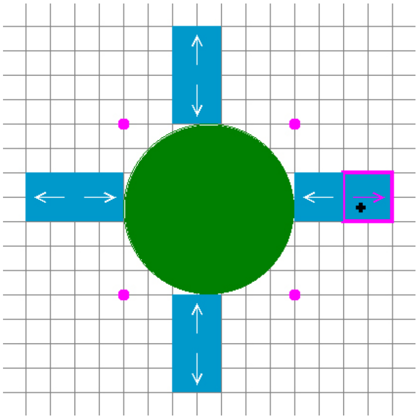


Fig. 6. The dwell buttons for the *Resize* tool appear around the selected shape in the dwell implementation

For the *Resize* tool, a pair of dwell buttons appears on each side of the selected shape (see Fig. 6). The button closer to the shape makes the shape smaller, and the one further from the shape increases the shape's size. When the participant has fixated on the dwell button for 400 milliseconds, it will be clicked and the size of the shape decreases or increases by one step.

The 400-millisecond threshold for dwell time was selected on the basis of literature from the field of eye typing, wherein dwell time is used to select letters from an on-screen keyboard. Majaranta and Räihä [12] concluded in their review that in eye typing studies with novice users, dwell times ranged from 450 to 1000 milliseconds. Majaranta et al. [13] performed a longitudinal eye typing study wherein participants were able to adjust the dwell time. In their study, none of the participants used a dwell time of 400 milliseconds or less during the first session. After five sessions (that is, 75 minutes' practice), most participants had decreased the dwell time to 400 milliseconds or less. For novice users, any dwell time shorter than 400 milliseconds would cause significantly more unintended commands.

4.4 Tasks

We asked participants to perform four tasks with each style of input. The tasks were the following:

1. Move the object drawn to the target area.
2. Use already-drawn objects to build a house in the target area.
3. Resize the object drawn until it matches the target area.
4. Use drawn objects to build a house in the target area. Resize the objects when necessary.

For Tasks 1 and 2, the participant had only movement tools, *Move* and *Nudge*, available. For Task 3, only the *Resize* tool was active. For the fourth task, the participant was able to use both the movement tools and the resizing tool. Tasks 1 and 3 were used to train the participants in use of the new tools. Task 2 was selected to reveal possible difficulties when there are several objects in the drawing area. With Task 4, we wanted to see how well switching from one tool to another works.

4.5 Procedure

Each test took 40–60 minutes. At the beginning of the test, the participant was asked to fill in a questionnaire form, for background information. Then the purpose and the procedure of the test were introduced, and informed consent was requested from the participant. The test had two parts, each using one of the two input styles. The two parts followed the same procedure; only the input style was different. The order of the input styles was counterbalanced.

At the beginning of each part, the experimenter demonstrated the input style with a mouse. Then the participant was seated in front of the eye tracker, at arm's length from the monitor, and the eye tracker was calibrated. After calibration, the experimenter started the testing software and the participant performed the four tasks. After completing the tasks, the participant was asked to fill in a user-satisfaction form. Meanwhile, the experimenter restarted the eye tracker. After a short break, the second part of the test was started. After the second part and the associated user-satisfaction form, the participant was briefly interviewed about the experiences during the test.

4.6 Results and Discussion

We calculated task-completion times, times for completing an action, and the number of unnecessary actions from the data collected. To test the statistical significance of our results, we used repeated-measures ANOVAs with Greenhouse–Geisser correction and paired-sample *t*-tests for *post hoc* pairwise comparisons.

Task-Completion Times. In comparison of the mean times for completion of a full task (see Fig. 7), the dwell implementation was revealed to be faster for tasks 1 and 2, wherein the participants only had to move objects. In tasks 3 and 4, which involved the need to resize the objects in addition, the two implementation types took equally long for completion, on average. This means that the resizing actions take so much longer to complete with the dwell implementation that the advantage gained in the moving actions is lost. Only the main effect for the task was significant ($F_{2,19} = 72.45$, $p < .001$), as can be expected from the nature of the tasks.

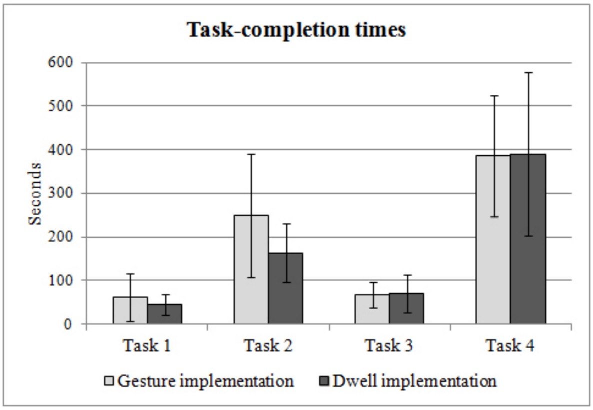


Fig. 7. Task-completion times, in seconds, for the two implementations. The error bars show the standard deviations of the means.

When examining the times to complete the first task with the two implementation types, we found that the first performance of Task 1 took significantly more time than the second one ($F_{1,10} = 5.95$, $p < .05$). This was independent of the implementation type used ($F_{1,10} = 1.47$, $p > .05$). The results demonstrate that it always takes time to figure out how to use one's eyes to control an application when a gaze-controlled application is used for the first time.

Completion Time for an Action. On average, performance of an action was almost equally fast in the two implementation types (see Fig. 8) for all actions. Statistical tests showed that task had a significant effect on the time taken per action ($F_{1,15} = 8.97$, $p < .01$). It also had an interaction effect with the implementation type on the completion times ($F_{1,15} = 7.96$, $p < .01$). Implementation type on its own did not have a statistically significant effect on completion times ($F_{1,11} = 1.19$, $p > .05$).

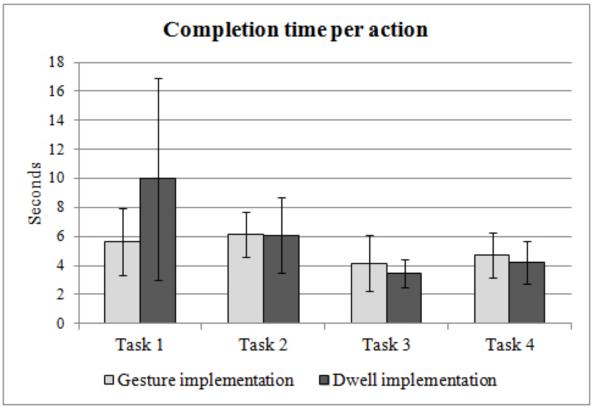


Fig. 8. Completion time per action, in seconds, for the two implementations. The error bars show the standard deviations of the means.

The exception is that in Task 1, the actions with the dwell implementation took almost twice as long as those performed with the gesture implementation. The implementation type had a significant effect on completion time for Task 1 ($F_{1,10} = 5.67$, $p < .05$), and it did not matter which implementation type the participant used first ($F_{1,10} = 0.00$, ns). We believe this result reflects the fact that the participants tried out the gaze gestures more than the dwell buttons before starting to perform the task. We observed in the tests that many participants made several gestures to get a shape to move and then stopped the movement before the shape had moved more than a couple of steps. With the dwell implementation, there was less behavior of this kind.

Excess Actions. We calculated the optimal number of actions for each task. Optimal performance in tasks 1–4 involved 2, 12, 10, and 54 actions, respectively. Only four times during the tests did a participant manage to complete a task optimally. When the task was only to move the shapes (tasks 1 and 2), the participants used more actions in the gesture implementation than in the dwell implementation. However, when the tasks included resizing of the shapes (tasks 3 and 4), more actions were employed in the dwell implementation than in the gesture implementation.

The data support our observations during the tests: the participants had difficulties in resizing the shapes with the dwell implementation, because the dwell buttons to make the shape smaller and larger were next to each other. Because of jitter in the gaze and small calibration errors, the participants often accidentally clicked the wrong dwell button. Then another action was needed to reverse this wrong action. Therefore, to complete one successful action, the participant had to perform three actions.

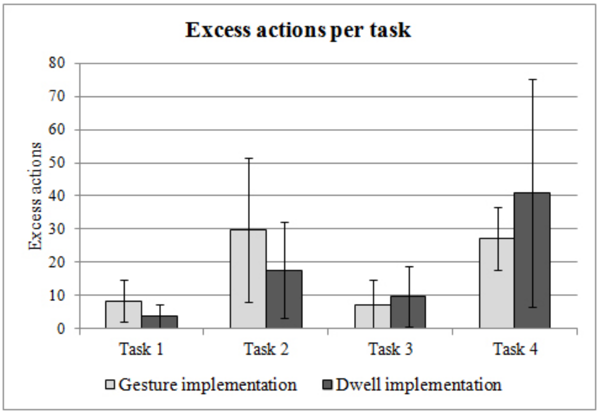


Fig. 9. Number of excess actions per task in the two implementations. The error bars show the standard deviations of the means.

As expected, task had a significant main effect on unnecessary actions ($F_{2,23} = 16.83$, $p < .001$), since the tasks were very different in how many actions were needed for their completion. Implementation type did not have a main effect on the number of excess actions ($F_{1,11} = 0.005$, ns), since whether a given implementation type fared better or worse varied from one task to the next. Statistical testing revealed that task and implementation type had a significant interaction effect ($F_{2,20} = 4.73$, $p < .05$). This supports what is visible in Figure 9. The result means that one implementation type is better for certain tasks than the other, and vice versa.

The participants used more actions than needed in their very first task, no matter which implementation type they started the test with. When the participants started with the gesture implementation, they used, on average, 11.3 actions more than needed for completion of the first task. When facing the same task in the dwell implementation later, they used only 2.3 actions more than the optimum. The participants who started with the dwell implementation performed 5.0 actions more than the number needed in their very first task, and only 5.5 actions more when they later completed the task with the gesture implementation.

The statistical tests showed that there was a statistically significant difference between the two types of implementation in the number of unnecessary actions for Task 1 ($F_{1,10} = 6.10$, $p < .05$). That is, for Task 1, the participants made more unnecessary actions with one of the implementation types (the gesture implementation) than in the other. As described above with regard to completion time per action, we observed during the tests that, with the first task, the participants tried out the use of gaze gestures more than they did the use of dwell buttons. Our results suggest that the use of gaze gestures needs more training in the beginning than that of dwell buttons.

Subjective Impressions. We asked the participants to evaluate their use experience on a seven-point Likert scale (with 1 indicating "strongly disagree" and 7 standing for "strongly agree"). They evaluated their experience on seven dimensions after using both implementation types. When one looks at the average scores, the gesture implementation appears better or at least equally good on all dimensions (see Fig. 10). The largest differences emerged for ease of resizing objects and in interaction speed. For ease of moving objects and on the natural-interaction dimension, the two implementation types were equally good. None of the differences was shown to be statistically significant in Wilcoxon signed-rank testing.

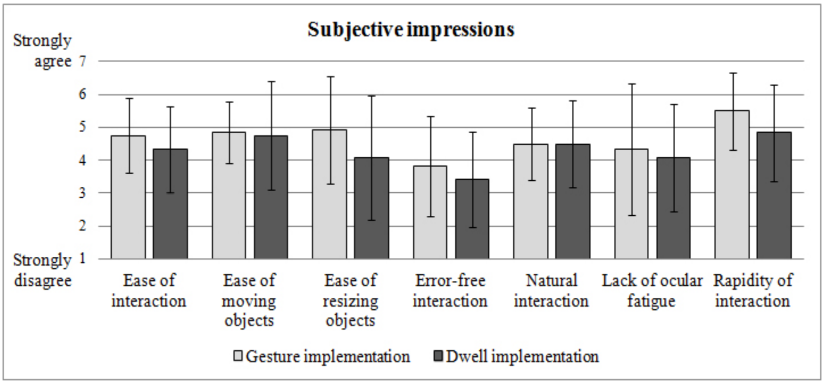


Fig. 10. Subjective impressions gathered after each implementation type. The error bars show the standard deviations of the means.

All participants who started with the dwell implementation rated the gesture implementation better than the dwell implementation or equally good. Participants who started with the gesture implementation were less unanimous with their scores. When we asked about the preference between the two in the interviews, seven participants preferred gaze gestures, four preferred the dwell implementation, and one was undecided. In particular, the difficulty in hitting the correct resizing dwell button in tasks 3 and 4 tipped the scale to gesture implementation. If we had increased the dwell-button size from 50 × 50 pixels or left space between the dwell buttons, such problems might have been avoided. However, seeing the drawing and accessing the objects drawn are essential to drawing applications. Also, in that solution, when several objects are placed close to each other and the user could readily select the wrong one by accident, the resize buttons for the wrongly selected object might hide the intended object and cause a dilemma. In our study, we already gave twice as much space from the drawing canvas to the dwell buttons as in the gesture implementation. Had we given them even more space, the situation would have been neither comparable to the gaze-gesture implementation nor appropriate for drawing applications anymore.

5 Conclusions

As described at the start of the paper, our motivation is to solve problems in existing gaze-controlled drawing applications by creating a new kind of drawing application that utilizes movable, resizable, and modifiable objects. The first step was to establish a functional way to move and resize the objects. For this purpose, we selected gaze gestures, since we wanted to keep the drawing canvas as free of buttons as possible.

The next step was to test whether the gaze gestures could work as well as the dwell buttons that are the traditional input style. The results from the experiment described in this paper are very encouraging. Although the dwell buttons were the better input style for moving shapes, in resizing tasks the gaze gestures proved to be an even better input style, solving all the problems from which the dwell implementation suffered. Furthermore, the participants were able to move and resize the shapes with gaze gestures even when the drawing canvas was half-filled with various shapes. We were excited to learn that most of the participants in our experiment felt that the gaze gestures worked well and that they preferred the gaze gestures to the dwell-button implementation.

In terms of time to issue a command, gaze gestures may never beat dwell buttons. The real advantage of gaze gestures is their ability to remain functional despite calibration errors and low accuracy of the eye tracker. Our resizing task showed how vulnerable the dwell-time input is to even small accuracy problems. Overall, our results showed that the simple, one-segment gaze gestures can be used for tasks other than switching between modes. The only limitation for one-segment gaze gestures is the small size of the gesture vocabulary. However, with adequate planning, the use cases could be numerous.

We have implemented a very usable way to move and resize shapes in a drawing application. Our next two steps are user tests with users from our target user group – i.e., with disabled users – and releasing the drawing application for the public.

Acknowledgements. This work was supported by the Finnish Doctoral Program in User-Centered Information Technology (UCIT).

References

1. Drewes, H., Schmidt, A.: Interacting with the Computer Using Gaze Gestures. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007, Part II. LNCS, vol. 4663, pp. 475–488. Springer, Heidelberg (2007)
2. Gips, J., Olivieri, P.: EagleEyes: An Eye Control System for Persons with Disabilities. In: 11th International Conference on Technology and Persons with Disabilities (CSUN 1996). Los Angeles, USA (1996)
3. Heikkilä, H., Räihä, K.-J.: Speed and Accuracy of Gaze Gestures. Journal of Eye Movement Research 3(2), 1–14 (2009)
4. Hemmert, F., Djokic, D., Wettach, R.: Spoken Words: Activating Text-To-Speech through Eye Closure. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 2325–2330. ACM, New York (2008a)
5. Hemmert, F., Djokic, D., Wettach, R.: Perspective Change: A System for Switching Between On-screen Views by Closing one Eye. In: Working Conference on Advanced Visual Interfaces, pp. 484–485. ACM, New York (2008b)
6. Hornof, A.J., Cavender, A.: EyeDraw: Enabling Children with Severe Motor Impairments to Draw with Their Eyes. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 161–170. ACM, New York (2005)
7. Hornof, A., Cavender, A., Hoselton, R.: Eyedraw: A System for Drawing Pictures with Eye Movements. In: 6th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 86–93. ACM, New York (2004a)
8. Hornof, A., Cavender, A., Hoselton, R.: EyeDraw: A System for Drawing Pictures with the Eyes. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1251–1254. ACM, New York (2004b)
9. Isokoski, P.: Text Input Methods for Eye Trackers Using Off-Screen Targets. In: 2000 Symposium on Eye Tracking Research & Applications, pp. 15–21. ACM, New York (2000)
10. Istance, H., Bates, R., Hyrskykari, A., Vickers, S.: Snap Clutch, A Moded Approach to Solving the Midas Touch Problem. In: 2008 Symposium onEye Tracking Research & Applications, pp. 221–228. ACM, New York (2008)
11. Jakob, R.J.K.: The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. ACM Transactions on Information Systems 9(2), 152–169 (1991)
12. Majaranta, P., Räihä, K.-J.: Text Entry by Gaze: Utilizing Eye-Tracking. In: MacKenzie, I.S., Tanaka-Ishii, K. (eds.) Text Entry Systems: Mobility, Accessibility, Universality, pp. 175–187. Morgan Kaufmann, San Francisco (2007)
13. Majaranta, P., Ahola, U., Špakov, O.: Fast Gaze Typing with an Adjustable Dwell Time. In: 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 357–360. ACM, New York (2009)
14. Meyer, A., Dittmar, M.: Conception and Development of an Accessible Application for Producing Images by Gaze Interaction - EyeArt (2009), `http://wiki.cogain.info/images/d/da/EyeArt_Documentation.pdf` (retrieved)

15. Møllenbach, E., Lillholm, M., Gail, A., Hansen, J.P.: Single Gaze Gestures. In: 2010 Symposium on Eye Tracking Research & Applications, pp. 177–180. ACM, New York (2010)
16. Tchalenko, J.: Free-Eye Drawing. Point (11), 36–41 (2001)
17. Van der Kamp, J., Sundstedt, V.: Gaze and Voice Controlled Drawing. In: 1st Conference on Novel Gaze-Controlled Applications, p. 9. ACM, New York (2011)
18. Yeo, A.W., Chiu, P.C.: Gaze Estimation Model for Eye Drawing. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 1559–1564. ACM, New York (2006)

Designing Gesture-Based Control for Factory Automation

Tomi Heimonen[1], Jaakko Hakulinen[1], Markku Turunen[1], Jussi P.P. Jokinen[2], Tuuli Keskinen[1], and Roope Raisamo[1]

[1] School of Information Sciences, University of Tampere, Finland
{tomi.heimonen,jaakko.hakulinen,markku.turunen, tuuli.keskinen,roope.raisamo}@sis.uta.fi
[2] Department of Computer Science and Information Systems, University of Jyväskylä, Finland
jussi.p.p.jokinen@jyu.fi

Abstract. We report the development and evaluation of a gesture-based interaction prototype for controlling the loading station of a factory automation system. In this context, gesture-based interaction has the potential to free users from the tedious physical controls but it must also account for safety considerations and users' perceptions. We evaluated the gesture interaction concept in the field to understand its applicability to industrial settings. Our findings suggest that gesture-based interaction is an emotional, physically charged experience that has the potential to enhance the work process. Participants' feedback also highlighted challenges related to the reliability of gesture recognition technology in the workplace, the perceived professionalism of gesture-based interaction, and the role of physical feedback in promoting feeling of control. Our results inform the development of gesture-based interaction for similar contexts.

Keywords: Gesture-based interaction, emotions, user experience, field study.

1 Introduction

Gesture-based interaction could provide many benefits as an alternative to physical controls in industrial settings, where the manual operation of machinery can be ergonomically challenging and hazardous or is governed by safety regulations. Several characteristics of gesture-based systems affect their usability in such settings, including technical aspects such as latency of operation, stability, resolution, and precision [1], as well as human-centered aspects such as intuitiveness, ergonomics, and ease of gesture performance and recall [2]. Appropriately designed systems must balance these sometimes contradictory concerns.

We report the findings from a year-long design and development process of a gesture-based interaction concept, focusing on the results of evaluations conducted in real context of use. The overall goal of the research was to create novel interaction concepts for metal workshops and to study their potential in real work tasks. Our research was aimed at answering the following questions:

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 202–209, 2013.

- Which factors affect the acceptability of gesture control in industrial settings?
- What is the user experience of gesture control in industrial settings, especially with respect to its emotional aspects?

2 Related Work

2.1 Gesture-Based Interaction

Gesture tracking technologies can be broadly divided into *perceptual* (utilizing, e.g., vision and audio) and *non-perceptual* (based on, e.g., mouse or touch surfaces) [3]. The taxonomy proposed by Karam and schraefel [3] divides gesture styles between *deictic* (pointing), *manipulative* (tight coupling between the controlled object and hand/arm movement), *semaphoric* (symbolic), *gesticulation* (combination of gesture and speech), and *language gestures* (e.g., sign languages). Due to the prevalence of symbolic gesturing in the real world (e.g., police and military hand signals), we chose it as the starting point for our design.

Empty-handed, or *touchless*, gestural interaction has been proposed to be advantageous in various real world contexts, such as medical environments and environments where physical input devices could be vandalized [4]. In our context of use, the situation is similar to the latter, as introducing input devices into the factory environment is not viable. For example, touchscreens could be smeared with residues, and remote controls could easily be damaged or misplaced. Examples of perceptual gesture-based interaction in a context of use resembling ours can be found in human-robot interaction [5-7]. However, to our knowledge, our work is one of the first user-centered studies of designing gesture-based control to operate factory automation systems.

2.2 Designing Gesture Interactions

Combining the technical and user-centered requirements makes the design of gesture interaction challenging. Although it may be possible to design an efficient and easy to recognize gesture vocabulary for human-machine interaction, one needs to also consider the social implications of gesturing in the workplace. Previous research suggests that gestures that are based on subtle movements and are similar to existing technology and every day actions are preferred to gestures that look uncommon, are physically uncomfortable, or interfere with face-to-face conversations [8]. Additionally, the general meaning of the gesture should be understandable to both the user and bystanders to be socially acceptable [9]. Early stage development that looks past technical restrictions is one way to avoid expending effort on gestures that are not acceptable [8]. However, as Wobbrock et al. [10] point out, reliability of recognition is nevertheless an important criterion for early prototypes. Our design approach is a hybrid of the technology-oriented and human-centered styles; we aimed for reliable recognition through designed gestures, but carried out a series of design workshops to ensure the gestures are also acceptable.

3 Gesture-Based Interaction for Factory Automation

The context of use for the proposed gesture interaction concept is the loading station area, which is a part of a manufacturing automation system (Figure 1). The operator's task is to attach raw materials to a pallet using various tools, which the automation system then moves to machining, and subsequently to remove the machined parts. Using a button panel situated to the side of the loading station, the operator can *open* and *close* the door separating the automated storage from the operating area, *drive* the pallet *in* or *out* of the storage, and *rotate* the pallet in either horizontal direction. The *emergency stop* can also be used at any time.

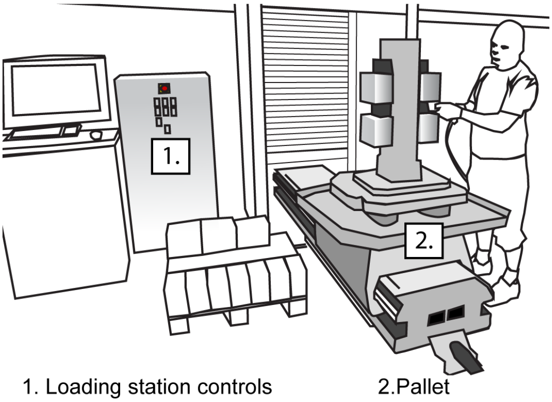


Fig. 1. User attaching materials to the pallet (original image courtesy of Fastems)

The loading station controls are placed away from the pallet due to safety regulations. In crowded workshop conditions, the controls can be hard to reach, and their operation requires constant movement from the pallet to the controls and back. These issues could be alleviated with gestures, as the loading station could be controlled by simply taking a step back to a safe distance from the machine and performing the appropriate gesture, even while grasping tools. The use of gestures would naturally also require strong safeguards, such as multiple means of tracking the operator's location with respect to the machine.

3.1 Gesture Vocabulary for Loading Station Control

The gesture vocabulary was designed in a series of workshops with representatives from factory automation manufacturing industry to ensure its validity in the realistic industrial context. This collaboration was grounded on prior user observations in industrial environments. During the design process we iterated different versions of the gesture vocabulary and visual feedback to ensure the operation was robust and the feedback informative. This process resulted in a gesture set of seven poses (Figure 2). The gestures were designed to be distinctive from one another, so that their recognition would be as reliable as possible using the template-based recognition algorithm implemented for the prototype. The poses also attempt to provide intuitive mappings to the functioning of the loading station through body-centric associations. For example, driving the pallet in and out is mapped to the user's horizontally extended arm, while door operation is mapped to vertically extended arm, both matching the direction of the resulting movement.

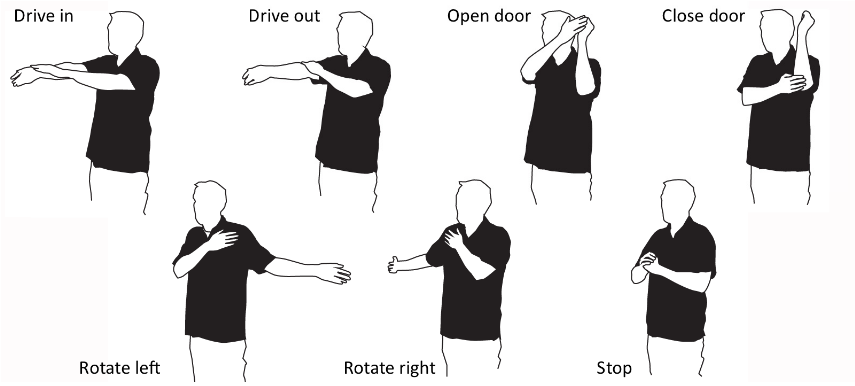


Fig. 2. The proposed gesture set consisting of static hand and arm poses

3.2 Loading Station Simulator

A simulator application was implemented to demonstrate the concept. It contains a 3D model of the loading station and an instruction screen, which would be visible on a separate display in a real world set up. Figure 3 shows the screens as displayed during the field tests. The instruction screen has icons, which depict the pose for each operation and indicate their activation. It provides continuous feedback on system state and gesture recognition, and proved valuable in letting users know that the system is successfully tracking their movements. We utilized the Microsoft Kinect sensor via the OpenNI framework (http://www.openni.org) in early versions of the prototype, and the official Kinect SDK in later iterations.

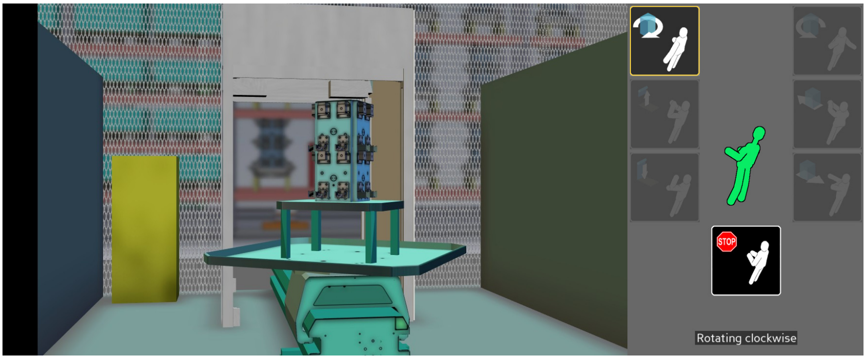


Fig. 3. Loading station simulator (left) and the gesture instruction screen (right)

4 Evaluating Gesture-Based Interaction in Real Contexts of Use

The emotional user experience and acceptance of the concept were evaluated in real workshop environment. The first study was conducted at a factory automation manufacturing site and the second at a local metal workshop, where loading stations similar to the simulated one are used daily. Externally, the prototype was similar in both studies, but internal data processing was updated to improve stability.

4.1 Method

Participants were opportunistically recruited on-site. A total of 22 participants, 2 female and 20 male, took part in the evaluation. They had an average work experience with factory automation loading stations of 6.0 years (range = 0-25 years, s.d. = 6.9). Participant backgrounds ranged from factory automation product designers to assembly line workers and metal workshop laborers.

After a demonstration of the gesture commands by a researcher, the participants were able to briefly practice the gestures. Next, they were asked to perform each of the gestures twice according to a script cued by a researcher, simulating the operation of a real loading station. Interaction took place in front of the simulator, which was presented on a 37-inch high definition television next to the loading station cell.

User experience was measured with a questionnaire, which contained a Likert scale of eleven questions concerning emotional experiences including alertness, anxiety, delight, desperation, determination, excellence, frustration, pride, skillfulness, success, and surprise. In addition, an item concerning the feeling of control was included. Another Likert scale of five statements concerned the acceptability of the gestures: their practicality, necessity, professionalism, intelligence, and physical demand. The questionnaire was filled after the participants had tested the prototype. In addition, short interviews were conducted with nine of the participants.

4.2 Results

Emotional User Experience. The responses to the questionnaire were analyzed using multidimensional scaling (MDS). The resulting configuration shows which emotions are related to each other (Figure 4). The assumption is that similar emotional items are correlated with the same underlying user experience factor, which can be calculated as their sum variable. The items were separated by their emotional valence into pleasant (α = .862) and unpleasant (α = .804) emotional groups, along the x-axis. Cronbach alphas over 0.8 were taken as a confirmation that the items were reliable indicators of the underlying emotional factor.

The results indicate the emergence of two distinct emotional user experience factors: *competence* [11] and *frustration* [12]. Competence, or *self-efficacy*, is linked to the participants being pleasantly surprised by the new suggested interface (Spearman ρ = .771, $p < .001$), and associating feeling of control with very positive emotional experiences, such as pride and success. In addition, a positive association between having a pleasant user experience and feeling in control was observed (ρ = .508, p = .019). Frustration was evident in even slightly negative, or unpleasant, user experience being strongly associated with not accepting the gestures. Accordingly, negative correlations were observed between feeling frustration and perceiving the gestures as intelligent (ρ = -0.542, p = 0.011), professional (ρ = -0.587, p = 0.005), safe (ρ = -0.671, p = 0.001), and needful (ρ = -0.509, p = 0.018).

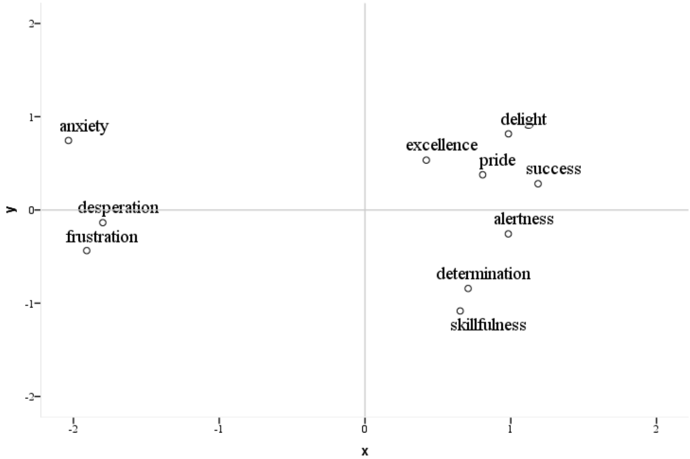


Fig. 4. Multidimensional scaling solution of experienced emotions. The configuration shows that two groups are formed along the x-axis, which indicates the valence of the emotions.

Interview Feedback. The interviews were conducted with metal workshop workers (4 respondents) that operate loading stations daily, and factory automation mechanics (5) that assemble the loading stations. Most felt that the new scheme of interaction was a welcome introduction to the usual work routine. The notable benefits mentioned were the increased freedom of movement, speeding up the work process, improving safety, and easing the learning process of the control interface due to easily learnable gestures. However, many also suggested that the current interface is more reliable, and the shift to non-physical controls would introduce errors without making them easier. Four notable items of doubt were extracted from the interviews:

1. Can the operator trust the delicate sensory in a factory environment?
2. How easily does the sensor break or get dirty?
3. Is it safe to perform elaborate gestures in factory environment?
4. The control interface is already easy. Would it truly be easier to perform gestures instead of walking to the control station to press a button?

When asked about the perceived professionalism of the proposed user interface, the participants made references to attributes not welcome in industrial environment such as "games" and "fooling around". Mostly these references were associated with the gestures themselves: moving arms and body around was perceived as being something not suitable in a factory environment, which suggests gesture control as currently designed might break social norms at the workplace. Physical controls also seem to significantly affect the feeling of professionalism. The most important factor associated with feeling of control was the lack of physical contact with the interface. Comments such as “it felt strange that I couldn’t touch it”, “I’m used to getting direct physical feedback from the button”, and “I feel more in control with the traditional physical buttons” indicated this to be one of the critical factors influencing the acceptance of the new gesture-based interface.

5 Discussion

Our research contributes to the development of gesture-based interaction for industrial workplaces. First, we found that the proposed gesture interface evoked clear emotional experiences as distinct groups of pleasant and unpleasant emotions were observed. These were shown to be associated with feeling of control and acceptance of the concept. In designing gesture-based interaction, it is thus critical to actively support the feelings of self-efficacy and minimize the experienced frustration.

Second, we identified doubts related to the professionalism of gesture-based interaction that are strongly linked to the absence of physical control and the social acceptability of gesturing in the workplace. The first aspect is closely associated with feeling self-efficant: is it possible to feel competent and in control without physical interaction? This resonates with findings from previous research, which suggest that it is important to understand the relationship between the physicality of control and the system being used, and how this affects the users’ preferences [13]. In this respect, an obvious shortcoming of touchless interaction is the lack of tactile feedback, and one must provide a form of feedback that promotes the feeling of control when physical controls are absent. With respect to acceptability, it should be noted that the focus of the research was to develop and study novel interaction concepts for *future* work settings. It is an open question to what extent the concerns with professionalism apply to new generations of users who already have experiences with using gestures in games and with mobile devices.

Finally, results from studies that are not a part of the users’ work practice need to be interpreted with care. Benefits and drawbacks with respect to existing interfaces are difficult to estimate without experiencing gesture-based interaction in the context of real work tasks. For example, would the audiovisual feedback from a real loading station compensate for the lack of physical control? Longitudinal, in situ studies of gesture-based production systems are needed to address the above issues.

6 Conclusion

Gesture-based interaction in industrial setting is a novel and physically charged activity that is also an emotional experience. When designing gesture-based interaction for industrial environments, attention should be paid to creating a system that promotes self-efficant experiences. An optimal interaction is challenging but practical in order to make the interaction an emotionally pleasurable and novel experience. Challenges to address in future work are related to studying the perceived robustness of the gesture tracking technology in real work tasks, designing the physical gesture movements to fit the work context, and compensating for the lack of physical control with alternative feedback methods.

Acknowledgements. This research was carried out as part of the Finnish Metals and Engineering Competence Cluster (FIMECC)'s UXUS program. Appreciation is extended to Matti Nurminen and colleagues from Fastems Oy Ab, Tuomas Kiviluoto from the Agora Center at the University of Jyväskylä, and Mika Käki from Idean.

References

1. Crowley, J.L., Coutaz, J., Bérard, F.: Perceptual User Interfaces: Things That See. Commun. ACM 43(3), 54–64 (2000)
2. 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)
3. Karam, M., Schraefel, M.C.: A Taxonomy of Gestures in Human Computer Interactions. Technical Report, Electronics and Computer Science, University of Southampton (2005)
4. de la Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., Ruschin, D.: Touchless Interaction-Novel Chances and Challenges. In: Jacko, J.A. (ed.) HCI International 2009, Part II. LNCS, vol. 5611, pp. 161–169. Springer, Heidelberg (2009)
5. Burger, B., Ferrané, I., Lerasle, F., Infantes, G.: Two-handed Gesture Recognition and Fusion with Speech to Command a Robot. Auton. Robot. 32(2), 129–147 (2012)
6. Cerlinca, T.I., Pentiuc, S.G., Vatavu, R.D., Cerlinca, M.C.: Hand Posture Recognition for Human-Robot Interaction. In: 2007 Workshop on Multimodal Interfaces in Semantic Interaction, pp. 47–50. ACM, New York (2007)
7. Suay, H.B., Chernova, S.: Humanoid Robot Control Using Depth Camera. In: 6th International Conference on Human-Robot Interaction, pp. 401–402. ACM, New York (2011)
8. Rico, J., Brewster, S.: Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 887–896. ACM, New York (2010)
9. Montero, C.S., Alexander, J., Marshall, M.T., Subramanian, S.: Would You Do That?: Understanding Social Acceptance of Gestural Interfaces. In: 12th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 275–278. ACM, New York (2010)
10. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined Gestures for Surface Computing. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1083–1092. ACM, New York (2009)
11. Gravill, J.I., Compeau, D.R., Marcolin, B.L.: Experience Effects on the Accuracy of Self-assessed User Competence. Inf. Manage. 43(3), 378–394 (2006)
12. Juutinen, S.: Emotional Obstacles of E-learning. Jyväskylä Studies in Computing, vol. 145. Jyväskylä University Printing House, Jyväskylä (2012)
13. Freeman, D., Hilliges, O., Sellen, A., O'Hara, K., Izadi, S., Wood, K.: The Role of Physical Controllers in Motion Video Gaming. In: Designing Interactive Systems Conference, pp. 701–710. ACM, New York (2012)

Metamodels Infrastructure and Heuristics for Metamodel-Driven Multi-touch Interaction

Víctor López-Jaquero, Elena Navarro, Francisco Montero, and Pascual González

Laboratory of User Interaction and Software Engineering
Computing Systems Department
University of Castilla-La Mancha
Albacete 02071 - Spain
{victor,enavarro,fmontero,pgonzalez}@dsi.uclm.es

Abstract. Novice users usually find it hard to manipulate models by using traditional Model-Driven Development techniques, because of the gap between the modeling tools and these users' mental models. In this context, multi-touch interfaces emerge as an alternative to make it easier for novice users to interact with the models by using natural gestures and taking advantage from the popularity that touch-based devices have achieved. In this paper, a metamodel infrastructure and a set of heuristics are presented to automatically generate multi-touch visual editors for manipulating models. The editor generated is driven by a metamodel that also prevents the user from creating not valid models. These heuristics have been validated while developing an environment for novice users, such as psychologists or physiotherapists, for the treatment of people with Acquired Brain Injury.

Keywords: model-driven development, multi-touch interaction, heuristics, Acquired Brain Injury (ABI).

1 Introduction

The exploitation of models during the software development process is a valuable tool for stakeholders to convey their ideas about design, needs or requirements of the system-to-be. Developers, designers or software architects are used to employ graphical node-link notations for the manipulation of models. UML is a clear example of this approach which is widely used, for instance, to specify class diagrams used at different stages of the development process.

However, it is frequently the case that users and/or clients of the system-to-be are required to manipulate models that are used during the development of the system to convey their expertise in the problem domain. This scenario can arise easily in the user-centered design, where the final users of the application are involved in the development. This can be a challenging, or even overwhelming, task for them as they usually do not have the required abilities to tackle modeling activities. This has been the problem we had to face during the development of HABITAT [24], a system to

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 210–227, 2013.

support the relearning process of people with Acquired Brain Injury (ABI). One of the main cornerstones of this system is its functionality to design new types of relearning activities so that the relearning process can be fully customized according to the specific needs of the people with ABI. During the development of this functionality, the exploitation of models emerged as a suitable solution. However, our users (psychologists, speech therapists, physiotherapists, etc) did not have the necessary abilities for manipulating models, although they did have the knowledge we needed about the problem domain. This was the challenge we had to face with the ideas presented in this work: can we provide users, who are novice users, with a tool for manipulating models that hides the complexity behind this task? The solution to overcome this complexity has been to exploit the benefits provided by the integration of two well-known approaches: Multi-touch interfaces [23] and Model-Driven Development (MDD, [13, 31]).

Multi-touch interfaces provide users with attractive and innovative facilities for the manipulation of applications by means of touch gestures. Thus, the interaction can be performed in a natural and intuitive way. Several works, such as [9][16], have shown in the experiments performed that this approach is actually suitable for novice users, as they perceived the interaction with multi-touch applications as more attractive and interesting than traditional applications. Therefore, its exploitation in this work emerged in a natural way, since cooperation with novice users is our main goal.

MDD is not only becoming increasingly popular among researchers, but also among practitioners. It has proved to have a positive influence on the reliability and productivity of the software development process due to several reasons, such as, the exploitation of techniques for the automatic generation of code or the use of models as drivers of the development process. Both reasons led us to its consideration in this work; first to introduce the necessary facilities for the manipulation of models and second to generate model-manipulation multi-touch interfaces in an automatic way. The generation of multi-touch interfaces is performed by automating a set of heuristics, which are presented in this work, that exploit the structure and semantics of the primitives used for domain modeling by using *Ecore* metamodels [11].

The paper is organized as follows. Section 2 provides an overview of ABI, by describing the target population and how this work was conducted. Section 3 presents and analyzes the related previous work. Section 4 describes the metamodels infrastructure to support our proposal. Section 5 describes the heuristics developed. Section 6 presents the initial results of the conducted exploratory evaluation. Finally, Section 0 rounds off the paper by presenting the conclusions drawn and some future work.

2 Case Study: Treatment of Acquired Brain Injury

People with Acquired-Brain Injury (ABI) have suffered “damage to the brain that occurs after birth and which is not related to congenital disorders, developmental disabilities, or processes that progressively damage the brain” [32]. There are several causes of ABI, such as cerebral vascular pathology, skull-brain trauma due to accidents, meningitis, brain tumours, etc. Therefore, it can be stated that just about everybody is exposed to this risk in its daily life. Cases affected by this disability are becoming increasingly common. According to the JCCM Health Council [10] 4 out of every 1000 persons suffer some kind of ABI at some time in their lives. Although people of all ages can be affected, it is more frequent among the younger and older members of the population, as they are more prone to accidents.

According to the experts, the process of integral ABI relearning must include cognitive treatment, in addition to physical and occupational therapies. It should also be emphasized that ABI associations, such as ADACE, which we have collaborated with in different projects, highlight that ABI victims should be provided with a proper treatment as soon as possible, since there is increasing evidence of its effectiveness during the first stages after injury [8]. However, identifying the proper treatment for each person is a difficult and time-consuming task, since brain injury has dramatically varied effects and no two people can expect the same resulting difficulties. This means that an individualized relearning process must be identified for each person. In this context, providing specialists with tools to create and customize the activities and tasks that they use in the processes of recovery [22] is a must. This is the aim of our system: HABITAT [24].

HABITAT enables ADACE specialists to create relearning activities by instantiating the implemented relearning patterns [22]. These relearning patterns were validated by the specialists and were put into practice thanks to the implementation we made in HABITAT. However, the specialists highlighted the need to customize these patterns and this led us to define the relearning pattern metamodel described in Section 4.1 and its implementation in HABITAT. This metamodel had to be instantiated and used by the specialists to define and modify the special relearning patterns they needed. Nevertheless, as these people were not used to manipulating models by using node-link representations, the alternative was to provide them with multi-touch User Interfaces (UI), which are automatically generated by means of the heuristics presented in Section 5 and using the metamodels infrastructure offered in Section 4.

3 Related Work

The generation of UIs or visual metaphors to edit models is not a new trend. All case-tools offer a means of manipulating models by providing a visual notation. Nevertheless, generating a UI to manipulate a domain model is not commonly available. One example of generation of UIs out of a domain model is the Graphical Modeling Framework (GMF, [12]) and Eclipse Modeling Framework (EMF, [11]). By using EMF a developer can create a domain metamodel by means of *Ecore* and then generate a graphical editor by using the GMF framework. The features of this editor are specified by means of a set of models. Another example is Executable UML [19] which aims at generating UIs out of a UML specification. The UIs generated are for standard desktop application, and the interaction is mostly based on *drag&drop* interaction in a tree. The author provides some guidelines regarding how UI is generated, the so called *interactive manifestations*. This approach is not intended for metamodel manipulation, but introduces interesting ideas regarding the generation of UI out of UML models. Nevertheless, the generated UI is not appropriate for novice users and the user has no feedback regarding multiplicities in the specification of the cardinalities of the relationships. Another approach pursuing the generation of UI out of object-oriented specifications is *Naked Objects* [26]. In Naked Objects, the applications are specified solely by using domain entity objects. A direct matching between domain objects and presentation is proposed. Unfortunately, the heuristics to generate the presentation are not described. As for Executable UML, this approach is not aimed at manipulating metamodels, but at supporting whole application user interface generation. In Naked Objects, the presentation cannot be customized; therefore it cannot be adapted to different user skills or preferences. There is no guidance regarding the order the tasks should be carried out, the target platform is desktop applications and it is not designed for novice users.

There are also other similar approaches from the human-computer interaction community. Model-based User Interface Development Environments (Mb-UIDE) [27] provide a mechanism to design the UI by means of a number of declarative models, which are latter translated into code directly executable on a specific platform or into an intermediate language (usually XML-based). Mb-UIDE has been in use since the beginning of the 90s and it is becoming increasingly integrated into the MDD approach [34].

In Mb-UIDE, the domain model represents the information required by the user to carry out the tasks through the UI. To express these models, different notations have been used, but undoubtedly, the most commonly used are entity-relationship notation and UML class diagrams. Some of the MB-UIDEs using class diagrams are OVID [29], Janus [2], AME [17], Teallach [4], OO-H [7] and IdealXML [21]. Two of the Mb-UIDEs that use entity-relationship notation for domain modeling are Trident [5] and Genius [14]. Just-UI [20] is a MB-UIDE that provides a set of patterns to generate a UI out of a domain model for standard desktop applications. Although these approaches are not aimed at generating UI for the manipulation/edition of a domain model, they provide interesting insights into what should be modeled to automatically generate a UI out of models.

We took inspiration from these approaches to identify the key features that should be modeled to generate a UI, including the requirements to have some extra models apart from the domain model to generate usable UIs to manipulate the models. These extra models enable the modeling of, for instance, the aesthetics of the user interface to be generated. These extra models are described in depth in section 4.

Since our goal is to be able to generate a UI for the manipulation of domain models by novice users, with no experience in using software design tools, the proposal used should not be the one used in most CASE tools. Therefore, some research was carried out on the different interaction techniques available. This study showed that multi-touch user interfaces were the most intuitive ones for novice users, such as ABI specialists [22]. The most important reason why ABI specialists chose multi-touch interaction was because they felt it was natural, and also because they were used to the techniques involved in multi-touch smart phones. Nowadays, multi-touch interaction and gestures are being widely used in many kinds of portable and fixed devices [23].

The concept of interaction style refers to the different ways the user can communicate or otherwise interact with an artifact. There are different interaction styles, for instance: command language, form filling, menu selection or direct manipulation. In our context, direct manipulation was chosen because it offers several advantages such as [30]: visually presenting task concepts, it is easy to learn, errors can be avoided more easily and recognition memory, as opposed to cued or free recall memory, can be emphasized. There are currently many electronic devices in which touch and motion-based gestures are supported and there are de facto standards related to this type of interaction, such as the Apple Human Interface guidelines [1]. These guidelines have been used for the definition of the heuristics presented in Section 5.

4 Metamodels Infrastructure

Several approaches have emerged to date describing the guidelines to execute the MDD paradigm. Perhaps, the most commonly known is the Model-Driven Architecture (MDA, [18]), an initiative of the Object Management Group (OMG). MDA promotes the separation of the domain model from the underlying technology to facilitate higher flexibility while designing and evolving software systems. One of the key elements of the MDA initiative is the Meta-Object Facility (MOF, [25]), a four-layer architecture used for the definition of the metamodels and models involved in the software development process, as shown on the right part of Fig. 1.

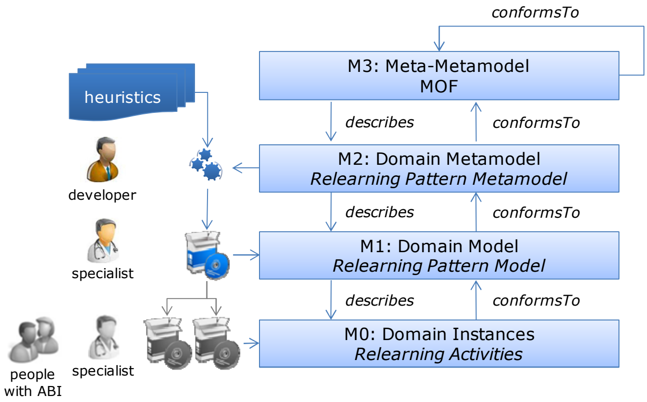


Fig. 1. Stakeholders, software products and MOF architecture in our approach

Fig. 1 shows the workflow of our proposal. The key element is the *instantiation process*: metamodels (M2) are instances of Ecore (M3), models (M1) are instances of metamodels, and domain instances (M0) are instances of models (this level is shown in gray to indicate it is out of the scope of this paper due to space limitations). In this instantiation process each stakeholder, specifically developers and ABI specialists, are focused on their interests and abilities. Developers are responsible for the specification of metamodels and the development of a flexible tool for relearning activity patterns. However, ABI specialists are focused on the definition of relearning activity patterns by using the tool with a high level of usability created by the developers. The UI of this tool was generated automatically by the developers, using the set of heuristics that are presented in Section 5. These heuristics were automated by using XPAND [33], a specialized language for code generation out of Ecore models. Moreover, the editor model was automatically created thanks to the capabilities offered by the Eclipse Modeling Framework (EMF, [11]).

As Fig. 1 shows, the MOF architecture has been used to describe both the domain metamodels (M2) and the models for the systems under development (M1). The domain metamodels are used to describe the core processes and domain concepts that are to be used by developers to convert design into code. According to our case study, the domain metamodel is the relearning pattern metamodel. This architecture has been used as follows:

- *Meta-metamodel (M3) level*. It offers a collection of primitives to define metamodels at level M2, that is, it is a meta-metamodel to describe metamodels. In this proposal, Ecore from EMF was used as meta-metamodel.
- *Metamodel (M2) level*. The elements in this metamodel are used to describe the elements of the model at level M1. ABI specialists demand software for the creation of relearning activities for people with ABI that is customizable enough. Although, these activities are known by specialists, they unfortunately cannot implement them. HABITAT provides a computer-based tool for relearning activity patterns specification for ABI specialists. With this aim, in this M2 level, developers created the domain metamodels that were used jointly with the heuristics presented in Section 5 to create a multi-touch tool in an automatic way. This tool is used by ABI specialists to manipulate the domain model, that is, the model for specifying relearning activity patterns.
- *Model (M1) level* is defined by instantiating the M2 metamodel, which is used to define the relearning patterns that enable the specialists to create relearning activities. ABI specialists work, at this level, with the computer–based tool developed by the developers. ABI specialists create the relearning activity patterns to provide tools for people with ABI. At this level, concrete interaction styles, deficits and resources are specified.
- *Instances (M0) level*. At this level, the instances of the domain model are created. In our case, the relearning activities are defined by the specialists and used by the people with ABI, as instances of the M1 model. ABI specialist instantiate and create different tools for people with ABI where different deficits and interaction possibilities are considered. People with ABI use specific software designed by ABI specialists for their treatment.

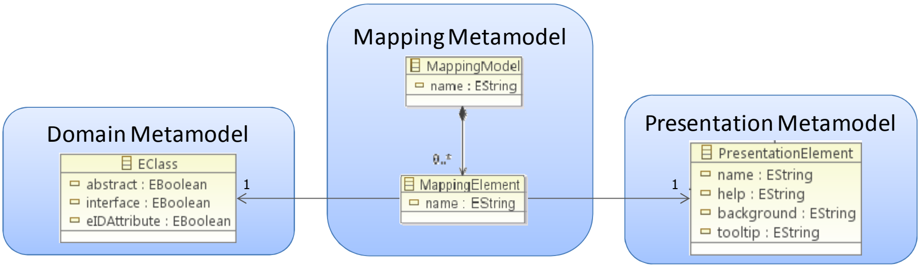


Fig. 2. Metamodels Infrastructure

Although it is not shown in Fig. 1, at M2 level, developers identify and specify three metamodels: domain, mapping and presentation. As it can be observed in Fig. 2, these metamodels are generic, so that they can be used for any general purpose domain. Next, all these three metamodels are described:

- *Domain metamodel*. This metamodel supports the specification of the facilities and services required for the target domain. As it can be observed, the main element of *Ecore*, *EClass*, has been used to facilitate that any domain metamodel defined as instance of *Ecore* can be used in this proposal. In our case study, developers are aimed at providing an environment for ABI specialists supporting the specification of any number of relearning activities patterns. Later, in Section 4.1 we present the HABITAT relearning activity pattern metamodel as an example of a concrete domain metamodel that can be used.
- *Presentation metamodel*. This metamodel is used to support the specification of the presentation details for each domain metamodel element. Section 4.2 describes the presentation metamodel developed in this proposal.
- *Mapping metamodel*. This metamodel is a mediator entity between domain and presentation metamodels. It relates both metamodels supporting loose coupling between the domain metamodel and the presentation metamodel.

4.1 Domain Metamodel: HABITAT Relearning Pattern Metamodel

At the metamodel level (M2), developers defined a specification useful for relearning activity patterns specification. These patterns constitute the domain to be defined by our novice users, ABI specialists. At the beginning, many ABI specialists, for instance in the ADACE association [22], documented the relearning activities using cards. These relearning activity descriptions had different elements or sections that were analyzed and abstracted away to create the metamodel shown in Fig. 3 (for the sake of clarity, class attributes are not shown in the figure). Purposes and descriptions of the elements of relearning activity patterns are the following:

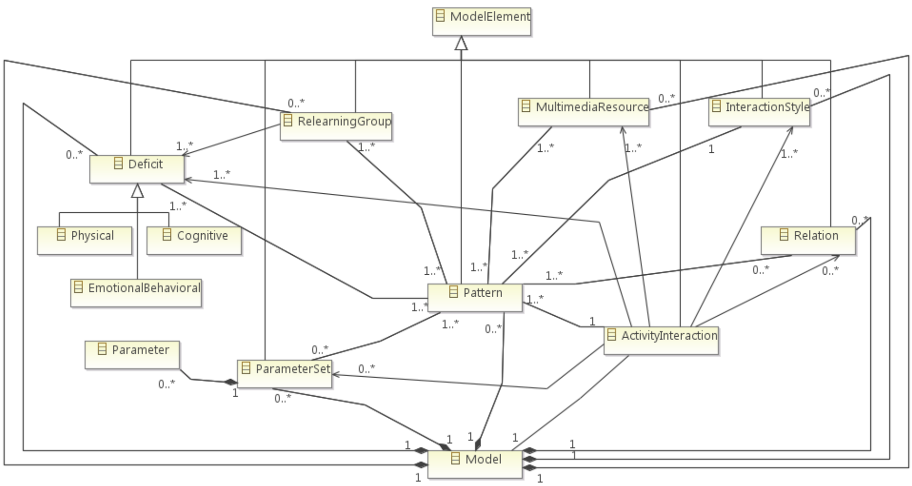


Fig. 3. Relearning pattern metamodel

- *Model*. This is a general-purpose element for organizing the domain metamodel. The definition of this kind of element is a constraint imposed by *Ecore*.
- *ModelElement.* This element was defined to abstract away some attributes that are shared by several metamodel elements.
- *Pattern*. This element represents a relearning activity pattern at a conceptual level, that is, concrete relearning activity patterns will be instances of this element. It represents an abstraction of a type of treatment or activity for people with ABI. It will be instantiated by ABI specialists.
- *Relation*. This element represents the relationships between relearning activity patterns. Few patterns live in isolation. Typically, they introduce new, hopefully smaller and more tractable activities which will lead you to other relearning activities. Or, there may be other patterns that treat the same set of deficits.
- *Parameter*. This element is used to represent the information of each pattern. Instances of this element that ABI specialists can specify are related to evaluation criteria, timing, etc.
- *ParameterSet*. This element represents sets of parameters with a common purpose.
- *Deficit*. Each relearning activity pattern is addressed to treat a specific set of deficits (e.g.: physical, cognitive and/or emotional/behavioral).
- *RelearningGroup*. In order to facilitate the treatment process, different groups were defined by ADACE, each one being characterized by a set of specific deficits.
- *MultimediaResource*. This element is used to associate multimedia resources to relearning activity patterns.
- *InteractionStyle*. This element represents the interaction style that the ABI specialists wants each person with ABI to use when he/she is doing a relearning activity.
- *ActivityInteraction*. This element represents the kind or type of activity that ABI specialist wants to associate with a relearning activity pattern. For instance, association or puzzle activities are instances of this element.

By using the previous elements developers provide ABI specialists with a software tool. Then specialists can document relearning activity patterns. Customization and personalization of this software tool can be achieved by the presentation metamodel that will be described in the next section.

4.2 Presentation Metamodel

This metamodel was defined by developers to achieve flexibility in the look and feel of the computer-based tool provided to the ABI specialists. Each domain model will be related to at least one presentation model, that is, an instance of the presentation metamodel illustrated in Fig. 4.

The presentation metamodel elements are described as follows:

- *PresentationModel*: This is a general-purpose element for organizing the presentation metamodel. It has a similar purpose to Model in the domain metamodel (see Fig. 3).
- *PresentationElement*. Each time a domain element is instantiated, a *PresentationElement* will be created to allow the customization of its presentation.

As can be observed in Fig. 4, it helps in describing different parameters, such as background color, tooltip, or label.

- *Font*. This element is used to allow the font customization of domain elements. It is related to alignment, size, etc.
- *Icon*. This element was included to associate each domain element with an icon.
- *Border*. This element is used to support border customization in domain elements.

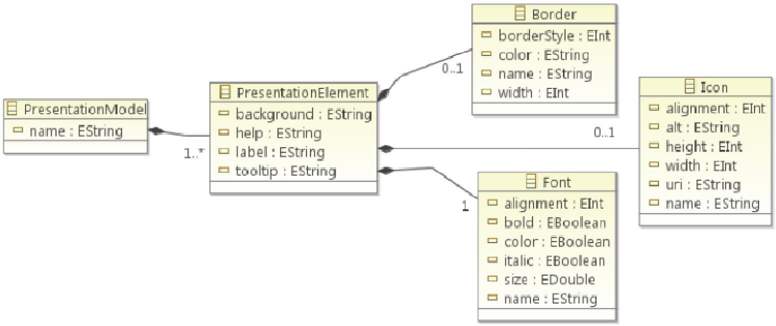


Fig. 4. Presentation Metamodel

Once domain and presentation metamodels were introduced we will describe in the following section the heuristics identified for generating multi-touch UIs that allow novice users to manipulate models.

5 Heuristics for Model-Driven Multi-touch Interaction

Heuristics are experience-based techniques for problem solving, learning, and discovery. These methods are used to speed up the process of finding a satisfactory solution, where an exhaustive search is impractical or conditioned in a certain way. Examples of this method include using a "rule of thumb" or common sense. When these heuristics are repeatedly used, and thoroughly tested, they can become patterns. We were able to identify heuristics by applying our previous experience in similar multi-touch interaction UIs developed in collaboration with ADACE. These UIs pursued hiding the complexity usually found when novice users have to manipulate metamodels in a MDD environment.

As shown in Fig. 3, after the developer has specified the domain metamodel, the tool that will be used by the novice users is generated. This tool is generated automatically by means of EMF and XPAND frameworks, and the heuristics proposed in this paper.

The heuristics gathered in this section proved useful for UI development out of a set of models. These UIs were suggested and evaluated by ABI specialists, who were our end-users. Multi-touch and direct manipulation were the interaction styles chosen to support the application. The inputs for the definition of these heuristics were the specialist's requirements. Our activities were driven by the metamodeling principles underlying *Ecore* metamodel.

One of the main elements considered by the heuristics presented in this work is the semantics of the relationships that can established in Ecore. To make easier understanding the heuristics proposed, the semantics of UML relationships [6], and how they can be expressed in Ecore, is described:

- *Association*: An association relationship is a structural relationship between two model elements that shows that objects of one classifier connect and can navigate to objects of another classifier. Even in bidirectional relationships, an association connects two classifiers, the primary (supplier) and secondary (client). This relationship is specified in Ecore by establishing an *EReference* between two *EClasses*. If a bidirectional relationship is needed, then two *EReferences* have to be created, one for each direction, and their *EOpposite* attribute must be set to specify that one is opposite of the other one.
- *Composition*: a composition relationship represents a whole–part relationship, and it is a specific type of aggregation, which is another type of UML relationship. An aggregation relationship depicts a classifier as a part of, or subordinate to, another classifier. A composition relationship specifies that the lifetime of the part classifier is dependent on the lifetime of the whole classifier. To specify this type of relationship in Ecore, an *EReference* between two *EClasses* is created. Besides, its attribute *containment* must be set to true.
- *Generalization*: A generalization relationship denotes that a specialized (child) model element is based on a general (parent) model element. Although the parent model element can have one or more children, and any child model element can have one or more parents, typically is the case that a single parent has multiple children. This relationship type is described in Ecore by setting the attribute *ESuperType* of the child *EClass* to its parent *EClass*.

Other important attributes used while describing Association and Compo*sition* relationships in Ecore are *LowerBound* and *UpperBound.* They are used to describe the cardinality, that is, how many instances from one of the entities is related to each instance in the other entity involved in the relationship.

While the ABI domain metamodel was being developed, and UIs were being discussed with ABI specialists, we identified empirically a set of heuristics correlating the semantics and structure found in the metamodel and the user interaction to manipulate it. These heuristics were gathered in specific scenarios, but we found that they can be reused again and again in other scenarios. In our proposal, multi-touch and motion-based gestures are used to design the following heuristics:

- *Heuristic #1* – [Root]: all the derivation of the user interface starts from the root of the model as this is a constraint imposed by *Ecore*. For instance, *Model* was the selected root for the Relearning pattern metamodel shown in Fig. 2.
- *Heuristic #2* – [Recursion]*:* as the user browses the elements of the domain model to create and manipulate an instance, all the heuristics are applied recursively. That is, the root where all the heuristics are being applied is the current element the user is browsing.
- *Heuristic #3* – [Association (2-directional) - Drag&Drop]. A bidirectional association relationship is used, for example, to specify that a *pattern* can be related to several *relearning groups*, and that a *relearning group* can be related to several *patterns* (see Fig. 3).

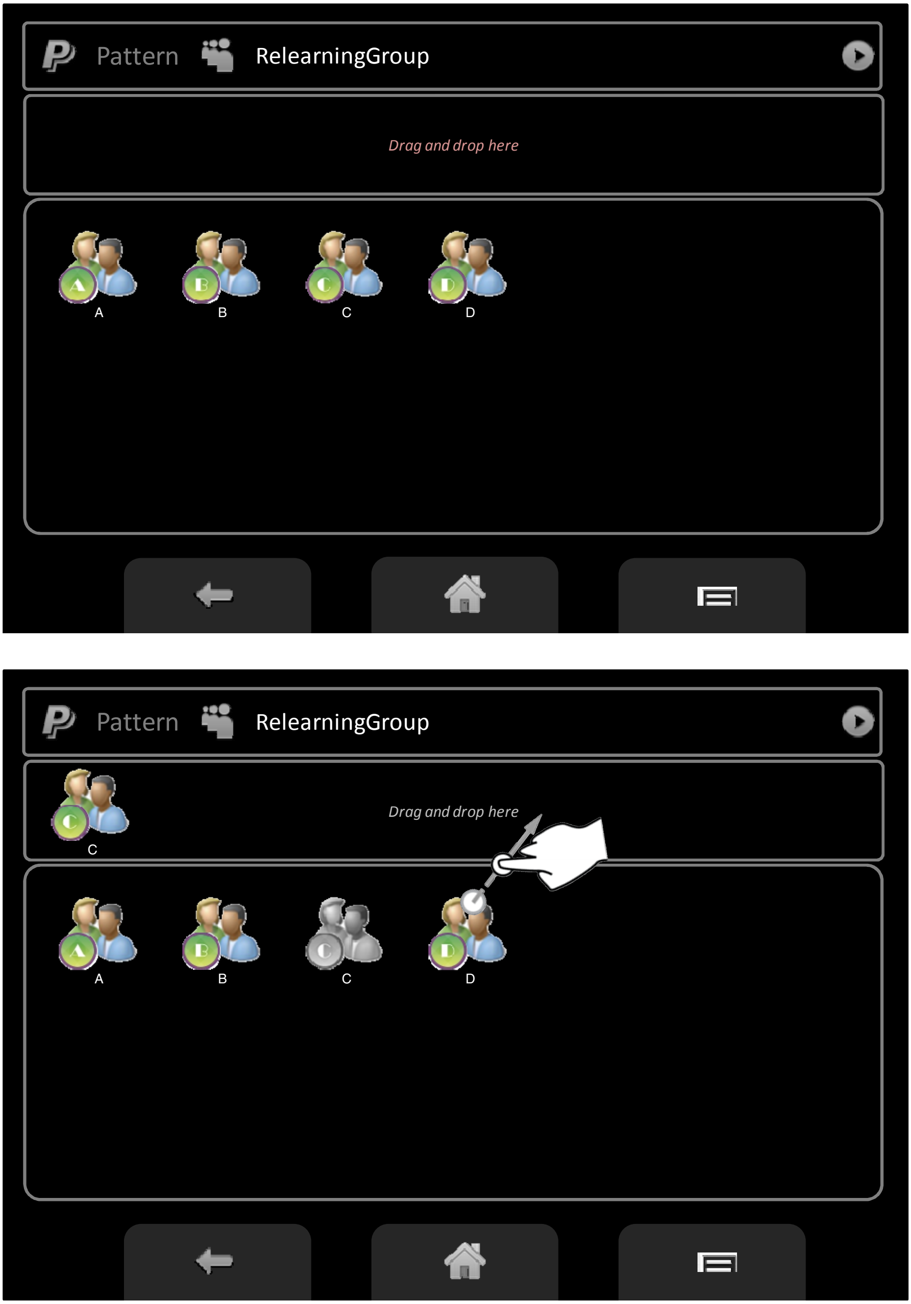


Fig. 5. Generated UIs and gesture according to Heuristic #3

Drag&Drop is the interaction style to be used whenever an association of the domain model has to be manipulated. As shown in Fig. 5, by dragging one instance from the lower part, and dropping it into the upper part, the user can easily create the relationship between a *pattern* and a *relearning group*. To provide more guidance to the user, and prevent errors, feedback related to the number of elements that the user can or must drag is provided. If the user can relate or not the elements (optional relationship), then a suggestion written in grey text is shown in the upper part. On the other hand, if it is mandatory for the user to create the relationship (mandatory relationship) then this text is written in red. Note that depending on the path followed to reach an element in the metamodel, this heuristic represents one direction or the other of the bidirectional relationship. Always the direction represented in the UI is the one whose origin is the current element the user is manipulating.

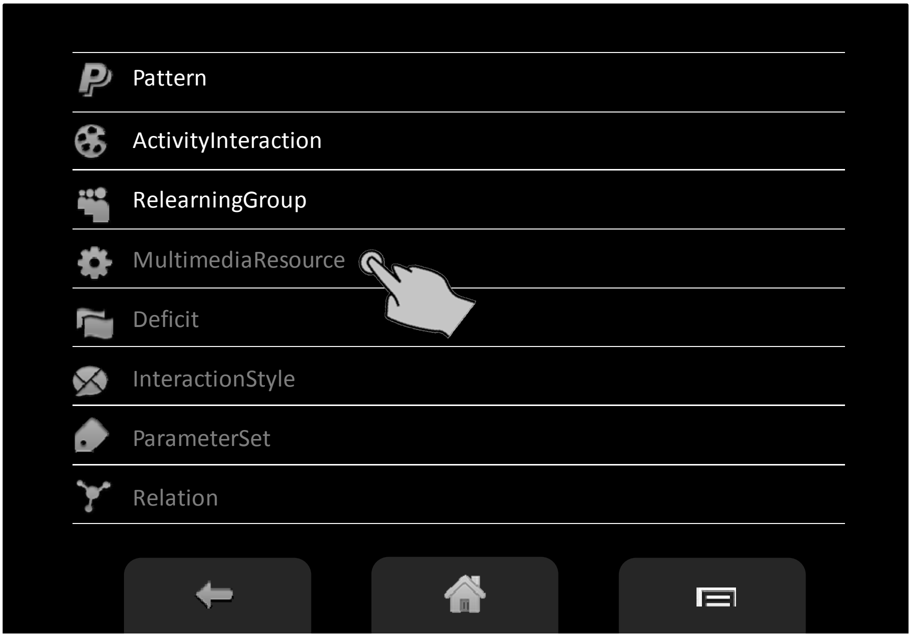


Fig. 6. Generated UIs and gestures for Heuristic #4

- *Heuristic #4* – [Association (1-directional) – Disable]. This association is used, for instance, to describe the unidirectional relationship between *activity interaction* and *multimedia resources* (see Fig. 3). As Fig. 6 shows, to specify a *multimedia resource* for a *relearning pattern*, an *activity interaction* must be previously defined. That is, unidirectional relationships describe dependencies in the order that instances from the metamodel can be created. *Disable* is the interaction used to reflect the dependencies created by unidirectional relationships, those elements having an incoming unidirectional relationship will be disabled in the user interfaces until the dependency is fulfilled. In the example of Fig. 6, once an activity interaction is defined, *multimedia resources* can be created. Once at least a multimedia resource has been created a UI similar to that presented in Fig. 5 will be offered to the user to relate it to an activity interaction.

- *Heuristic #5* – [Generalization – Tap]. In the example, different kinds of ABI *deficits* are shown, such as cognitive, physic or emotional/behavioral. In the UI generated from the metamodel, the user should be able to navigate through the types of *deficits* and to specify special deficits for each *relearning pattern*. *Tap* is the gesture used to navigate the generalization hierarchies. In our example, the user navigates the *deficits* types of our metamodel. Once a particular *deficit* type is selected (see Fig. 7), the user can manipulate its instances.

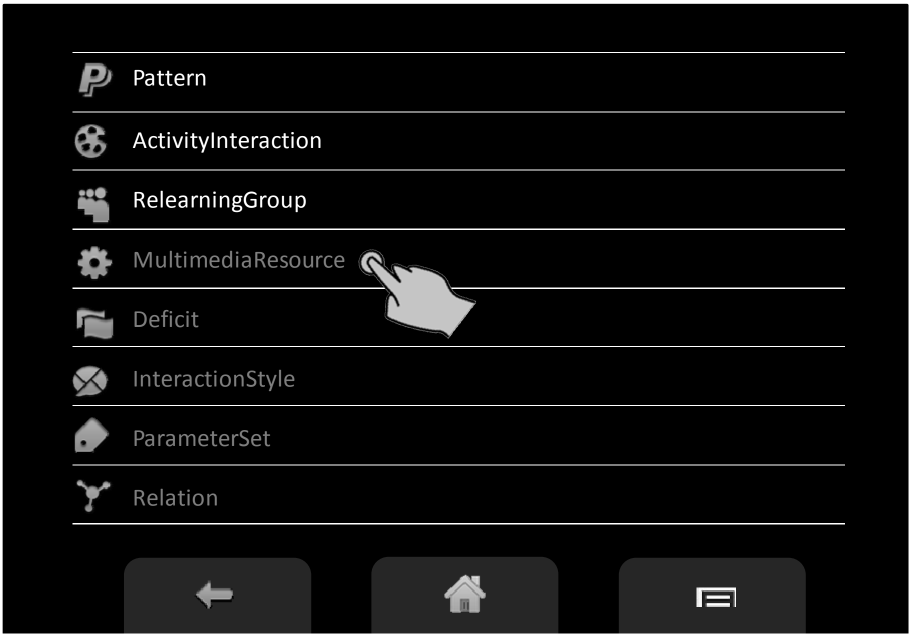


Fig. 7. Generated UIs and gestures for Heuristic #5

- *Heuristic #6* – [Composition – Press]. This relationship is used, for instance, in the ABI domain metamodel to specify that *model* is composed of 0 or many *ActivityInteraction*. This means that the user should be supported in creating, editing or deleting *activity interactions* instances. *Press* is the interaction style to be used for the manipulation of these relationships. A press is a touch in a surface for an extended period of time that is used to select the element to be modified, and then it can be drag to the bottom side of the UI for its edition or deletion. Fig. 8 shows how an *activity interaction* is manipulated. If a new instance has to be created, then the user presses on the free area of the UI so that a UI to specify the new instance is shown by the system. For instance, when the user presses on the free area of the Fig. 8 then the UI depicted in Fig. 9 is shown to specify a new *activity interaction*.

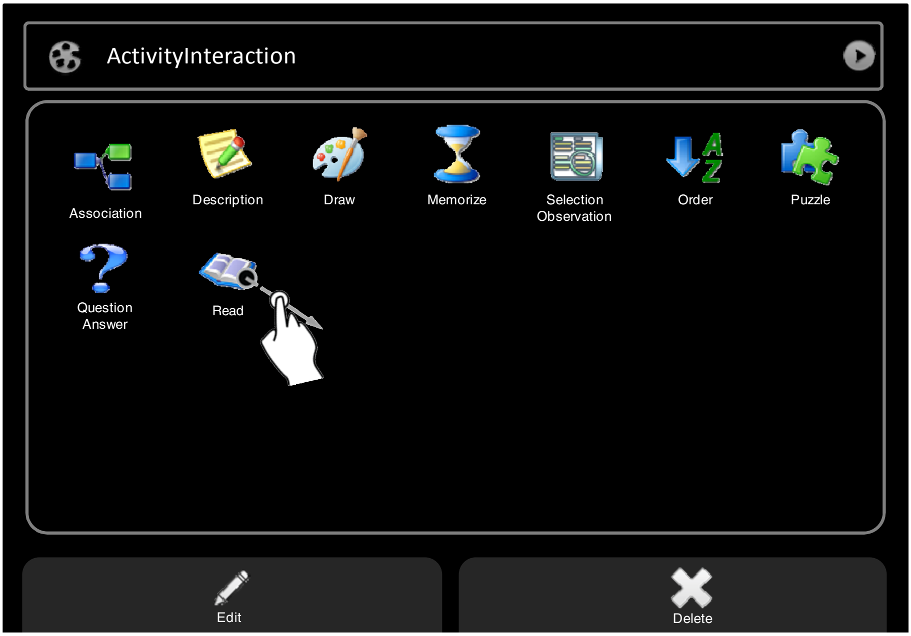


Fig. 8. Generated UIs and gestures for Heuristic #6 while editing or deleting an *activity interaction*

Fig. 9. Generated UIs and gestures for Heuristic #6 while creating an *activity interaction*

It is worth noting that although the ABI domain has been used as guiding example for the explanation of the proposed heuristics, they are totally independent of any domain. No information or relation to the domain has been considered while they were identified and automated by means of XPAND, so that they can be reused in different contexts by only feeding the generation tool with new metamodels. In addition, this proposal has a direct impact on productivity, as most MDD proposals do. This means that as the domain metamodel evolves, the multi-touch UI changes in an automatic as well without coding effort.

6 Exploratory Evaluation

Both the heuristics and the metamodel infrastructure were validated and refined during the development of the HABITAT project by ABI specialists. This project aims at helping people with ABI. ABI patients have brain injuries caused by damage to the brain after birth that can involve the loss of cognitive, physical and/or emotional capabilities. The relearning process to improve their quality of life and help them in recovering their lost capabilities is a long and hard road.

This relearning process should be flexible since every single patient has specific needs to which the relearning process should be tailored as far as possible. This personalization should be achieved by using the specialists' normal vocabulary so that both the knowledge and the experience gathered during the relearning process can be reused. The issue of common vocabulary has already been addressed in previous studies [22].

The exploratory evaluation was conducted in collaboration with three ABI specialists. An iterative process was used for refining the heuristics on the basis of the feedback collected from the subjects. First, we met with the specialist to detect their needs in terms of modelling concepts. After several meetings, a metamodel was created that gathers the knowledge about the problem that is used to support the ABI specialists in creating their own relearning activity patterns to help during the relearning process of people with ABI. Second, a first version of the heuristics was defined, and they were used to create the first prototype that was tested by the ABI specialist. After several iterations the final version of the heuristics presented in this paper was produced. In each iteration, the cognitive walkthrough usability testing technique was used to detect possible usability issues in the prototype created by applying the heuristics.

Then, the engine to automate the generation of the user interface by applying the heuristics was developed. This engine was used to generate the multi-touch user interfaces for the ABI metamodel. This tool was used by the ABI specialists to instantiate the metamodel to create new relearning patterns.

The interviews with the ABI specialists revealed positive results, since all the specialists told that they were able to create the intended relearning pattern by using the generated user interface. Nevertheless, a thorough evaluation is required to fully validate all the heuristics, involving subjects from different domains.

7 Conclusions and Further Work

Model-Driven Development is a powerful tool to develop software. Nevertheless, some of the models involved in a model-driven development are actually dependent on a specific domain, where specific experts are actually the ones that should manipulate them. Nevertheless, usually domain experts are novice users. By providing a multi-touch based interaction, fully driven by the underlying metamodel, we are supporting novice users in the manipulation of metamodels. Our experience with ABI specialist have thrown successful results, supporting the specialist in creating relearning patterns [22] out of UI automatically derived from the domain metamodel.

In this paper a metamodels infrastructure and a set of heuristics are proposed. The heuristics drive the generation of multi-touch user interfaces that support the manipulation of metamodels to manipulate domain models. These heuristics exploit the semantics and structure of any metamodel specified by using *Ecore* to generate these multi-touch user interfaces.

The multi-touch user interfaces are generated so novice users can create models from metamodels, bridging the gap between the developer's language and what novice users understand. Furthermore, a metamodel infrastructure is also introduced to support the customization of the presentation of the different metamodel elements. Thus, icons or other aesthetics features can be specified to make the generated user interfaces more attractive to the user. Generating automatically a multi-touch user interface from metamodels also helps in improving certain basic usability principles, such as: visual presentation of task concepts, easy to learn, error prevention (by supporting the user in performing only valid actions), consistency (since the same tasks, such as adding items to an aggregation, is always carried out in the same way) or presentation structuring (by using the relationships in the metamodel to group related concepts).

As further work, we are applying these heuristics to other domains to gain further experience and improve our current tools for the generation of multi-touch user interfaces from domain metamodels.

Another future work is related to the evaluation of the proposal. Although the exploratory evaluation threw positive results, a more thorough evaluation is required to fully validate the heuristics presented. Currently several experiments are being designed, involving users from different domains.

Acknowledgements. We would like to thank the ADACE Association (http://www.adaceclm.org) for their cooperation in our research on ABI. This work has been partially supported by the grant insPIre (TIN2012-34003) from the Spanish Government Department of Science and Innovation and the grant AVANZA TSI-020400-2011-20 from the Ministry of Industry, Tourism and Commerce of Spain.

References

1. Apple. iOS Human Interface Guidelines (p. 156), http://developer.apple.com/library/ios/documentation/userexperience/conceptual/mobilehig/MobileHIG.pdf (retrieved) (last access February 05, 2011)
2. Balzert, H.: From OOA to GUI – The JANUS-System Human-Computer Interaction. In: IFIP TC13 Int. Conf. on Human-Computer Interaction (INTERACT 1995), Lillehammer, Norway, June 27-29, pp. 319–324 (1995)
3. Balzert, H., Hofmann, F., Kruschinski, V., Niemann, C.: The JANUS Application Development Environment - Generating More than the User Interface. In: 2nd Int. Workshop on Computer-Aided Design of User Interfaces (CADUI 1996), Namur, Belgium, June 5-7, pp. 183–208 (1996)
4. Barclay, P.J., Griffiths, T., McKirdy, J., Paton, N.W., Cooper, R., Kenne, J.: The Teallach Tool: Using Models for Flexible User Interface Design. In: 3rd Int. Conf. of Computer-Aided Design of User Interfaces (CADUI 1999), Louvain-la-Neuve, Belgium, October 21-23, pp. 139–158 (1999)
5. Bodart, F., Hennebert, A.M., Leheureux, J.M., Vanderdonckt: Computer-Aided Window Identification in TRIDENT. In: IFIP TC13 Int. Conf. on Human-Computer Interaction (INTERACT 1995), Lillehammer, Norway, June 27-29, pp. 331–336 (1995)
6. Booch, G., Rumbaugh, J., Jacobson, I.: Unified Modeling Language User Guide. Addison Wesley (1999)
7. Cachero, C.: OO-H: Una expresión a los métodos OO para el modelado y generación automática de interfaces hipermediales. PhD. Alicante University (2003)
8. Cicerone, K.D., Dahlberg, C., Kalmar, K., Langenbahn, D.M., Malec, J.M., Bergquist, T.F., Felicetti, T., Giacino, J.T., Harley, J.P., Harrington, D.E., Herzog, J., Kneipp, S., Laatsch, L., Morse, P.A.: Evidence-based cognitive relearning: Recommendations for clinical practice. Archives of Physical Medicine and Relearning 81(12), 1596–1615 (2000)
9. Ciocca, G., Olivo, P., Schettini, R.: Browsing museum image collections on a multi-touch table. Information Systems 37(2), 169–182 (2012)
10. Consejería de Sanidad de la Junta de Comunidades de Castilla-La Mancha. Conjunto mínimo de datos básicos de la Región de Castilla La Mancha (2006)
11. EMF, http://www.eclipse.org/modeling/emf/ (last access October 05, 2011)
12. GMF, http://www.eclipse.org/modeling/gmf/ (last access October 05, 2011)
13. Hailpern, B., Tarr, P.: Model-driven development: The good, the bad, and the ugly. IBM Systems Journal 45(3), 451–461 (2006)
14. Janssen, C., Weisbecker, A., Ziegler, J.: Generating User Interfaces from Data Models and Dialogue Net Specification. In: INTERACT 1993 and CHI 1993, Amsterdam, The Netherlands, April 24-29, pp. 418–423 (1993)
15. López Jaquero, V., Montero, F., Molina, J.P., González, P., Fernández Caballero, A.: A Seamless Development Process of Adaptive User Interfaces Explicitly Based on Usability Properties. In: 9th Conf. on Engineering for Human-Computer Interaction, EHCI-DSVIS 2004, Hamburg, July 11-13 (2004)
16. Luebbe, A., Weske, M.: Tangible Media in Process Modeling – A Controlled Experiment. In: Mouratidis, H., Rolland, C. (eds.) CAiSE 2011. LNCS, vol. 6741, pp. 283–298. Springer, Heidelberg (2011)
17. Märtin, C.: Software Life Cycle Automation for Interactive Applications: The AME Desing Environment. In: 2nd Int. Work. on Computer-Aided Design of User Interfaces (CADUI 1996), Namur, Belgium, June 5-7, pp. 57–76 (1996)

18. Miller, J., Mukerji, J.: MDA Guide Version 1.0.1, Object Management Group, Needham, Massachusetts (June 2003)
19. Milisev, D.: Model Driven Development with Executable UML. Wiley Publishing, Inc., Indianapolis (2009)
20. Molina, P.J., Meliá, S., Pastor, Ó.: User Interface Conceptual Patterns. In: Forbrig, P., Limbourg, Q., Urban, B., Vanderdonckt, J. (eds.) DSV-IS 2002. LNCS, vol. 2545, pp. 159–172. Springer, Heidelberg (2002)
21. Montero, F., López Jaquero, V.: IdealXML: An Interaction Design Tool and a Task-Based Approach to User Interface Design. In: 6th Int. Conference on Computer-Aided Design of User Interfaces (CADUI 2006), Bucharest, Romania, pp. 6–8 (June 2006)
22. Montero, F., López-Jaquero, V., Navarro, E., Sánchez, E.: Computer-aided relearning activity patterns for people with acquired brain injury. Computers & Education 57(1), 1149–1159 (2011)
23. Müller-Tomfelde, C. (ed.): Tabletops - Horizontal Interactive Displays. Springer, London (2010)
24. Navarro, E., López-Jaquero, V., Montero, F.: HABITAT: A Web Supported Treatment for Adquired Brain Injured. In: IEEE Int. Conference in Advanced Learning Technologies (ICALT 2008), Santander, July 1-5, pp. 464–466 (2008)
25. OMG, Meta-Object Facility (MOF), ptc/03-10-04 (MOF 2.0 Core Final Adopted Specification), `http://www.omg.org/mof/` (last visit 2011)
26. Pawson, R., Matthews, R.: Naked Objects. In: Companion of the 17th Annual ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications (OOPSLA 2002), pp. 36–37. ACM, New York (2002)
27. Puerta, A.R.: A model-based interface development environment. IEEE Software 14(4), 40–47 (1997)
28. Reichart, D., Forbrig, P., Dittmar, A.: Task models as basis for requirements engineering and software execution. In: 3rd Int. Work. on Task Models and Diagrams for User Interface Design (TAMODIA 2004), Prague, Czech Republic, November 15-16, pp. 51–58 (2004)
29. Roberts, D., Berry, D., Isensee, S., Mullaly, J.: Designing for the User with OVID: Bridging User Interface Design and Software Engineering. New Riders Publishing (1998)
30. Schneiderman, B.: Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces. In: Proceedings of Intelligent User Interfaces, Orlando, USA (1997)
31. Selic, B.: The pragmatics of model-driven development. IEEE Software 20(5), 19–25 (2003)
32. Toronto Acquired Brain Injury Network, http://www.abinetwork.ca/definition.htm (last access May 02, 2011)
33. XPAND, http://wiki.eclipse.org/Xpand (last access May 02, 2011)
34. 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)

TactiPEd: Easy Prototyping of Tactile Patterns

Sabrina Panëels, Margarita Anastassova, and Lucie Brunet

CEA, LIST, Sensorial and Ambient Interfaces Laboratory
91191 - Gif-sur-Yvette CEDEX, France
{sabrina.paneels,margarita.anastassova,lucie.brunet}@cea.fr

Abstract. We present the design and evaluation of a tactile editor, TactiPEd for the rapid and easy prototyping of vibrotactile patterns. It is based on the graphical metaphor of the shape of the device, which is used for the tuning of the main tactile characteristics, including amplitude, frequency and duration of tactile sequences. The editor includes file systems functionalities using the XML format along with playing and recording functionalities. The editor was thoroughly evaluated: a usability evaluation was conducted with 9 participants, the designed metaphor-based patterns were analyzed for insights on cross-device design and finally the editor was tested with several devices. TactiPEd was successfully and easily used with little training and enabled users to design patterns in little time. The resulting patterns shared common characteristics across the devices for a given metaphor.

Keywords: tactile feedback, tactile pattern authoring, vibrotactile pattern.

1 Introduction

The development of wearable tactile devices, in the forms of bracelets [1, 2, 3], vests [4] or belts [5], has been increasing, in particular for mobile and wearable computing. By providing an alternative channel for communication, these devices enable the transmission of information in an eyes-free, ear-free and discrete manner, useful not only in contexts where other modalities are not available but also when they are heavily used, all the while letting the user focus on the surrounding environment. For instance in mobility, it frees the user from staring continuously at his smartphone screen for information whilst walking [6] and thus enables him to stay aware of the potential dangers incoming from traffic of cars and people. The haptic modality has successfully been used for example for providing navigation cues [4, 5, 6], status information about mobile phone applications (e.g. activity of email inbox [1]) or for enhancing the education and learning [7], among many other existing applications.

However, a lot of effort is still spent on designing appropriate, intuitive and discriminable vibrotactile cues, through numerous development iterations and user evaluations [8], often specific to one device and one application, and usually requiring specific device and programming knowledge. Therefore few research prototypes reach the market or the industry, though they can potentially benefit the areas of accessibility, inclusivity and user mobility. In fact, the general public is only confronted

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 228–245, 2013.

to basic haptic interaction (e.g. vibrations from mobile phones) whereas interface design has been giving increasing importance to user experience [9]. Solutions to integrate user experience in design frameworks are currently being investigated. One such solution lies in making the prototyping of interactions accessible to all, thus enabling a greater involvement of the user, for example using participatory design [10, 11]. This could potentially foster the creativity and intuitiveness [11].

Therefore, to ease the design and testing, and to help promote tactile research and make it more widely accessible, an interface is needed that enables the rapid prototyping and the easy authoring or tuning of patterns, not only by developers but also non-developers, such as ergonomists, designers and users, without the need for specific signal processing or signal communication knowledge. Such an interface should also support several types of devices including multi-actuator devices, i.e. a device with several actuators. Few such interfaces are currently available and widely used. In order to fill this gap, this paper proposes a novel tactile editor, which is based on using parameters that are common knowledge, namely activation, duration, amplitude and frequency as well as a graphical layout of the device for tuning these. The details of its design, implementation and evaluation are further described in the following sections after presenting existing interfaces.

2 Tactile Authoring and Prototyping

2.1 Direct Signal Representation

A few editors have been developed that are based on the direct representation of the vibration signal as editable waveforms. Enriquez and Maclean [12] developed the Hapticon Editor, for the creation and editing of haptic icons for a 1-DOF force-feedback device (e.g. a rotary knob). A pattern could be created by concatenating simple waveforms where length, frequency and amplitude were specified for each waveform and their shape modified through control points. Each icon could be subsequently played on the knob for testing. This work was extended and led to the Haptic Icon Prototyper [13], which uses a more graphical approach with a streamlined interaction sequence for composition of basic haptic icons (see left of Fig. 1).

Similarly, in the vibrotactile domain, Ryu and Choi [8] developed the posVibEditor [8] (see right of Fig. 1). It enables the prototyping of vibrotactile icons and can be used with multiple vibration motors. It also supports the creation of a library of patterns using XML-based pattern files. On the commercial side, Immersion developed the Haptic Studio (previously known as MOTIV Studio), a tactile editor for mobile platforms [14]. It provides templates of haptic effects (e.g. periodic effect) and a timeline interface for their combination into complex vibrotactile patterns. It also has a convenient feature of automatic vibrotactile pattern generation from music files in the MIDI format. It supports a variety of actuators (vibration motor, voice coil, Piezo element), i.e. mechanical devices that convert energy into motion and thus in these cases vibration, but it is unclear if it can be used for a device with multiple actuators.

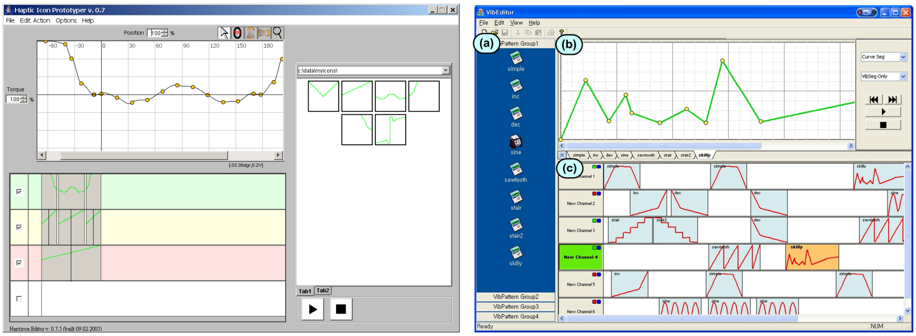


Fig. 1. On the left, the Haptic Icon Prototyper [13], with the waveform editor at the top, the streamlined sequence of signals at the bottom and the palette of haptic icons primitives on the right. On the right, the PosVibEditor [8], with the pattern manager (a) on the left of the image, the editor at the top (b) and the multichannel timeline interface (c)

These low-level interfaces enable finer control over the output due to a direct signal-based mapping. However, they are too low-level and as Lee and Choi [15] mention, they are far from being intuitive and time-efficient and rather ill-suited for non-experts who may lack the understanding required for creating the desired result. This is why another set of interfaces have been geared towards indirect representations or metaphor-based editing of tactile icons.

2.2 Indirect Representation – Metaphor Based

Audio – Musical Metaphor. Inspired by the similarity between sound and tactile vibration, Lee et al. aimed to provide a higher-level interface, the VibScoreEditor [16] (see left of Fig. 2), and therefore more widely accessible by using a musical metaphor. The vibrotactile score metaphor can represent the vibration pitch, by vertical location on the staff lines, the duration by the shape of the note and the strength by an integer inside the head of the note. The metaphor and authoring tool were thoroughly evaluated [15]: both programming and scripting by experts and waveform editing (using the PosVibEditor [8]) by non-experts were compared to using the tool. For all qualitative and quantitative measures, the vibrotactile score performed better than the other design methods, both for experts and non-experts. However, the drawbacks of the VibScoreEditor raised by the non-experts participants include the issue of users unaccustomed to musical notation as well as the difficulty to read strength and duration simultaneously. We believe these issues would be exacerbated with multiple actuators, not currently supported by the editor.

Jonas [17] used another graphical representation of music by drawing inspiration from music and video editing software with their timeline interface. In this regard, he developed a tactile editor with a track-based graphical user interface for non-experts, who can simply create and edit patterns by switching motors on and off for certain durations and at certain intensities (see right of Fig. 2). It also supports multiple actuators through Arduino and MakeController platforms. Participants positively evaluated the tool underlining that the timeline metaphor helped them design and test the vibrotactile patterns.

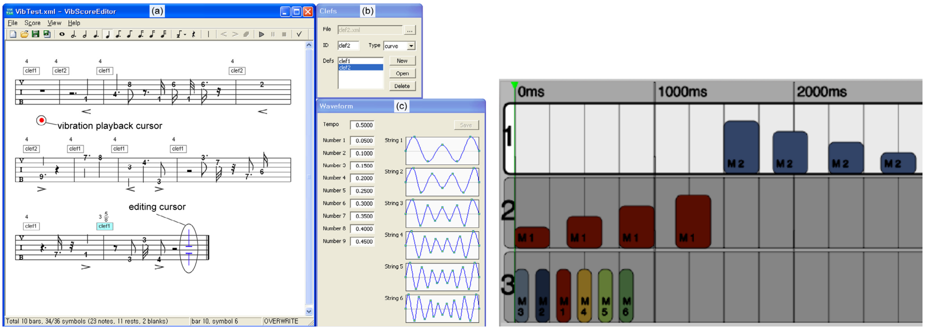


Fig. 2. Left: the VibScoreEditor [16] uses a musical metaphor for editing (main window) that is an abstraction of the physical parameters associated with a vibrotactile signal (right windows). Right: the Tactile Editor [17]

Another radically different and original approach to authoring haptic content is the TECHTILE toolkit [18], which aims to be a prototyping tool for easy and fun design and education of haptic media. It is based on a direct mapping from the audio signal to the vibration. Through a microphone, voice-coil actuators and a signal amplifier, users can easily record the sensation via audio and transfer it to a haptic sensation. First experience with haptics is surely improved, but fine-tuning still requires using audio editing software, which is not necessarily intuitive to all. Moreover, the audio channel is richer than haptics and mapping sounds to haptics is not straightforward, in particular for systems including multiple actuators. However, the transfer between audio and haptics is an obvious source of inspiration for designing haptics patterns; this is why in future work the possibility to transpose audio content to haptic patterns using TactiPEd will be investigated.

Touch Interaction. Given the advent of touchscreen interfaces, the possibility to program patterns directly by mapping touch properties to the vibration parameters has been explored. On the commercial side, the iPhone provides users with the possibility to record a custom-made vibration for contacts by simply tapping on the screen and recording the rhythm. Hong [19] investigated the mapping of user touch input to other vibration parameters to facilitate vibrotactile design. In his tool, duration of motion, touch pressure and the y-coordinate of the touch position are mapped to duration, amplitude and frequency respectively (see left of Fig. 3). This is an interactive and fun way to create patterns according to users; however, due to the low accuracy of touchscreens, gestures are hard to control and difficult for setting exact values. Furthermore, once recorded, a pattern is also difficult to edit as it requires re-recording of motion. This is why the future work proposes a hybrid method combining the 'Demonstration-Based' tool for a first recording and then a waveform editor for finer editing. Cuartielles et al. [20] proposed an editor where haptic patterns can be created by touching iconographic representations of the body on a touchscreen interface (see right of Fig. 3). Their previous editors used timelines to control each actuator but testing showed that editing the timeline for each haptic stimulus was time-consuming, in particular when designing high resolution haptic patterns (i.e. with a high number of actuators). We share the same point of view and focus rather on the shape of the device to resemble the output device [20] and to be independent of the position of the device on the body (e.g. some wristbands are developed to be placed on any position on the arm, such as the VibroTac [2]).

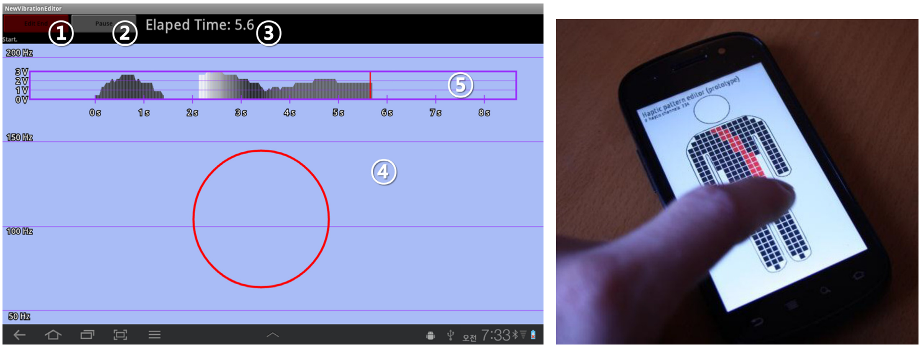


Fig. 3. On the left, tapping is mapped to frequency, amplitude and duration with the Demonstration-based editor [19]. Numbers 1 and 2 refer to editing buttons, 3 to elapsed time, 4 to the touch recording workspace and 5 to the display of the vibration signal graph. On the right, the Visual Editor for high resolution (i.e. high number of actuators) haptic patterns [20].

2.3 Summary

Waveform-based editors enable fine-tuning and precise editing and result. However, they require signal knowledge and are thus inaccessible to non-experts. This is why a number of approaches have investigated other representations, drawing inspiration from audio/music, touch or graphical interactions. However, these techniques usually restrict themselves to the time dimension of the vibration and omit the spatial distribution of the device. We believe this plays an important role in the design of patterns. For instance in navigation contexts, patterns often use the precise localisation of the actuators to form a pattern. A “right” message is often conveyed by a dynamic movement to the right or by simply activating the actuator located on the right of the device and the user. In other cases, some patterns are specific to a device layout and not easily transferrable to any device shape. For instance, for a diagonal pattern on a vest with eight actuators, its transfer to an 8-actuator bracelet around the wrist is not straightforward: what is a continuous diagonal in that case and is the meaning kept in the transfer? Moreover, few of these interfaces support both a variety of actuators and multiple actuators devices. A summary of the different metaphors and actuator support is depicted in Fig. 4.

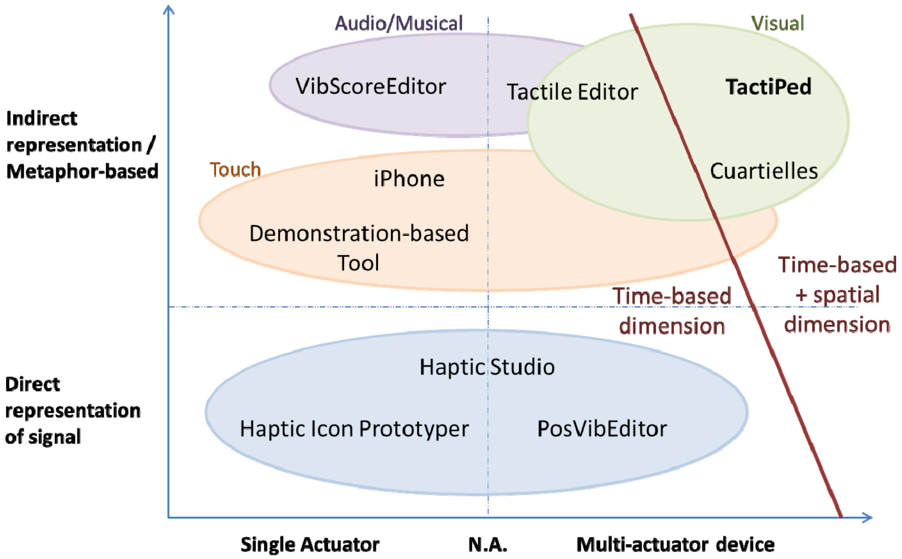


Fig. 4. Summary of the prototyping interfaces by their type of representation and the support for multiple actuators. N.A. refers to the case where the information is not available; e.g. for Haptic Studio the support for a multi-actuator device is unclear.

Therefore, we propose a tactile editor that uses the spatial layout of the device to set actuators activation along with other intuitive parameters such as duration, frequency and amplitude and that supports several types of actuators and multi-actuator devices.

3 TactiPEd: Tactile Pattern Editor

3.1 Overview

TactiPEd aims to provide a high-level interface accessible to non-technical experts (e.g. ergonomists and users) that facilitates the rapid prototyping and testing of tactile patterns. For that purpose, the editor is based on a graphical metaphor using the spatial layout of the device to tune its parameters. It also supports different types of actuators as well as multi-actuator devices so that pattern metaphors can be tested onto several devices. It has been implemented in Python and has been used on the Windows platform. The editor consists of four modules:

- *Graphical user interface*: for creating, editing, opening device templates; creating and editing patterns; and providing access to the other modules.
- *File handler*: for opening and saving device templates and pattern files via XML.
- *Communication module*: for interfacing the hardware, by mapping pattern values from the GUI to the hardware format, which is abstracted in the GUI through a defined device template.

- *Player/Recorder*: patterns can be tested on the fly and also recorded on the device if the hardware allows it.

Each module and its features are further detailed in the following subsections.

3.2 Graphical User Interface

The GUI provides the graphical view to create and edit the pattern and provides the user with access to the rest of the modules through a top bar menu (see Fig. 5).

Fig. 5. Top bar menu of the graphical user interface

The pattern creation and editing relies on using graphical items that can be modified (see Fig. 6 and Fig. 7) using mouse and keyboard interactions. The pattern is created by first opening a device template file (set by hardware constraints), which defines the number of sequences, the shape of the device and the parameters and their values (amplitude and frequency) and is the equivalent of a new file. The maximum duration of a sequence is currently set to 1.6 seconds as it was the maximum value for the devices tested; however it can easily be added as a parameter in the template. For each sequence, the parameters that can be modified are actuator activation, amplitude for each actuator or for all, duration and frequency.

The duration (see region 5 of Fig. 6) can be set by moving a slider, which gives a temporal indication through the length of the associated horizontal bars, similarly to a music equalizer. The actual value is displayed within the horizontal bar.

The amplitude can be either modified for each actuator if these are independent (see region 4 of Fig. 6) or for one sequence with the same value for all actuators (see region 2 of Fig. 7). In the first case, the amplitude values are mapped to a color gradient and set by either clicking on the corresponding actuator on the device shape until the desired value or for faster editing by right clicking the actuator (selection shown with a visual feedback) and setting directly the value via the keyboard. In the latter case, the amplitude is similar to a volume widget with a color gradient and the same mouse/keyboard interactions are used (see region 2 of Fig. 7). When the amplitude is defined per sequence, the activation of actuators is selected the same way as the amplitude per actuator, except with only two values: on and off.

The frequency is displayed within the duration slider to prevent a crowd of widgets using a bitmap with colored bars and white gaps (see region 1 of Fig. 7). The higher the frequency, the smaller the gap between the bars is, and inversely. The same interactions are used for setting the values: touch for successive increase and right click with keyboard value for direct setting.

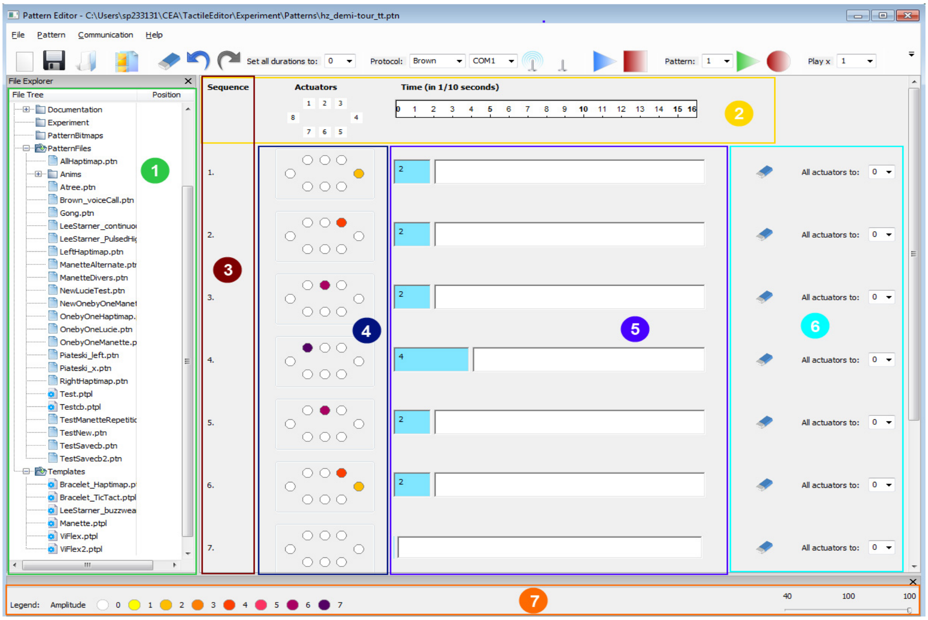


Fig. 6. Screenshot of the interface: Region 1 is a file explorer listing device templates and pattern files. Region 2 is a static ribbon explaining what each column is. The first column (region 3) is the sequence number. The second (region 4) represents the shape of the device where each actuator-circle is a clickable item that changes color according to its value. The next column (region 5) holds the duration bars while the last (region 6) holds widgets for editing (e.g. the eraser resets the sequence). Region 7 contains the legend and zoom.

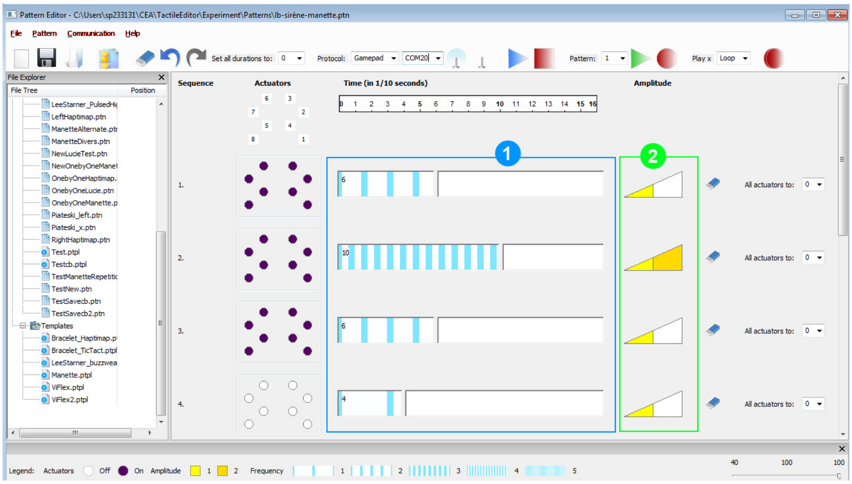


Fig. 7. Screenshot for a pattern tested on a custom built gamepad. Region 1 shows the duration bars displaying the frequency and region 2 shows the widget to set the amplitude per sequence.

The frequency and amplitude can also be copied and pasted by dragging and dropping the values from one sequence to another using the mouse. Finally, a legend is displayed at the bottom to show the matching between colors and values.

3.3 File Handler

The file handling module is based on the XML format for easy management of device templates and pattern files, including saving and opening them. Given the wide use of XML, the files could easily be used in other applications or extended to contain additional information. Small adjustments can also be made directly into the file, if preferred. The pattern file is divided into the definition of the device template (which is the same content as the template file) with the shape, the order of actuators and type of parameters, as described above, and the pattern parameters (activation, amplitude, duration, frequency and repetitions) for each sequence.

3.4 Communication Module and Internal Player/Recorder

The communication module establishes the connection to the device via a defined communication protocol and sends the data appropriately to the hardware. In the GUI (see Fig. 5), in order to test a pattern with the internal player, the user first needs to select a protocol and the port where the device is connected before attempting a connection. Each protocol in the list corresponds to a protocol file. Each protocol file follows a protocol template that implements a defined set of functions used by the GUI, namely *init* (for initialization and opening the connection), *closeSocket* (for closing the connection), *play* (for testing the displayed pattern), *stop*, *playRecordedPattern* (for playing a pattern recorded on the device) and *recordPattern* (to record a pattern at a certain memory position on the device). These functions are called upon pressing the corresponding button of the internal player (see Fig. 5).

The protocol list can also be updated 'on the fly' by adding a new protocol. The new protocol simply needs to be written in python by a device expert, implementing the functions cited above, and then added via the GUI. The complexity of the protocol is hidden to the novice end-user who simply uses a device template file to create and edit the pattern, as well as the connection and internal player buttons to test the patterns. The device template file reflects the data required by the protocol template and should also be created by someone knowledgeable. It provides information about the number of maximum sequences that can be sent to the hardware, the different parameters available (amplitude, frequency) and the number of values, and the shape of the device with the order of each actuator for data transmission.

4 Evaluation

The evaluation focused on assessing two main aspects: evaluating the usability of the interface with three user profiles (ergonomists, engineers and non-expert users), and gaining initial insights about the creation of patterns on different devices.

4.1 Methodology

Participants. Nine participants (5f-4m), aged 25 to 60 (*M*=37.67), were recruited to evaluate the tool. Their background ranged from ergonomists (PhD Student to research engineers), engineers in electronics, mechanical design and computer science, and a retired commercial director. Most of the participants use devices with tactile feedback on a regular basis for phone or sports watch notification, for improving immersion in games and in the context of developing and testing haptic devices. Based on their background and subjective level of expertise rating, the participants were divided into three groups: three non-experts with little or no use/knowledge of tactile devices, three haptic engineers and three haptic ergonomists.

Tutorial. After filling a background questionnaire, the participants were instructed about how to use the interface through a tutorial read by the experimenter. This step-by-step tutorial introduced the notion of a haptic pattern, before describing interactively how to create a new pattern using a device template, set its parameters, test it and open/save pattern files. It was available to the user during the experiment along with a one-page visual summary. Two devices were used for the evaluation: a haptic bracelet and a gamepad presented in Fig. 8 and further described below. At the end of the tutorial, the participants were asked to create from scratch a pattern corresponding to the "no" nodding metaphor on both devices to assess whether the participant understood the tutorial and could proceed to the tasks. Metaphors were chosen to describe the patterns to create as they are useful for the design of intuitive patterns [21, 11].

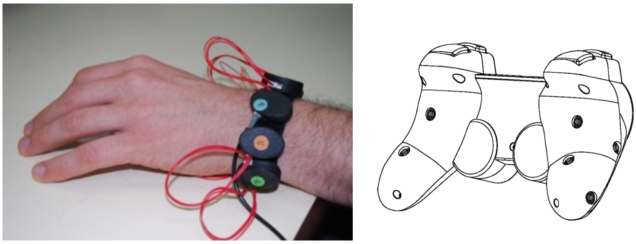


Fig. 8. Left: custom-made bracelet with vibration motors. Right: Gamepad from [7] with 8 electromagnetic actuators (black dots) integrated in the back of an off-the-shelf gamepad.

Devices. The editor was tested with two different devices: a custom-built bracelet [11] and a gamepad developed by Changeon et al. [7]. The custom vibrotactile bracelet (Fig. 8) provides stimulation through eight independent actuators mounted on a velcro band wrapped around the wrist. Each actuator is composed of a commercially available coin motor (Precision Microdrives 310-113), a microcontroller and a power circuit to control the amplitude of the vibration (7 levels). The actuators are linked in series to a supervisor microcontroller that synchronizes the actuation level, the spatial distribution and the timing of the tactile stimuli. The gamepad [7] is composed of eight actuators, with independent frequency and amplitude, and a microcontroller to drive the actuators. Each actuator contains a coil, in a ferromagnetic core, that interacts with a magnet. The magnetic interaction produced generates a controlled attraction/repulsion force and thus produces a controlled vibration. The main parameters for designing patterns are amplitude per sequence (2 levels), frequency (5 levels) and duration (16 values between 0 and 1.6s).

Tasks. At the end of the do-it-yourself example, the participants were instructed to create patterns from scratch for each of the device based on three different metaphors, an audio metaphor “Siren”, a visual metaphor “Turn back” and a tactile metaphor “Stroking”. Each metaphor was described with a sentence and an image. They were each chosen as a representative metaphor for each modality commonly used in the design of patterns: audio, visual and tactile. Participants were given the choice about which device and which metaphor to start with. The metaphors were completed in the following order for most participants: siren, turn back and stroking. At the end, participants were asked to fill out satisfaction and cognitive dimensions questionnaires.

Data. Time for task completion, notes about comments and general use, and answers to questionnaires were collected. The experiment took two hours on average.

4.2 Results

Effectiveness. Participants completed all the tasks without difficulty, leading to patterns they felt satisfied with, no matter the device used and the arrangement of actuators. Participants took longer on average to program on the gamepad but a t-test on the log-transformed data (for the assumption of normal distribution) revealed that the difference was not significant between the bracelet (*M*=245s, *SE*=45.2) and the gamepad (*M*=285.4, *SE*=54.24) with $t(8)=-1.73$, $p > .05$ (see Fig. 9). There are also no significant differences between the user groups (see Fig. 9). Interestingly, some participants commented that it was more difficult to program patterns on the gamepad as the arrangement of actuators was less intuitive than a bracelet.

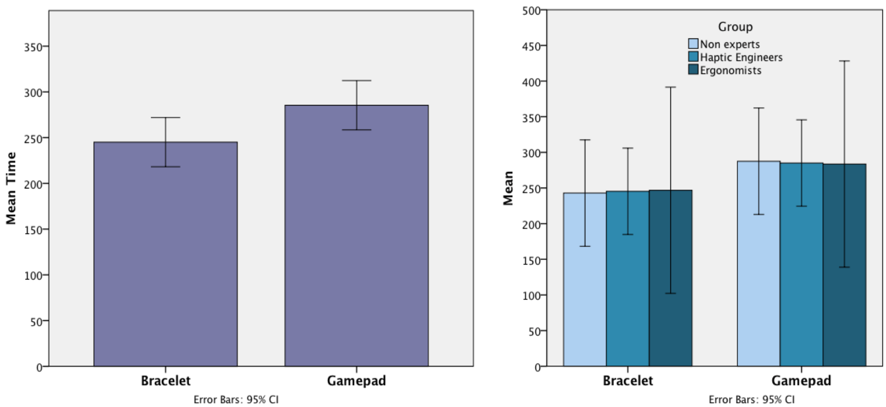
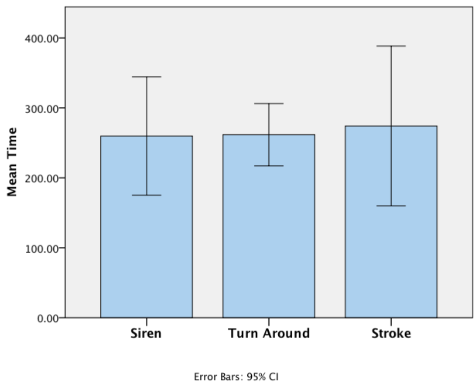


Fig. 9. Results for time completion (in seconds) for each device

Efficiency. The time taken to program a pattern was 265s on average [237s-291s]: patterns were prototyped in less than 5 minutes with the editor. The time taken to create patterns was compared according to the metaphors. We hypothesized that the audio metaphor (Siren) would be the quickest to prototype while the tactile one (Stroking) would be the longest. Indeed, audio can be easily mapped to tactile properties through rhythm and pitch/intensity, whereas a tactile sensation is very subjective and much richer than what current vibrotactile devices provide. The results showed a slight increase in time between the metaphors as we hypothesized (see Fig. 10). We performed a repeated one-way ANOVA on the log-transformed data (for a normal distribution). Mauchly's test indicated that the assumption of sphericity was not violated $\chi^2(2)=3.98$, $p > 0.05$. Though the results are in line with the hypothesis, there are no significant differences between the Siren (M=259.78, SE=42.2), the Turn back (M=261.67, SE=46.5) and the Stroking metaphors (M=274.17, SE=85.5), $F(2,16)=0.36$, $p> 0.05$. There was also no significant effect of expertise, $F(2,6) < 1$ ns. These results show that the editor was used successfully (prototyping < 5min) with no significant differences in time to complete the different tasks for each metaphor.

During the experiment, we noted that participants had more trouble associating the tactile metaphor (Stroking) to a pattern. They thought the shape of the devices was not appropriate and/or the levels of parameters were not fine-grained enough. Interestingly, only the haptic engineers followed our hypothesis and took longer with the tactile metaphor. The non-experts and ergonomists were getting faster in prototyping patterns. This can be explained by the fact that participants were getting more familiar with the interface and with the process of associating a pattern to a metaphor but also in some cases by the lack of ideas for the stroking pattern.



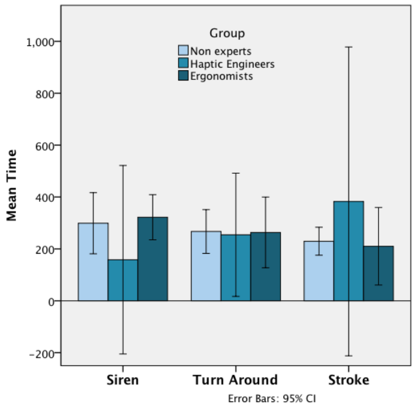


Fig. 10. Results for time completion (in seconds) for each metaphor

To evaluate efficiency along cognitive parameters, we used the cognitive dimensions framework [22], which can be used as an evaluation technique for visual programming environments. It characterizes the system along a set of system properties or cognitive dimensions to illustrate its strengths and weaknesses and trade-offs to realize for effective use. An adapted version of the questionnaire (available online[1]) was provided to the participants with continuous scales assessing each dimension and further questions to elicit feedback and possible improvements. The different dimensions that we used include: visibility (VIJU) for the visibility of the different parts of the system, viscosity (VISC) for the ease of making changes, diffuseness (DIFF) for the verbosity of language, hard mental operations (HMOS) for the demand on cognitive resources, error-proneness (ERRP), closeness of mapping (CLOS) for the closeness of representation to domain, role expressiveness (ROLE) to qualify whether the purpose of a component is readily inferred, progressive evaluation (PROG) for the ability of checking work-to-date at any time, provisionality (PROV) for the ability to play around with ideas, premature commitment (PREM) for constraints on the order of doing things, consistency (CONS) for the expression of similar semantics in similar syntactic forms and secondary notation (SECN) for the possibility to add extra information.

Ratings for each of the used dimensions are depicted in Fig. 11 both on average and for each group of users. Overall the ratings are positive: error proneness (3.2) and hard mental operations (3.6) received low scores while the other dimensions are rated highly, in particular viscosity (8.6), visibility (8), provisionality (8.6), progressive evaluation (8.85) and closeness (8). These are the key features the tool is aiming at, i.e. a clear notation close to the end result where changes can be made easily and tested at any time, letting the user freely play around with ideas. The tool was rated poorly on the secondary notation dimension, justified by the fact that no feature was available to annotate the patterns. This will be added in future versions. It is also worth noting, as expected, that ergonomists were more critical of the system with higher expectations while non-experts were more easily satisfied.

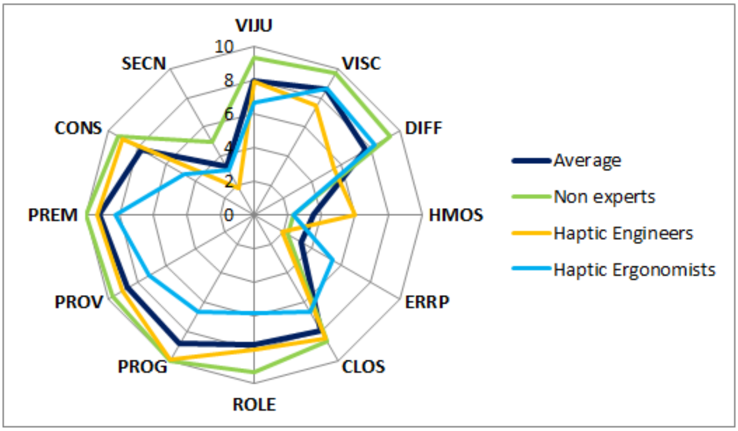


Fig. 11. Averaged ratings for each of the cognitive dimensions scales on a scale from 0 to 10, both overall and for each user group. 0 is generally a bad score except for error proneness (ERRP) and for hard mental operations (HMOS).

[1] http://www.cl.cam.ac.uk/~afb21/CognitiveDimensions/

Overall participants successfully used the system in little time and thought it was easy to use and intuitive, but there was still room for improvement with:

- **General layout:** with suggestions on the placement and look of icons. For instance, users requested to change the play button color and position (1 participant), to move it along with stop and repetitions to the pattern window (1) or just repetitions (1), to keep a single icon for connect/disconnect (1), to make the duration bar shorter (1) and to place the "reset all" icon above the reset sequence icons (1).
- **New functionalities:** such as the possibility to add comments (5 participants), the ability to open several patterns for comparisons (4), the ability to convert a pattern from one protocol to another in the possible extent (1), right click on a file directly from the library to play it (1), a visual dynamic representation of the pattern (1) (this feature is already available by generating an animated gif), and copy/paste a pattern from one window to another (1).
- **Interactions:** covering the improvements of existing widget interactions, e.g. the "All actuators to" a chosen value should apply to the ones selected (4 participants) and the "all duration to" to active sequences (1), copy/paste drag and drop of sequences (2 participants), of time (2) or from one actuator to another (1), the ability to move sequences (2), the possibility to click directly in the slider to change time (2) or using the roller of the mouse (1) or the ability to set the duration via the keyboard (1), separating the time from the frequency (1) and changing the interactions and widgets so that there are no differences between amplitude per sequence and per actuator (2).

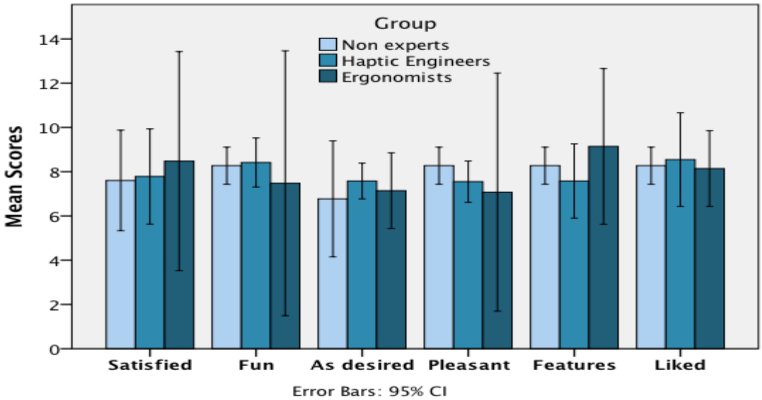


Fig. 12. Satisfaction for each question per user group (out of 10)

Satisfaction. General satisfaction about the interface was evaluated through six statements on a continuous scale from 0 to 10 spanning the general satisfaction of the interface, whether it is fun to use, functioning as desired, pleasant to use and the satisfaction with the features and use. The satisfaction was rated quite highly with an average score of 7.9 out of 10 (see Fig. 12). There is no particular trend according to the user group. Apart from one ergonomist that was critical in her ratings to give room for improvement (see the high variations in the error bars); most participants were happy with the features and enjoyed using the interface.

Pattern Analysis. During the usability study, results about the properties for designing patterns on each device and for each metaphor were also collected. There were various patterns designed but common characteristics emerged for each metaphor (see Table 1 for a detailed list and Fig. 13 for representative patterns).

Table 1. Description of the main characteristics of the user-created patterns for each metaphor. The number in brackets corresponds to the number of patterns with such features.

Siren	Turn back (U-turn)	Stroking

Activation & Movement:
• All (7)
  – Continuously (2) and with amplitude variation (4)
  – Intermittent (1)
• Circular (1)
• Alternating up/down (1)
 	
Amplitude: mostly constant high (4), ↗ (2) and ↗↘ (2), varied (1)
	Activation & Movement:
	• Circular (7)
	  – Alternating half circle (6)
	  – Full circle (1)
	• All intermittently (1)
	• Alternating up/down (1)
	
	Amplitude: mostly medium to high (5), ↗↘ (2), ↗ (1) or low (1)
		Activation & Movement:
		• Circular (7)
		  – Half circles (4)
		  – Full circle (1)
		• Wave around both sides (2)
		• Slight long vibration (2)
		
		Amplitude: low (7) to medium (2)

Activation & Movement:
• Alternating / Intermittent (7)
  – Intermittent all (3)
  – Alternating (3)
  – Both (1)
• ↗↘ the perceived intensity (2) (with number of actuators and/or the frequency and amplitude)
 	
Amplitude & Frequency:
• Both ↗↘ (5)
• High amplitude and medium to high frequency (3)
• Both low (1)
	Activation & Movement:
	• "wave", i.e. actuators playing one after the other (6)
	  – Half circle (3)
	  – Laterally or vertically (3)
	• All-intermittent (1),
	• Alternating left to right (1)
	• Down/side/up motion (1)
	
	Amplitude & Frequency:
	• Low to medium frequency (5)
	• ↗↘ (2) and ↗ (1) frequency
	• High frequency (1)
		Activation & Movement:
		• "Wave" (6)
		  – In one direction (4)
		  – Alternating wave (1)
		  – Both (1)
		• Alternating sides (1)
		• All-intermittent (1)
		
		Amplitude & Frequency:
		• Both low (8)
		• High frequency (1)

- **Siren:** Overall, the patterns reflected the sound effect of sirens, either with an increasing/decreasing intense wailing or an intermittent sound similar to alarm clocks.
- **Turn back (U-turn):** Overall, the resulting patterns were continuous movements either mimicking the “visual” representation of a U-turn or the actual physical movement of going back.
- **Stroking:** Similarly to the “turn back” metaphor, the patterns were based on continuous movements. However, they were combined with low amplitude and frequency to imitate the soft continuous movement of a stroking gesture.

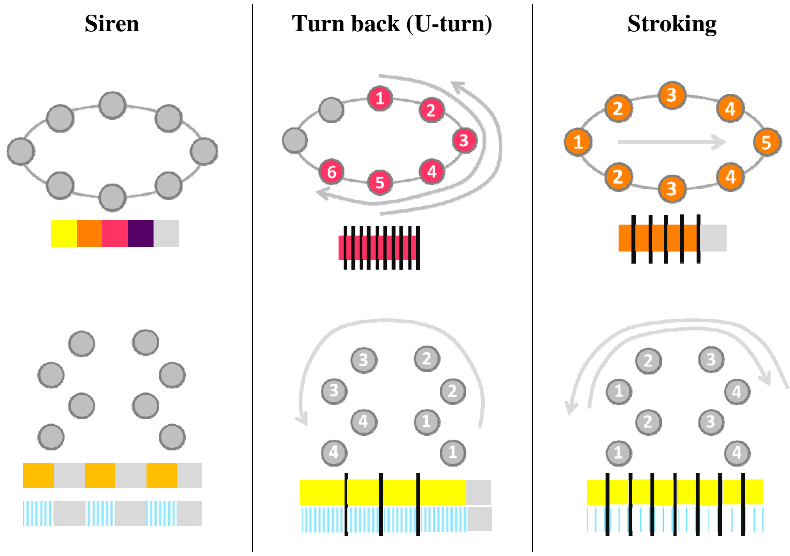


Fig. 13. Examples of the most representative patterns for each metaphor and each device. The grey blocks indicate a pause whereas the black lines indicate a change of actuator with the same amplitude. The colours correspond to the actual values in the interface.

The resulting patterns are, as we anticipated, heavily influenced by the common representations of metaphors, leading to a consensus on the representation, no matter the type of device, as long as they permit it. Given an appropriate set of metaphors, the patterns could potentially be transferrable from one device to another after identifying the dominant parameter(s) related to a metaphor. This would enable the creation of a library of metaphor-patterns that would be suitable to any device. This also supports the hypothesis that original and intuitive patterns can be created from well-chosen metaphors [11]. Further testing with various devices and configurations, various metaphors including ones that can involve localization on the device is needed to investigate further this hypothesis.

5 Conclusion

This paper presented the initial work undertaken to help facilitate the development of vibrotactile patterns and make it accessible to non-experts. It resulted in the development of a visual interface based on the spatial layout of the device, as we believe this plays a role in the pattern design. The interface enables the creation, editing and testing of new patterns on multi-actuator devices and supports various types of vibrotactile actuators. It has successfully been used with a number of custom-developed devices, including a prototype bracelet that consists of six electromagnetic actuators (presented in [23]), and for reproducing patterns described in the literature [4, 3] by adapting the protocol files to match the characteristics of the corresponding devices. The interface can potentially be extended to support any device by providing the corresponding communication protocol and device template. It was also successfully used by nine participants during a usability evaluation that resulted in positive feedback about satisfaction and ease of use, all the while highlighting several areas for improvements (e.g. enabling comparison of patterns, adding annotations, etc.). The resulting patterns provided positive initial insights about the possibility of designing patterns that are transferrable from one device to another.

The future work will not only focus on improving the interface with the collected suggestions but also on investigating other more direct and intuitive approaches to prototype patterns, especially whilst being mobile. A larger study on designing metaphor-based patterns on different devices will also be conducted to validate the initial insights and possibly provide guidelines about a "universal metaphor-based design" of tactile patterns, which takes into account the cultural differences for metaphors.

References

1. Pasquero, J., Stobbe, S.J., Stonehouse, N.: A Haptic Wristwatch for Eyes-Free Interactions. In: SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 3257–3266. ACM (2011)
2. Weber, B., Schätzle, S., Hulin, T., Preusche, C., Deml, B.: Evaluation of a vibrotactile feed-back device for spatial guidance. In: WorldHaptics, pp. 349–354. IEEE Press (2011)
3. Lee, S.C., Starner, T.: BuzzWear: Alert Perception in Wearable Tactile Displays on the Wrist. In: SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 433–442. ACM (2010)
4. Jones, L.A., Lockyer, B., Piateski, E.: Tactile display and vibrotactile pattern recognition on the torso. Advanced Robotics 20(12), 1359–1374 (2006)
5. Heuten, W., Henze, N., Boll, S., Pielot, M.: Tactile Wayfinder: A Non-Visual Support System for Wayfinding. In: Nordic Conference on Human-Computer Interaction: Building Bridges (NordiCHI), pp. 172–181. ACM (2008)
6. Pielot, M., Poppinga, B., Heuten, W., Boll, S.: Pocketnavigator: studying tactile navigation systems in-situ. In: SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 3131–3140. ACM (2012)
7. Changeon, G., Graeff, D., Anastassova, M., Lozada, J.: Tactile Emotions: A Vibrotactile Tactile Gamepad for Transmitting Emotional Messages to Children with Autism. In: Isokoski, P., Springare, J. (eds.) EuroHaptics 2012, Part I. LNCS, vol. 7282, pp. 79–90. Springer, Heidelberg (2012)

8. Ryu, J., Choi, S.: posVibEditor: Graphical Authoring Tool of Vibrotactile Patterns. In: Workshop on Haptic Audio Visual Environments and Games (HAVE), pp. 120–125. IEEE Press (2008)
9. Forlizzi, J., Ford, S.: The building blocks of experience: an early framework for interaction designers. In: Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS), pp. 419–423. ACM (2000)
10. Sanders, E.: From User-Centered to Participatory Design Approaches. Taylor & Francis Books Limited (2002)
11. Brunet, L., Megard, C., Panëels, S., Changeon, G., Lozada, J.: "Invitation To The Voyage": The Design of Haptic Metaphors to Fulfill Occasional Travelers' Needs in Transportation Networks. In: WorldHaptics, 5th Joint Eurohaptics Conference and IEEE Haptics Symposium (WHC). IEEE Press (2013)
12. Enriquez, M.J., MacLean, K.E.: The Hapticon Editor: A Tool in Support of Haptic Communication Research. In: Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems (HAPTICS), pp. 356–362. IEEE Press (2003)
13. Swindells, C., Maksakov, E., MacLean, K.E., Chung, V.: The Role of Prototyping Tools for Haptic Behavior Design. In: Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems (HAPTICS), pp. 161–168. IEEE Press (2006)
14. Immersion Corporation: Haptic Studio, `http://www2.immersion.com/developers/index.php?option=com_content&view=article&id=507&Itemid=835`
15. Lee, J., Choi, S.: Evaluation of Vibrotactile Pattern Design Using Vibrotactile Score. In: Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems (HAPTICS), pp. 231–238. IEEE Press (2012)
16. Lee, J., Ryu, J., Choi, S.: Vibrotactile Score: A Score Metaphor for Designing Vibrotactile Patterns. In: WorldHaptics, Third Joint Eurohaptics Conference and IEEE Haptics Symposium (WHC), pp. 302–307. IEEE Press (2009)
17. Jonas, M.: Tactile Editor: A Prototyping Tool to Design and Test Vibrotactile Patterns. Master's thesis, Computer Science Department, RWTH Aachen University (2008)
18. Minamizawa, K., Kakehi, Y., Nakatani, M., Mihara, S., Tachi, S.: TECHTILE toolkit: a prototyping tool for design and education of haptic media. In: Virtual Reality International Conference (VRIC), pp. 26:1–26:2. ACM (2012)
19. Hong, K.: Demonstration-Based Vibrotactile Pattern Authoring. Master's thesis, Division of Electrical and Computer Engineering, Pohang University of Science and Technology, Korea (2012)
20. Cuartielles, D., Göransson, A., Olsson, T., Stenslie, S.: Developing Visual Editors for High-Resolution Haptic Patterns. In: Workshop on Haptic and Audio Interface Design (HAID) – Posters and Demos, pp. 42–44 (2012), `http://www.english.certec.lth.se/haptics/HAID12/-proceedings.pdf`
21. Pirhonen, A., Brewster, S., Holguin, C.: Gestural and audio metaphors as a means of control for mobile devices. In: SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 291–298. ACM (2002)
22. Green, T.R.G., Petre, M.: Usability Analysis of Visual Programming Environments: a 'cognitive dimensions' framework. Visual Languages and Computing 7, 131–174 (1996)
23. Panëels, S., Anastassova, M., Strachan, S., Van, S.P., Sivacoumarane, S., Bolzmacher, C.: What's Around Me? Multi-Actuator Haptic Feedback on the Wrist. In: WorldHaptics, 5th Joint Eurohaptics Conference and IEEE Haptics Symposium (WHC). IEEE Press (2013)

Towards Many Gestures to One Command: A User Study for Tabletops

Yosra Rekik[1,*], Laurent Grisoni[1,2], and Nicolas Roussel[1]

[1] INRIA Lille Nord Europe, France
[2] LIFL, University of Lille 1, CNRS, France
{yosra.rekik, nicolas.roussel}@inria.fr,
laurent.grisoni@univ-lille1.fr

Abstract. Multi-touch gestures are often thought by application designers for a one-to-one mapping between gestures and commands, which does not take into account the high variability of user gestures for actions in the physical world; it can also be a limitation that leads to very simplistic interaction choices. Our motivation is to make a step toward many-to-one mappings between user gestures and commands, by understanding user gestures variability for multi-touch systems; for doing so, we set up a user study in which we target symbolic gestures on tabletops. From a first phase study we provide qualitative analysis of user gesture variability; we derive this analysis into a taxonomy of user gestures, that is discussed and compared to other existing taxonomies. We introduce the notion of *atomic* movement; such elementary atomic movements may be combined throughout time (either sequentially or in parallel), to structure user gesture. A second phase study is then performed with specific class of gesture-drawn symbols; from this phase, and according to the provided taxonomy, we evaluate user gesture variability with a fine grain quantitative analysis. Our findings indicate that users equally use one or two hands, also that more than half of gestures are achieved using parallel or sequential combination of atomic movements. We also show how user gestures distribute over different movement categories, and correlate to the number of fingers and hands engaged in interaction. Finally, we discuss implications of this work to interaction design, practical consequences on gesture recognition, and potential applications.

Keywords: Tabletop, multi-touch gesture, gesture recognition, interaction design.

1 Introduction

Tabletops have become very good candidate systems for interactive setups. Users in such systems tend to use different class of gestures (e.g. [12]), which implies that any application shall potentially integrate ways to handle such variety. In the meantime, multi-touch gestures are often thought by application designers for a one-to-one mapping between gestures and commands, which does not take into account the high

* Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 246–263, 2013.

variability of user gestures; it can also be a design choice that leads to simplistic interaction choices. There is, to our knowledge, no solution yet to this issue, nor even any path of research drawn yet to solve it. Our motivation is to make a step toward many-to-one mappings between user gestures and commands, by understanding user gestures variability for multi-touch systems. Most studies are task oriented, and allow to exhibit best matches between gesture type, and elementary tasks or commands. Some results show little agreement among users in mapping between gestures and their effect [40]. In order to provide application designers with knowledge that will help designing good many-gestures-to-one-command mappings, we need another experimental approach, which is to exhibit, for a specific type of gesture, all the possible gestural representations that users may achieve. Instead of studying relation between gestures and tasks, we propose to study relation between gesture and underlying symbolic pattern. We advocate, following [1], for the need to construct designed interaction languages, and we provide, in this article, elements that describes how gesture variability may be integrated in such interaction design. We think that a good question, as a start, is: *how do people draw symbolic gestures using their hands on tabletops, and how do they express variability for the same symbol?*

For answering this, we set up a user study in which we target symbolic gestures on tabletops. We have designed our experiment as a two-fold work (see Section 3). From a first phase study, we provide a taxonomy of user gestures; this taxonomy integrates three different views of user gesture (semantic, physicality*, movement structure);* we discuss this in regard to other existing taxonomies, and exhibit its quality. We introduce, through this taxonomy, the notion of *atomic* movement, that may be combined throughout time (either sequentially or in parallel), and can be used to structure user gesture. In a second phase, a user study is conducted with specific class of gesture-drawn symbols; from this study, and according to the provided taxonomy, we evaluate user gesture variability with a fine grain quantitative analysis. Our findings indicate that user equally use one or two hands, also that more than half of gestures are achieved using parallel or sequential combination of atomic movements. We also show how user gestures distribute over different movement categories, and correlate to the number of fingers and hands engaged in interaction. Finally, we discuss implications of this work to interaction design, practical consequences on gesture recognition, and potential applications.

2 Related Work

In the process of developing surface computing technologies, several studies can be found in an attempt to grasp and unify the rich vocabulary of gestural multi-touch interaction, e.g., [28, 22, 16]. We structure the description of related work into three parts: user-centered studies of interaction gestures for tabletops, taxonomy propositions, and formalisms for multi-touch gestures.

A lot of work has been done these past years on user-defined interaction for multi-touch. Given the versatility of free-hand multi-touch gestures and the high variety of users behaviors in producing them [26], user-centric approaches have been at the heart of many research studies on gestures for multi-touch systems. Rather than bounding users to an arbitrary set of gestures defined by system designers (e.g., [25, 31, 37]), Wobbrock et al [40], followed by others (e.g., [29, 5]), adopt a guessability methodology [39] to build up a user-defined gestures set for classical control actions and object manipulations. Finding its fundamentals in [30, 33, 6], this approach consists in presenting the effect of an action to users and then asking them to invent the corresponding gesture. The gesture which was consistently performed by the largest number of users is then retained to be representative of the corresponding action. Within these studies, valuable discussions are reported about the different characteristics of user-defined gestures, e.g., number of fingers, hand pose, etc. However, the focus is on the design and analysis of a one-to-one mapping between gestures and their actions. In [11], Henze et al suggest to derive and compare multiple gesture sets rather than a single one. Their findings indicate that this is a beneficial approach to reduce the risk to exclude promising candidates for gestures. In a field study investigating the variety of gesture performed by people, Hinrichs et al [12] found that users choice of gestures was influenced by the interaction context in which the current action occurred and not only based on preferences for a given gesture for a particular action. They suggest that a many-to-one mapping is also desirable to strengthen the design of gestural interaction techniques. While being specific to pen gestures, Long et al [22,23] studied perceived gesture similarities. Since evaluation of similarities is a complementary problem to the one addressed in this article, such a work can help explaining differences between gestures classes.

Several works also target taxonomy of multi-touch gestures. Wobbrock et al. [40] are among the first to establish a unified taxonomy for surface gestures. They provide a coarse-grain classification along four categories: form, nature, binding, and flow. While these categories are all important, the main focus of our work falls within the *form* category, which captures *how* gestures are performed by users, and its relationship to the *nature* category, which captures the users *semantic* interpretation of gestures. In Wobbrock's taxonomy, the nature category distinguishes symbolic, physical, metaphorical and abstract gestures. The form category distinguishes static or dynamic pose and path for each hand. In the same spirit, Wu et al. [41] describe the process of gesture performance as a finite state machine, with start position (registration), a dynamic phase (continuation), and end position (termination), similar in concept to that described in Charade [3]. Freeman et al [7], in the context of the design of a gesture-learning tool, expanded the form category along three dimensions: registration pose, continuation pose, and movement. Although that taxonomy of Freeman et al presents a sound picture of the large variety of multi-touch gestures, it is not adapted to model variability of association between gestures and symbols.

Remarkable recent researches are being conducted on the formal specification and the reliable recognition of multi-touch gestures. GeForMT [15] provides a formal abstraction of multitouch gestures using a context-free grammar. A discussion of that formalism is given in respect to Wobbrock et al. [40] taxonomy in an attempt to show how it can capture users gestures. Gesture Coder [24] recognizes multitouch gestures via state machines. Proton [19, 18] describes multi-touch gestures as regular expressions modeling a whole sequence of touch events. GestIT [35] is a proof of concept library implementing a meta-model based on compositional operators and Petri Nets to describe multi-touch gestures. All these software-oriented frameworks and languages provide system sound specifications allowing to express complex multi-touch gestures. Nevertheless, it is not obvious how they can apply to capture in a comprehensive and faithful manner the behavior and variability of non technical users in producing gestures. The implications for multi-touch interaction systems to support the variety of users choices in a transparent manner open in fact new opportunities but raises many challenges. For example, recent studies have tackled the difficult issue of designing robust multi-touch recognizers abstracting away the use of multiple fingers [14], or multiple differently ordered strokes [2, 20, 38] for the same symbol. Interestingly, many of those recognizers were designed without prior in-depth analysis of users behaviors thus only taking into account a limited designer-oriented vision of users choices. In this context, a better understanding of users choices is relevant to reduce the dualism of how and who must be adapted to the other: the user or the system [8]. As discussed later in this paper, our findings enlighten the features that deserve deeper modeling efforts for future system-oriented gesture formalizations.

3 User Study

3.1 Overview and Rationale

Eliciting user behaviors have been proved extremely useful to help the design of new, strong and flexible interaction tools (e.g., [33, 40, 5, 9]). Our goal is to broaden the range of possible responses we can get from users and to gain insights into users' variability when issuing a multi-touch command. More generally, we want to elicit the different ways users perceive and issue a multi-touch gesture with the ultimate goal of determining the rules leading to a better definition of what shall be a gestural language for multi-touch interaction. For that purpose, we ask participants to appeal to their imagination to perform different gestures, at the aim of grasping and analyzing the variability and dynamic of user behavior.

As a result, our proposed study is divided in two phases: a first phase, that has two goals: familiarize participants with the interactive surface, and more importantly, observe and analyze their 'intuitive' interaction styles, within an uncontrolled experimental procedure where user can both choose symbol and gesture that represents it. This phase is intended to construct a taxonomy that is used as a basis for the remainder of the study. In the second phase, we achieve quantitative analysis of how users draw symbols, using explicit instructions (name of the symbol, number of variations) and asking participants to explore the different ways to achieve specific symbols. The detailed experimental procedure and context is described in the following.

3.2 Participants

A call for participation has been made using mailing lists and the advertising lobby screens available at our lab and its institutional partners. The call targeted people who were not user-interface designers. In final, we collected results from 30 volunteers, among them 14 were female. Age of participants ranged from 20 to 57 years (average age was 28.4 years). All participants were right-handed. Participant occupations included secretary, chemists, biologists, electronic and mechanics experts, researcher in networks and telecommunications and graduate students. Participant nationalities include different European, African and Asian countries. Self-reported expertise of participants with touchscreen devices were found to significantly differ in the type of interactive surface ($\chi^2 = 62.27$, $p < 10^{-11}$, $\varphi = 0.72$). Fig. 1 summarizes this; we can see that none of our participants previously used a Tablet Pc nor Tabletops. As a result they are completely novice to the interactive surface used in our experiment.

Table 1. Distribution of usage of touchscreen devices among our participants

	Smart Phone	Tablet	Tablet PC	Tabletops and surfaces
Never	8	18	30	0
Occasional	9	9	0	0
Regular	13	3	0	0

3.3 Apparatus

The study took place in our lab, where we had set up a Microsoft Surface 1 measuring 24"x18". Only the experimenter (one author) and the subject were present during the study. During each experiment session, participants' hands were videotaped and the experimenter observed user behavior and took detailed notes. Author notes and videos from each participant were then used in our analysis.

3.4 Procedure

As sketched previously, our study consisted of two phases:

- **First Phase:** The interactive surface is presented to the participant and he/she is explained that the surface accepts multiple fingers. The participant is then asked to perform any multi-touch gesture that comes to his mind and meaningful to him/her. No further comment or request is made that may suggest degrees of freedom, such as number of hands to use, number of fingers, etc. For each performed gesture, users are asked to describe it in a think-aloud protocol. Users have 3 minutes to represent all the gestures they can think of; they are also free to stop before 3 minutes in case they consider the task to be over.
- **Second Phase:** we explicitly provide the participants with the following sequence of 8 symbolic forms (corresponding to a subset of Microsoft Application gestures [27]): circle, square, triangle, vertical line, horizontal line, corner, V and Caret. For each symbol, the participant is asked to perform the symbol using four different manners. Participants are only told the name of the symbols to perform, by oral instruction, and we do not show them any image of the required symbols.

Both phases are conducted consecutively for each participant with a small pause in between. Participants were not constrained by any timing issues when performing their gestures. To prevent any screen content from influencing the gestures participants were performing, we provided no visual feedback of gesture input, e.g., [40,14].

4 Results from the First Phase

4.1 General Observations

In the first phase of our study, we collected 618 user-made gestures. The number of gestures per participant ranged from 8 to a maximum of 46 gestures. Although collected gestures were having broad properties, similar features were observed among different participants. Without surprise participants produced different forms using interchangeably one or more fingers and one or two hands, to draw different kinds of symbols (e.g., line, circle, square, triangle, etc), alphanumeric characters (i.e, letters and numbers), shaped (e.g., tree, heart, flower, star, bird). 6 participants, being regular users of iPads or Smart Phones, additionally performed gestures mimicking standard control actions such as double tape, or rotational, translational and scaling patterns. From our collected data, we were also able to extract several observations about the physical engagement of participants. In particular, 26 (resp. 27) over the 30 participants have used at least once a single (reps. two) hand(s). 24 participants moved simultaneously both hands in symmetric poses. 4 participants alternated from one hand to another in a sequential style. 6 participants used one hand to perform a gesture while their second hand was hold in a stationary pose as to draw a static reference guiding the other hand. This observation holds for gestures performed with a single hand using the thumb and the index. 1 participant used to move hands in the air and touching the surface with her fingers from time to time. 1 participant used exclusively static hand posture on the surface. Except for these two cases, the relative movement of participants fingers was the rule guiding the achievement of gestures. Neither the number of fingers nor their type seemed to us as a conscious parameter that participants were intentionally thinking about. We did also notice no particular preference on the start and the end positions of performed movements. Participants mostly used their right hands when moving from left to right, and inversely they used their left hands when moving from right to left. However, we did not notice other apparent rules applying to the direction of movements nor to the size of their trajectories.

4.2 Qualitative Analysis: User Gesture as Atomic Movements Combined over Parallelism and Sequentiality

From the gestures collected during the first phase, we extracted several observations that provide elements about how users perform gestures.

- Atomic movements: The very recurrent observation in participants' behavior is that they grouped their fingers into unitary blocks moving in a consistent manner, while being completely free from the microscopic timeless notion of touch as may be handled by the system. We found that number of contact fingers does not impact the accomplishment of their movements, as long as involved fingers are close to each others. The interesting observation is that the notion of proximity is relative to user-proper referential and seems to be hardly definable in absolute and universal manner from a system point-of-view. Users referential can in fact be substantially scaled up or down from the performance of one gesture to another one. However, it tends to stay constant and consistent over time and through possibly multiple movements composing the same single gesture. From this observation, we introduce the notion of *atomic movement* which reflects users' perception of the *undividable* role that a group of fingers is playing when performing a gesture. From our observations, users atomic movements are mostly in reference with the imaginary trail of a group of fingers which position is evolving in closely related movements. An atomic movement can have an internal state that can change depending on hands shape, fingers arity, velocity, direction, etc. However, state changes do not alter the role an atomic movement is playing in users' mind and its primary intention. From our observations, we distinguish between two categories in participants movements depending on whether (i) the trail corresponding to fingers is stationary or (ii) it implies an embodied motion. As a practical examples, variable number of fingers, from one or two hands, moving together following the same path or being held stationary to delimit or point a region in the interactive surface, are among the most frequently observed atomic movements.

- Parallelism: For some gestures, participants combined the movements of their fingers simultaneously, in either a symmetric or an asymmetric style. From a geometrical perspective, symmetry occurs mostly between two atomic movements performed in parallel on the surface such as the trajectory of the one was the mirrored image of the other. For instance, users are observed to produce circled pattern by moving fingers from both hands in parallel such us the trajectory of each hand forms a semi-circle. On the other side, asymmetry in gestures occurs when participants were holding some fingers stationary upon the surface and simultaneously moving some others. For instance, users are observed to produce circled patterns by touching a region of the surface with one hand and simultaneously moving fingers from the other hand all around. From a physicality perspective, bi-handed parallel movements are mostly attended with the use of the same fingers combination on each hand, while the use of one hand mostly engages the use of the index and the thumb. From these observations, we found that symmetry in users gestures can be described by the parallelism expressed by atomic movements.
- Sequentiality: We observed that some participants often operate in a sequential manner by iteratively posing and moving fingers on the surface, then releasing and posing fingers again at a new location binding the set of already performed movements. Users sequential movements imply more than a time pause or direction change. They are performed using one hand, as well as alternating different hands or fingers, and mixing parallel atomic movements with single atomic movements. In this class of interaction style, movements are mixed and matched both in time and in space according to users specific referential. This referential does not map perfectly with the system. For example, the boundaries of strokes induced by the atomic multi-finger movements are never perfectly matching with one another, though we think that participants intended to do so in their minds.

From this analysis, user gestures can be modeled using atomic movements, possibly combined along with parallelism and sequentiality. Fig. 1 provides a simple situation illustrating this with three gestures produced by different participants.

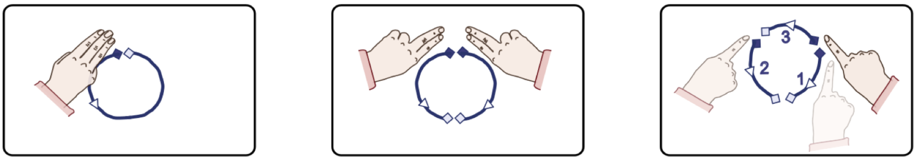


Fig. 1. Example of Atomic movements combination through time. From left to right: Elementary atomic movement, parallel movement, sequential movement.

4.3 An Embodied Taxonomy of Multi-touch Gesture

To capture the space in which our participants were conceiving and producing gestures, we propose the multi-level layered taxonomy summarized in Fig. 2. It is worth noticing that the levels of our taxonomy do not model separable attributes to be characterized individually. Instead, they represent the different aspects of a single unified dynamic mechanism ruling users in the achievement of a multi-touch gesture.

SEMANTIC-CONCEPT		
Mental meaning, Users' thoughts		
↕		
PHYSICALITY		
Enabling Motor Skills (e.g., hands, fingers)		
Posture, Arity (e.g., single, multiple, mixed)		
↕		
MOVEMENTS		
A set of Atomic Movements		
Elementary (E)	Ref (R)	
	Motion (M)	
Compound (C)	Parallel (P)	$P := P_1 * P_2 ; P_1, P_2 \in \{E, P\}$
	Sequential (S)	$S := S_1 – S_2 ; S_1, S_2 \in \{E, C\}$

Fig. 2. A multi-level model for users' gestures. The '*' (resp. '-') refers to movements performed simultaneously (resp. sequentially) in time.

At the high level of our taxonomy, we capture the fact that a multi-touch gesture emerges from what user's mind is modeling before even touching the surface. In this respect, an external observer can only try to guess the *semantic concept* hidden in user's gesture, since it might be the case that the gesture it-self is not sufficient to fully reveal user's thought — which is in accordance with previous studies [36,13,40]. From a neurological perspective, hands and fingers are controlled and coordinated by human motor system at the aim of achieving a desired task. The *physicality* level thus captures the motor control allowing users to project the semantic level into the interactive surface. The *movement* level is the consequence of the motor goal expressed by hands and fingers motions in order to infer unitary blocks building the whole gesture.

The movement level is at the core of our taxonomy since it constitutes the interface between the user and the interactive surface/system. Consistent with our observations, we propose to structure this level according to two generic classes built in a recursive manner. At the low level of the recursion, we find the class of gestures formed with an elementary atomic movement. An elementary atomic movement can be either of type stationary (Ref) or Motion as discussed previously in our qualitative observations. The *Compound* class refers to the recursive composition of a set of atomic movements. It is expanded in two categories depending on the time combination of composing atomic movements. The *Parallel* category refers to users making two or more different but simultaneous movements. The *Sequential* category refers to users performing a set of atomic movements, being possibly parallel or elementary, holding and releasing hands or fingers, on and from the surface, in a discrete iterative manner.

4.4 Users Variability

Our taxonomy is the result of a qualitative empirical synthesis of a wide range of collected gestures. We found that the three levels of our taxonomy contribute leveraging and unifying the high variety of users gestures. In fact, users gestural variations can be elicited as the result of the mental picture and the time-space composition of atomic movements, as well as their physical mapping into users fingers and hands.

At the semantic level, the global pattern induced by movements is the most apparent attribute that users where instantiating in several different manners. However, gestures with similar global patterns can have different properties, e.g., their composing atomic movements can be in different classes. At the physical level, variations in the number of fingers and hands are a natural outcome for most participants. Finally, users variations can be captured at the movement level by eliciting the different possible time combinations of atomic movements (Motion and Ref) as well as their number which can vary from a gesture to another and from a user to another.

4.5 Comparison with Existing Taxonomies

Comparing to previous taxonomies, the semantic concept of our taxonomy relates to the nature category defined by Wobbrock et al [40]. In that study, users were shown the effect of a gesture, then they was asked to issue the gesture. Hence, the nature category is tightly related to the action of the gesture. In our first-phase study, we did not ask participant to perform any precise action. Thus, the semantic concept level only reveals the meaning of the gesture without mapping it to the type of a particular action. On the other hand, physicality in our taxonomy relates to the form category sketched in [40] and expanded by Freeman et al [7]. The registration, continuation and movement dimensions described within the form category there-in did not directly result from a specific user-centric study, since the intention of Freeman et al work was primary focusing on teaching the user how to perform a gesture. Although, those dimensions provide a sound picture of *how* users may perform a gesture, we find that the physicality and movement levels of our user-centric taxonomy complements and refines in many aspects the empirical work of Freeman et al. For example, Freeman's distinguishes between two types of movements: *path* and *no path,* depending on whether the hand moves along a surface path or not. In our work, we explicitly distinguish between how users perform gestures (Physicality) and the notion of movements. In this respect, the Movements level introduces a new dimension in users gestures and consistently renders the embodiment of gestural multi-touch interaction. The semantic concept behind users gestures can then be captured within an embodied and coherent flow engaging the cooperation of users fingers and hands to materialize the inter-relation between a set of unitary atomic blocks composing the gesture.

5 Results from the Second Phase

In our second-phase experiment, participants were asked to produce symbols in four different manners. We were able to adequately classify all the gestures from this experiment phase using our taxonomy. In addition to that, no pair of gestures associated to the same symbol, and identified as different by one user, falls into the same category according to our taxonomy. As an example, appendix shows the set of gestures that we noticed in the experiment in the case of the circle symbol.

In this section, we discuss taxonometric breakdowns of the variety of gestures proposed by participants. 2 participants were excluded from the provided statistics since they made less than the four gestures required per symbol in our experiment. 5 participants produced more than four different gestures for some symbols. We constraint our analysis to only the four first ones. Overall, we have retained 28 x 8 x 4 = 896 gestures that are analyzed in the following.

5.1 Physicality/Movement Inter-dependency

Fig. 3 shows the ratio (averaged over all users) of gestures performed with one and two hands for each symbol and overall. We also incorporate the amount of fingers (single or multiple) engaged per each single hand. If the gesture is movement-compound, we count it multi-finger if at least one hand was engaged with more than one finger. We can see that users-gestures are fairly distributed over one hand (52.77%) and two hands (47.22%). Although, participants used more often a single finger per hand (78.17%), a significant ratio of gestures where multiple fingers are used per hand can still be reported (21.83%). A Friedman test revealed that Symbols' type does not have a significant effect on the ratio of two handed gestures performed by users.

Fig. 4 shows gestures ratios classified by movement categories, where we further distinguish between one-handed and two-handed gestures. A Friedman test revealed a significant effect of Symbols' type on the ratio of movement categories (Elementary: $\chi^2 = 32.55$, $p < 3.10^{-5}$; Parallel: $\chi^2 = 21.98$, $p < 2.10^{-3}$; Sequential: $\chi^2 = 50.12$, $p < 1.10^{-8}$). In the elementary atomic category, a post-hoc test using Wilcoxon test showed the significant differences of the couple of symbols (vertical line, horizontal line) and the other symbols. We attribute this to the fact that this couple of symbols does not imply direction change in fingers movements so that elementary atomic movements are the more natural to conceive for users. In the Parallel category, significant differences were found between the couple of symbols (V, Caret) and the other symbols. Actually, the ratio of parallel two-handed gestures performed for these two symbols is higher compared to the other symbols. This can be explained by the fact that these symbols can be more easily mapped into users two hands. In the Sequential category,

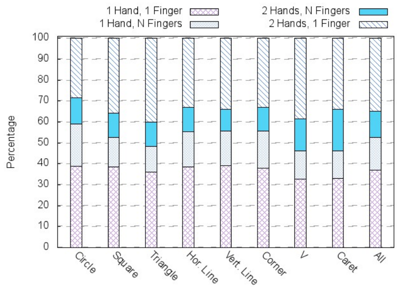


Fig. 3. Hands and Fingers (per hand) ratio

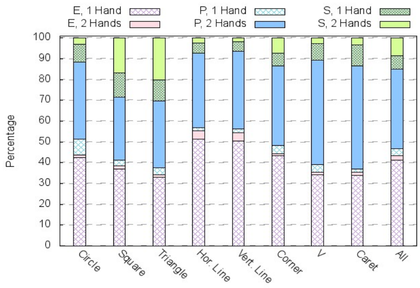


Fig. 4. Movement and Hands ratio (E: Elementary, P: Parallel, S: Sequential)

significant differences were found between the couple of symbols (square, triangle) and the other symbols. These two symbols are in fact clearly different from the others by the number of stroke combinations that can be used to perform them. Overall, we can see that users produced atomic and compound parallel gestures in approximately the same proportion (resp. 43.34% and 41.76%), while the compound sequential category is represented in a relatively non-negligible ratio of 14.89%. A Chi-square test with Yates' continuity correction revealed that the percentage of two-hand and one-hand gestures significantly differed by movement category ($\chi^2 = 523.34$, $p < 2.10^{-16}$, $\varphi = 0.78$). We can in fact remark the high correlation between two-handed gestures (resp. one-handed) and the parallel movement category (resp. elementary).

Finally, Fig. 5 shows the different combinations of hands fingers and their mapping to the movement categories. Overall we observed 18 possible one-handed and two-handed finger combinations. The parallel movement category is represented in 12

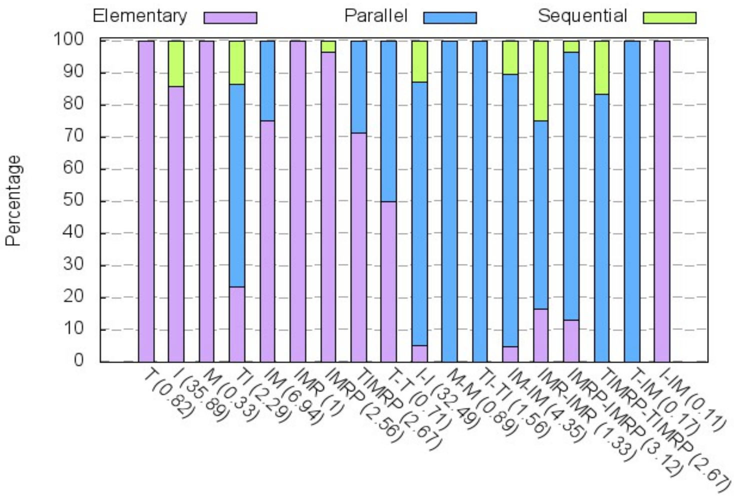


Fig. 5. Movement category ratio according to Hands-Fingers Combinations. T, I, M, R and P, denote resp. the thumb, index, middle, ring, and pink. The sign '-' distinguish between left and right hand. Numbers in braces refer to the ratio of the finger combination over all users.

combinations among them only the index-thumb and index-middle is one handed. The other parallel two-handed combinations show a high similarity in the type of fingers used per hand. We can notice the absence of gestures engaging the pink or the ring in an elementary atomic movement. Whenever these two fingers are used, they appear in combination with the middle, index, and/or thumb fingers, by inducing the same multi-finger atomic movement. These combinations reflect the natural (comfortable) motor capabilities of users as well as the affordance of hand movements and their dependencies – which is consistent with previous studies on the mechanical/neurological relationship between fingers and their kinematics, e.g., [15, 27, 41].

5.2 Users' Transition-Frequency Automatons

Gestures' properties were not random over the four trials allowed per symbol. In order to capture users thoughts and priorities in conceiving the different manners of producing a gesture, we study in this section the evolution of gestures properties over time using probabilistic automatons [34], mapping gestures properties into states and users variations into transitions. Fig. 6 shows four such automata in a comprehensive informal manner. The first three automata are user-centric while the fourth one provides a more system-centric perspective as it will be discussed in the following.

Every initial state of the automata depicted in Fig. 6 refers to participants starting the experiment. Columns refer to subsequent gestures produced by participants. The rows of the first automaton classify gestures according to whether they are one-handed or two-handed. Those in the second automaton classify gestures according to the movement category. The third automaton distinguishes between gestures where every composing atomic movement is single-finger and those where at least one atomic movement is multi-finger. Finally, the fourth automaton classifies gestures depending on whether exactly one touch is involved throughout the whole gesture, or multiple touches are involved. The main difference with the third automaton is that touches are viewed relative to the system and not to users. The numbers in each cell is then computed as the average ratio over all users of gestures found in the corresponding state. This provides gestures distribution over time and can be interpreted as the empirical probability of user's gesture property being mapped the corresponding state. Similarly, transitions depicted by labeled rows show the average ratio of participants moving from a state to another, which can be interpreted as the empirical conditional probability of falling in the subsequent gesture type knowing the type of the present gesture. For example, the initial state of the first automaton reads as 0.74 of participants perform the first gesture with one hand, or alternatively as users perform a two-handed gesture first with probability 0.26. Given that a user performs the first gesture with one hand, there is a probability of 0.45 that the outcome of the second gesture is two-handed. Notice that the cells in each column sum to one, which provides the empirical probability distribution (and thus the average ratio) of corresponding gesture types. These automata are actually averaged over users and symbols so that they only represent the behavior of an 'average' user over all symbols. We did make a more fine symbol-dependent analysis, not shown here, which revealed that different states for the four gesture trials are observed within every symbol; however the probability transition is different from a symbol to another and from a user to another.

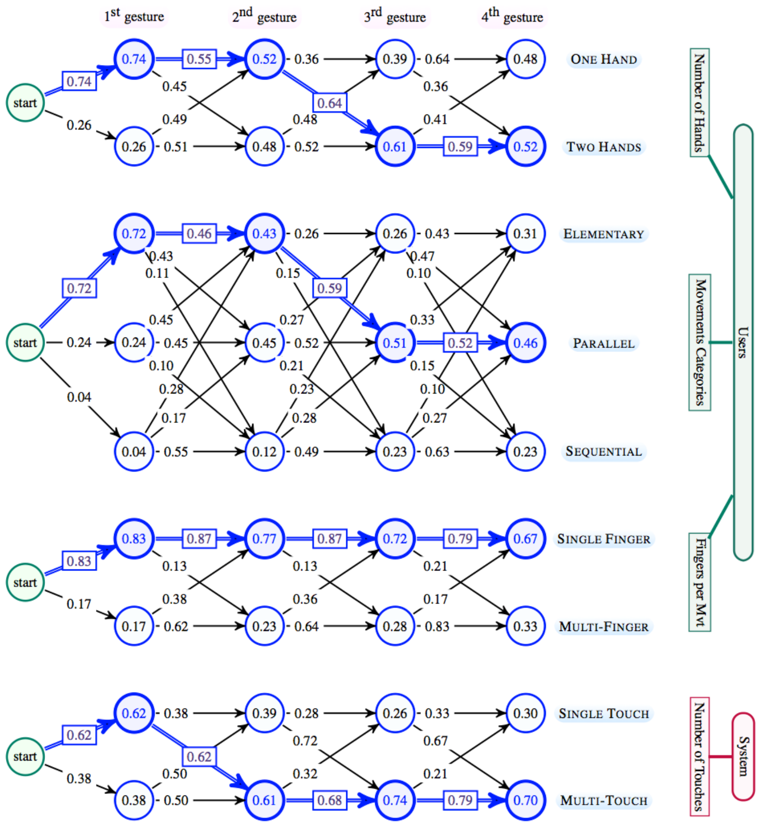


Fig. 6. Gestures Transition-Frequency Automatons (average over all symbols and users) according to (from top to bottom): Number of Hands, Movement categories, Number of fingers per atomic movement, and Number of contacts

Fig. 6 provides a time-dependent information about gesture type frequencies and variations. For example, as one can see in bold lines referring to the most likely gesture transitions, users start more likely with one-handed, elementary-atomic-movement, single-fingered, single-touch gestures. Then-after, they are more balanced in their choices consistently switching to two-handed and parallel movements. The empirical probability that users start with two-handed or parallel-movement gestures is relatively significant (0.26 and 0.24) and grows sharply as users advanced in the experiment. This is to contrast with fingers usage since a single finger per movement is most often used all along the produced gestures. We interpret this as bi-manual usage and movement variations being the most significant features ruling users' mind in performing the different set of gestures. Although the sequential-movement strategy is unlikely as a starting strategy (0.04), it is interesting to remark that users falling in this state are more likely to produce the same type of movements in subsequent gestures. We can interpret this as the sequential mode offering more degree of freedom in producing different gestures by consistently playing with hands movement combination. Finally, we remark that from the system perspective, gestures involving multiple touches on the surface are significantly represented all along the experiment mostly because users are either engaging their two hands or performing parallel movements.

6 Implications for Gestures-Based Application Design

In this section, we discuss the implications of our results for gesture design, surface technology, and user interfaces. In particular, we address points that seem important in order to design application that authorize several gestures for one command.

6.1 Movement Matters More Than Posture

Our user study demonstrates that the movement induced by fingers motion matters for participants more than hand posture. Over all our participants, only two participants performed static gestures where fingers or hands were maintained stationary (Ref). Only in this case, the motor skills (blob type, posture, arity) used to structure the gesture are important, while their movements are not. Static gestures (set of Ref atomic movements), where hands/fingers posture is crucial, go beyond available classical multi-touch surfaces and need further sensing and input processing technology. In contrast, gestures where movements on the surface are crucial, are more accurate to the available knowledge and expertise on processing multi-touch input. Therefore it is our opinion that movement-based gestures provide more space to fully take advantage of nowadays multi-touch technology, so that their study and understanding should be the priority in the short term. However, current trends in augmenting surface computing technologies with new sensing facilities are also compatible with our user study. Advances in these directions would allow to enrich multi-touch surface input vocabulary so that gestures embodiment and versatility can be better encapsulated and exploited within gesture interfaces.

6.2 Interaction Gesture: A Multi-level, Multi-view Phenomenon

Our study reveals that the variations of users gestures for the same command can be structured and classified by the specific properties of a set of atomic movements. Although the notion of atomic movement and the role it plays in our gesture taxonomy constitutes a low level abstraction of what users them-selves are modeling, it should as well serve for designers as a basic tool in the process of thinking, formalizing and setting up multi-touch gestural interaction techniques. Designing for multi-touch gestures as multi-movement entities embodied in users thoughts would then push a step towards shortening the gap between designers vision and the way multi-touch gestures are perceived and produced by end-users. In particular, an atomic movement is by definition not sensitive to the number of fingers or the number of hands being used, so that it enables to unify and to leverage previous studies recommending to not distinguish gestures by number of fingers, e.g. [40]. Thinking about multi-movement multi-touch gestures, one have to keep in mind that the interdependency between the set of atomic movements forming a multi-touch gesture highly depends on users motor control over time and over space. Two main alternatives can be elicited depending on whether one hand or two hands are considered. In both cases, it is more likely that the movements occur in parallel that is simultaneously in time. In addition, users are more likely to engage a single finger in the performance of one elementary atomic movement, thought this should not serve as *the* rule.

Besides allowing to provide guidelines for the design of multi-touch gestures, the atomic movement perspective allows to expand in a comprehensive, yet precise, manner the space of possible mappings between a command and users gestures. By investigating the different possible combinations at the movement level, a variety of single and multi-finger, single and two-hand gestures can be supported, which can: (i) improve flexibility, (ii) not penalize users by offering adequate response, and (iii) make sure that the variety of users choices leads to a gratifying interactive experience.

6.3 Gesture Recognition Needs to Be Deeply Rethought

From the aspect of system feasibility, our study raises new challenges for the generic encoding and the reliable recognition of multi-movement multi-touch gestures. In fact, a formal and rigorous system-computable definition to what is an atomic movement is first needed. From the quality of such a definition depends the design of system embedded programs that determines a faithful representation of users' atomic movements and enable a consistent processing and interpretation of users gestures. One promising research path is to augment existing multi-touch frameworks based on formal grammars (such as proton++ [18] and others [24, 15, 35]) with both (i) declarative language elements that capture the notion of touch closeness in an elementary atomic movement, as well as with (ii) new compositional operators that render the time and space relation of atomic movements. The goal would be to automatically encompass users mental model features like: the independence of movements from the number of fingers, the possible variations in the combination of movements etc, within such formal frameworks. For patterned shape gestures, the challenge is more on the extraction of the different strokes implied by users movements. For example, it is not clear how recognizers in the $-family [38] can handle the fact that a stroke could be constructed by users using a variable number of fingers. Recently, Jiang and al. [14] proposed an algorithm to extract a single stroke from the different trajectories of multiple fingers on the surface. However, this reduction is incompatible with the fact that multiple strokes can interleave in time, e.g., drawing a circle or a square or triangle or V or caret using two symmetric parallel atomic movements will be recognized as a line. We argue that the state-of-the-art recognizers for multi-touch gestures have to be rethought to support usability and consistently take into account the variety of users gestures in issuing a command. One path can be to take advantage from the consistency of the notion of touch closeness with respect to every user global time-space referential when performing atomic movements.

7 Conclusion and Future Works

In this paper, we investigated the different gestures that users adopt to issue the same symbol. We provided direct implications of our findings for the design of tabletops gesture-based applications. For reusability purposes, it is important to elicit the 'natural' alternatives available for users to perform a symbolic command. In future work, it would be interesting to investigate all the potential uses of such a variability integration, and its practical impact on gestural interaction techniques. Besides, it can be interesting to explicitly ask users to perform gestures in different classes and to evaluate their preferences and ranking of each class. Taking variability into account within gesture recognition, and integrating it into available gesture recognizer for tabletops and multi-touch systems is also a challenging issue which still has to be addressed in the future. A more general issue related to our user-study is to investigate to what extent our findings on user's variability can be applied to other type of gesture detection devices which do not require a contact surface, e.g., Kinect.

Acknowledgments. This work is partially funded by Interreg IV-A 2-seas SHIVA project.

References

1. Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., Zhao, S.: Tzu Kwan Valino Koh, J.: Understanding Mid-Air Hand Gestures: A study of Human Preferences in Usage of Gesture Types for HCI, Microsoft Research TechReport MSR-TR-2012-111
2. Anthony, L., Wobbrock, J.O.: $n-protractor: a fast and accurate multistroke recognizer. In: GI, pp. 117–120. Canadian Information Processing Society (2012)
3. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. ACM Commun. 36(7), 28–35 (1993)
4. Bevans, A.: Investigating the effects of bimanual multitouch interaction on creativity. In: Proc. of Creativity & Cognition, pp. 451–452. ACM (2011)
5. Cohe, A., Hachet, M.: Understanding user gestures for manipulating 3d objects from touchscreen inputs. In: GI, pp. 157–164. Canadian Information Processing Society (2012)
6. Epps, J., Lichman, S., Wu, M.: A study of hand shape use in tabletop gesture interaction. In: CHI EA, pp. 748–753. ACM (2006)
7. Freeman, D., Benko, H., Morris, M.R., Wigdor, D.: Shadowguides: visualizations for in-situ learning of multi- touch and whole-hand gestures. In: ITS, pp. 165–172. ACM (2009)
8. Good, M.D., Whiteside, J.A., Wixon, D.R., Jones, S.J.: Building a user-derived interface. Comm. of the ACM, 1032–1043 (1984)
9. Guilford, S.J.: The nature of human intelligence. McGraw-Hill series in psychology
10. Hager-Ross, C., Schieber, M.H.: Quantifying the independence of human finger movements: Comparisons of digits, hands, and movement frequencies. The Journal of Neuroscience 20, 8542–8550 (1967)
11. Henze, N., Löcken, A., Boll, S., Hesselmann, T., Pielot, M.: Free-hand gestures for music playback: deriving gestures with a user-centred process. In: MUM 2010, pp. 16:1–16:10. ACM (2010)
12. Hinrichs, U., Carpendale, S.: Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. In: CHI, pp. 3023–3032. ACM (2011)
13. Huang, T.S., Pavlovic, V.I.: Hand gesture modeling, analysis, and synthesis. In: IEEE International Workshop on Automatic Face and Gesture Recognition, pp. 73–79 (1995)
14. Jiang, Y., Tian, F., Zhang, X., Liu, W., Dai, G., Wang, H.: Unistroke gestures on multi-touch interaction: supporting flexible touches with key stroke extraction. In: IUI, pp. 85–88. ACM (2012)
15. Kammer, D., Wojdziak, J., Keck, M., Groh, R., Taranko, S.: Towards a formalization of multi-touch gestures. In: ITS, pp. 49–58. ACM (2010)
16. Karam, M., Schraefel, M.C.: A taxonomy of gestures in human computer interactions. Technical report, Univ. of Southampton (2005)
17. Kin, K., Hartmann, B., Agrawala, M.: Two-handed marking menus for multitouch devices. ACM Trans. Comput.-Hum. Interact., pp. 16:1–16:23 (2011)

18. Kin, K., Hartmann, B., DeRose, T., Agrawala, M.: Proton++: acustomizable declarative-multitouch framework. In: UIST, pp. 477–486. ACM (2012)
19. Kin, K., Hartmann, B., DeRose, T., Agrawala, M.: Proton: multitouch gestures as regular expressions. In: CHI, pp. 2885–2894. ACM (2012)
20. Li, Y.: Protractor: a fast and accurate gesture recognizer. In: CHI 2010, pp. 2169–2172. ACM (2010)
21. Lin, J., Wu, Y., Huang, T.S.: Modeling the constraints of human hand motion. In: HUMO, pp. 121–126. IEEE Computer Society (2000)
22. Long, A.C., Landay Jr., J.A., Rowe, L.A.: Implications for a gesture design tool. In: CHI, pp. 40–47. ACM (1999)
23. Long, A.C., Landay Jr., J.A., Rowe, L.A., Michiels, J.: Visual similarity of pen gestures. In: CHI, pp. 360–367. ACM (2000)
24. Lü, H., Li, Y.: Gesture coder: a tool for programming multi-touch gestures by demonstration. In: CHI, pp. 2875–2884. ACM (2012)
25. Malik, S., Ranjan, A., Balakrishnan, R.: Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In: UIST, pp. 43–52. ACM (2005)
26. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. Univ. of Chicago Press (1992)
27. Microsoft. Application gestures and semantic behavior (windows) Website, `http://msdn.microsoft.com/en-us/library/windows/desktop/ms704830v=vs.85.aspx`
28. Morris, M.R., Huang, A., Paepcke, A., Winograd, T.: Cooperative gestures: Multi-user gestural interactions for co-located groupware. In: CHI, pp. 1201–1210. ACM (2006)
29. Morris, M.R., Wobbrock, J.O., Wilson, A.D.: Understanding users' preferences for surface gestures. In: GI, pp. 261–268. Canadian Information Processing Society (2010)
30. 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)
31. Rekimoto, J.: Smart skin: an infrastructure for free hand manipulation on interactive surfaces. In: CHI, pp. 113–120. ACM (2002)
32. Schieber, M.H., Santello, M.: Hand function: peripheral and central constraints on performance. J. Appl. Physiol., 2293–2300 (2004)
33. Schuler, D., Namioka, A.: Participatory Design: Principles and Practices. L. Erlbaum Associates Inc. (1993)
34. Segala, R., Lynch, N.: Probabilistic simulations for probabilistic processes. Nordic J. of Computing, pp. 250– 273 (1995)
35. Spano, L.D., Cisternino, A., Paternò, F.: A compositional model for gesture definition. In: Winckler, M., Forbrig, P., Bernhaupt, R. (eds.) HCSE 2012. LNCS, vol. 7623, pp. 34–52. Springer, Heidelberg (2012)
36. Thieffry, S.: Hand gestures. In The Hand (R. Tubiana, ed.), pp. 488,492, Univ. of Chicago.
37. Tse,E.,Shen,C.,Greenberg,S.,Forlines,C.: Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. AVI, pp. 336–343. ACM (2006)
38. Vatavu, R.-D., Anthony, L., Wobbrock, J. O.: Gestures as point clouds: a $P recognizer for user interface prototypes. ICMI, pp. 273–280, ACM (2012)
39. Wobbrock, J. O., Aung, H. H., Rothrock, B., Myers, B. A.: Maximizing the guessability of symbolic input. CHI EA, pp.1869–1872. ACM (2005)
40. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. CHI, pp. 1083–1092. ACM (2009)
41. Wu, M., Shen, C., Ryall, K., Forlines, C., Balakrishnan, R.: Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. ITS, pp. 185–192. IEEE CS (2006)

Appendix

In Fig. 7, we summarized a representative set of gestures produced by users for the circle symbol by adequately classifing them according to our taxonomy. We show different movement categories and their mapping into fingers and hands. Although Fig. 7 shows gestures relative to the circle symbol, the depicted fingers and hand poses, as well as the induced atomic movements fairly holds for the other symbols.

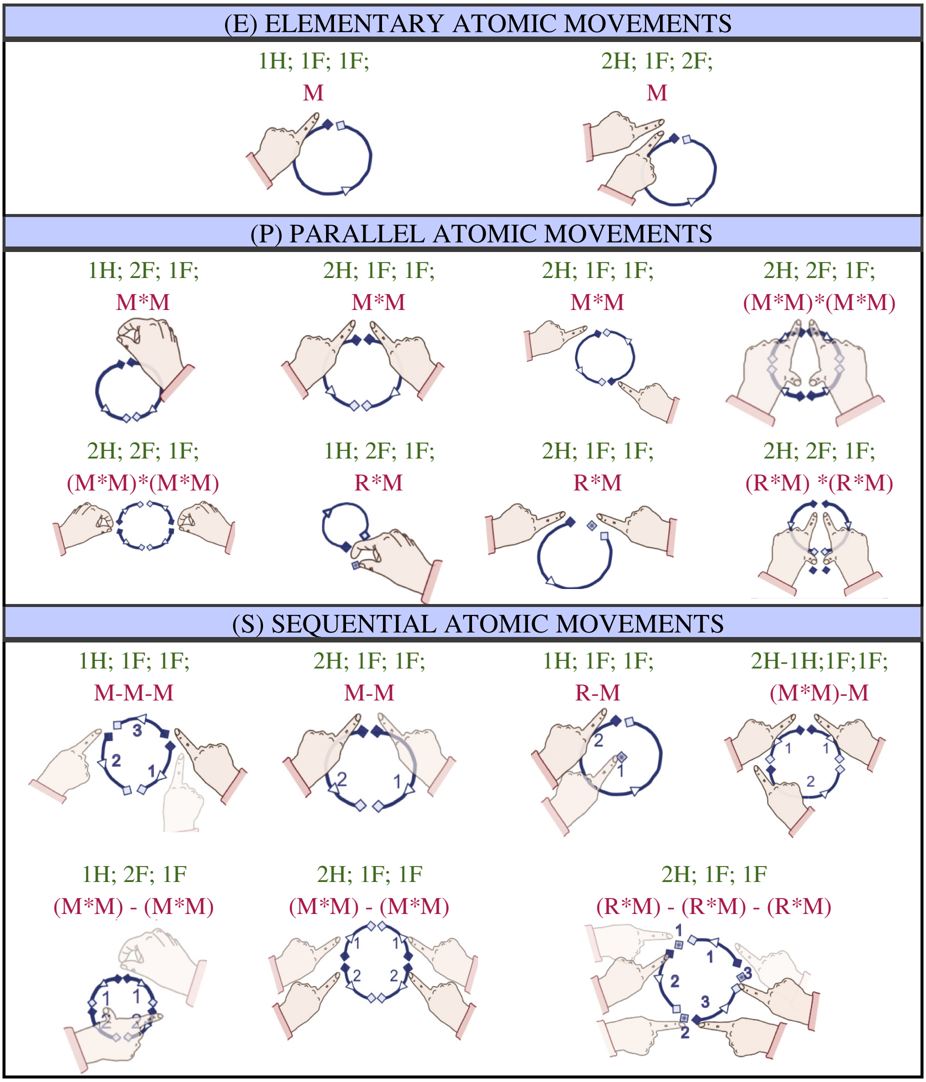


Fig. 7. A representative set of gestures for the circle symbol. We show respectively the number of hands used; the number of fingers per hand; and the number of fingers per movement (e.g., 2H; 1F; 1F; reads as: two hands, one finger per hand and one finger per movement). The atomic movements (R: Ref or M: Motion) and their time composition is also explicited.

User-Defined Body Gestures for an Interactive Storytelling Scenario

Felix Kistler and Elisabeth André

Human Centered Multimedia, Augsburg University,
Universitätsstr. 6a, 86159 Augsburg, Germany
{kistler,andre}@informatik.uni-augsburg.de

Abstract. For improving full body interaction in an interactive storytelling scenario, we conducted a study to get a user-defined gesture set. 22 users performed 251 gestures while running through the story script with real interaction disabled, but with hints of what set of actions was currently requested by the application. We describe our interaction design process, starting with the conduction of the study, continuing with the analysis of the recorded data including the creation of gesture taxonomy and the selection of gesture candidates, and ending with the integration of the gestures in our application.

Keywords: User Defined Gestures, Kinect, Full Body Tracking, Depth Sensor, Interaction, Interactive Storytelling.

1 Introduction

Creating intuitive interaction has always been a difficult, but important task for developers of games or interactive storytelling applications. As depth cameras have become broadly available consumer products with the Microsoft Kinect sensor[1], opportunities emerge for such systems to make use of novel full body interaction techniques. This in turn poses new challenges for the interaction designer. Various researchers have already started to integrate this new kind of interaction in their interactive storytelling system [8, 1], but usually the gesture set for interacting with the system is chosen by the developers themselves according to their imagination and preferences. However, a gesture that is intuitive for the developers does not necessarily have to be intuitive for the majority of users. A different approach that creates a user-defined gesture set has been addressed by several other domains, such as surface computing [18], public displays [11], or human robot interaction [14].

In this paper we adapt the method for creating a user-defined gesture set for full body gestures in an interactive storytelling scenario. We describe the full process from conducting the user study, analysis of the recorded data, to the integration of the gesture set in our application.

[1] http://www.xbox.com/KINECT

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 264–281, 2013.

1.1 In-Game Actions and Input Gestures

In computer games and interactive storytelling exist several types of in-game actions users can trigger via interaction. Two common action types are: *navigation* and *dialogue*. Basic navigation includes changing the position and orientation. Dialogue actions are usually involved when embodied conversational agents exist in the scenario that users can speak with. There are also other action types, e.g. manipulation actions which change the properties of an object, but our work will focus on navigation and dialogue. Furthermore, we focus on full body gestures that users need to perform for triggering those two types of actions. For navigation, this might be a quite straight-forward choice, as body movements are also used for navigation in real-life. For dialogue the main interaction modality in real-life is speech and it therefore might seem a bit awkward to use body gestures for it. However, conversational gestures are used in real-life for emphasizing or enhancing speech utterances, and sometimes even to replace them, e.g. when performing a head nod instead of saying “yes”. In addition, we want to keep the variety of different input modalities as small as possible because full body interaction itself can be quite difficult for users that are used to mouse and keyboard interaction and Kurdyukova et al. [10] have also shown that too many different input modalities can be more distracting than engaging for users. Overall, interactive storytelling applications can include a huge variety of specific actions for navigation and dialogue. It therefore seems quite impossible to find a generic set of actions for those two types. Instead, we investigate an action set created for our specific scenario. Nevertheless, it should represent a combination of actions typical to interactive storytelling scenarios and we have the hope that our findings also apply to other scenarios without major differences.

2 Related Work

In the following, we first describe work on classifying human gestures that helps us build our gesture taxonomy. Afterwards, we summarize an influential approach for acquiring a user-defined gesture set that we will adapt and use for our own work. The last area of research we analyze is that of full body gesture recognition, as we eventually aim to implement the recognition of our user-defined gesture set.

2.1 Human Gestures

There exist different taxonomies for the classification of conversational human gestures. One of the first was introduced by Efron [3] who presented five categories: *physiographics*, *kinetographics*, *ideographics*, *deictics*, and *batons*. Further, Kendon [5] tried to link his gesture taxonomy to the relation with speech and defined the following categories: *gesticulation*, *language-like gestures*, *pantomime*, *emblems*, and *sign language*. Another popular taxonomy was proposed by McNeill [12] who presented five types of gestures: *iconic*, *metaphoric*, *deictic*, *cohesive*, and *beat* gestures. The three taxonomies have a considerable overlap in their covered concepts, and they all focus on hand gestures during a conversation. However, the properties they describe can also be used to categorize full body gestures for human computer interaction as in our case.

None of the three taxonomies perfectly suits this purpose, but we have decided to use the terminology by McNeill and therefore describe the items of this taxonomy in more detail in the following. *Iconic* and *metaphoric* gestures both try to convey information by visually depicting an icon. However, they do this at different abstraction layers: *iconic* gestures are more concrete and directly represent a physical, spatial or temporal property of a real-world referent, e.g. when moving two fingers to indicate somebody is walking. *Metaphoric* gestures refer to abstract properties of a referent. For example, somebody might depict a container to refer to the contents of a story. *Deictic* gestures are pointing gestures that indicate a position or direction. The last two types of gestures do not convey meaning but accompany speech to emphasize parts of it (*beat gestures*) or to keep up the continuity (*cohesive gestures*). Therefore, those two types are less suited for direct interaction, whereas the other three categories seem to be good candidates for representing in-game actions. That is why we use them in our later described taxonomy. McNeill also defined three phases of a gesture: *preparation*, *stroke*, and *retraction*. *Preparation* is the phase in which the body is brought from its rest to a position that is suitable for executing the gesture. The *stroke* phase contains the main part of the gesture, while in the *retraction* phase the body is brought back to its rest. In terms of gesture recognition, the *stroke* phase is the most important part of a gesture as it contains the actual information.

2.2 User Defined Gestures

In the last couple of years, gestural interfaces have become more and more popular, ranging from small multitouch phones to large interactive tables or walls, but also to freehand and full body interaction using a depth sensor as the Microsoft Kinect. Nevertheless, gesture sets are often designed without sufficiently taking into accounts the preferences, habits, and needs of the actual users. For this reason, several researchers started to involve the user into the design process of the gestural interaction. Wobbrock et al. [18] presented a seminal approach to develop intuitive gestures for surface interfaces. They gathered a gesture set by presenting the wanted effect within the system to the user, and then asking the user to perform a gesture that should trigger this effect. After that, they determined gesture candidates by looking at all gestures performed by the users for a specific effect, and calculating an agreement score based on how often the same gestures were used. Researchers already adopted this process for other areas, e.g. Kurdyukova et al. [11] used it to design gestures for transferring data between tablet computers and a multi-display environment, and Obaid et al. [14] applied the process to create a full body gesture set for the navigational control of humanoid robots. In this paper, we adapt the process by Wobbrock et al. to identify intuitive gestures for an interactive storytelling scenario.

2.3 Full Body Gesture Recognition

A lot of work was already done in the field of gesture recognition in general, and different approaches were developed and extensively tested, including statistical classifiers [15], Hidden Markov Models [16], dynamic programming [13], and many more. Recognition algorithms of this kind were used in various application areas, although most of those methods are quite complicated to implement, are computational expensive, or need to be trained with a lot of example data [19]. An effort to make gesture recognition more accessible to developers without a strong background in pattern recognition is the $1 recognizer presented by [19]. However, this approach was – as the aforementioned ones – still targeted at gestures defined by the movement of a single point in 2D space. It also relied on manual data segmentation, so the start and end of a gesture had to be indicated to the algorithm, either in an implicit way, e.g. by touching the surface of a PDA with the pen [19], or in an explicit way, e.g. by pressing a button on a Wiimote [9]. There have also been approaches to get rid of those shortcomings by porting them to 3D space [9], or removing the need for manual data segmentation [20]. However, the research for those methods is partly still in an early phase, and they are quite complicated to apply for practitioners as well.

The release of the Kinect sensor again motivated researchers to investigate gesture recognition for the more complex motion capturing data provided by full body tracking. The challenge with this data is that it contains multi-point data in 3D space (position and orientation of multiple important joints of one or more users). In addition, there is no obvious way to apply manual data segmentation (no device in the users' hands to press a button on), and the data itself also is rather noisy [6]. On the other hand, this means that the data itself already contains more information as in other interaction modalities. In this way, one single data frame for one tracked user already can be seen as a gesture, or more precisely a posture, as it defines a specific configuration of the user's skeleton. For this reason, researchers developed easy to use techniques for full body gesture recognition in application prototypes. One of the first was the Flexible Action and Articulated Skeleton Toolkit (FAAST) [17] that bound gestures defined by simple text scripts to key and mouse events, and therefore enabled to control arbitrary applications via full body interaction. We developed a similar approach in our FUBI framework[2] of which an earlier version was already presented in [7]. It tries to achieve more powerful gesture recognition by giving more complex configuration options in an XML-based definition language. We use this framework to implement the user-defined body gestures in section 4.

3 Interactive Storytelling Scenario and Gesture Study

In this section, we describe the scenario of our application and its intended user interaction. We further explain our interaction study and present the results of its analysis.

3.1 Scenario and User Interaction

Our interactive storytelling scenario aims to provide intercultural training for young adults (18-25 year olds). The users learn by participating actively in the narrative in which they have to interact with virtual characters from different cultures. However, the characters do not represent real cultures, but synthetic ones as defined by Hofstede [4]. The users adopt the role of a character that has not traveled too much for most of his life. The scenario starts at the café of the character's grandmother, in which he

[2] http:/www.hcm-lab.de/fubi.html

receives a letter from his deceased grandfather. In this letter, the grandfather, who liked to travel the world, promises the grandson a "lost treasure" that he should find in a journey through different countries. In each country the grandson has to interact with locals in so-called critical incidents to progress. To be successful, the users have to select the correct interaction options depending on the agents' simulated synthetic culture. The selection itself should eventually be done by performing a corresponding full body gesture. In the final country the users will find out that the promised treasure is the experience that the grandfather had while travelling. The scenario is implemented using the cross-platform game engine Unity3D[3] and an agent architecture for culturally adaptable behaviors [2].

We conducted our interaction study for the introduction in the café and the first country of our scenario which at this state included two critical incidents. The users' first task was to find out the way to their hotel by interacting with people in a bar (first critical incident, cf. Fig. 1 left-hand side). In the subsequent incident users had to find the responsible supervisor in a nearby museum in order to receive entry permission for a park (second critical incident, cf. Fig. 1 right-hand side). The scene in the grandmother's café and the mentioned two critical incidents together included the following in-game actions to be triggered by the users: *yes, no, sit at bar and wait, approach group, ask for directions, leave bar, ask about supervisor, ask guard to talk to supervisor, approach supervisor, ask permission.*

Fig. 1. Virtual environment of the two investigated critical incidents

3.2 Study Setup, Procedure, and Participants

The experiment was arranged in a room of about 3 meters width and 6.5 meters depth. The participants were standing at a distance of about 2.5 meters in front of a 50 inch plasma display. A camera was placed in a height of about 1.5 meters left of the display to record the users' front from a slightly tilted view. The participants were told that they should place themselves at the initial position, but that they were still allowed to freely move within the camera's field of view during the study. The experimenter was sitting to the left of the participant and controlling the application running on the display via mouse and keyboard.

[3] `http://unity3d.com`

After a short introduction and a demographic questionnaire that also included a question about the users' experience with body gesture based interaction, the experimenter explained the participants their role in the study. The experimenter ran through the story script of the application and as soon as a user input would have been requested by the application, text boxes with the currently available in-game actions were displayed as overlays on the virtual scene as depicted in Fig. 1. At this point, the participants' task was to invent and perform a gesture for each displayed action, one after the other. The participants were told that they were allowed to use their full body for gesturing, but that the gesture itself should mainly be intuitive for them to trigger the requested action. It should, however, have a semantic relation to the action and not consist of simply pointing towards the action label on screen. To keep the process as reproducible as possible, the experimenter always spoke out the action that the user should investigate next and also gave a short explanation about the meaning of the action to avoid misunderstandings. After performing their invented gesture, the participants should indicate on a questionnaire how easy it was for them to come up with that gesture on a 7-point Likert scale.

22 participants took part in the study including 4 females and 18 males. Their age ranged from 22 to 35 with an average of 26.23 (SD 3.80). All except for one were right-handed. The participants were recruited from our university campus and therefore all had a computer science background. They stated themselves a medium experience with body gesture based interaction of 2.18 (SD 0.85) on a scale from 0 (no experience) to 4 (practically daily usage).

3.3 Results

The next chapters depict the results of our study, including our gesture taxonomy, a description of the gesture set, user ratings, agreement scores and time performances.

Table 1. Full body gesture taxonomy for our interactive storytelling scenario

Form	static gesture	A static body gesture is held after a preparation phase.
	dynamic gesture	The gesture contains movement of one or more body parts during the stroke phase.
Body parts	one hand	The gesture is performed with one hand.
	two hands	…with two hands.
	full body	…with at least one other body part than the hands.
Gesture type	deictic	The gesture is indicating a position or direction.
	iconic	The used gesture visually depicts the meant in-game action or a part of it directly.
	metaphoric	The gesture visually depicts an icon and describes the in-game action in an abstract way.

Gesture Taxonomy

The recorded videos were analyzed and annotated using the ELAN annotation tools[4] to extract the stroke phases of all gestures performed by the study participants for each in-game action. We manually classified all performed gestures according to three dimensions: *form, gesture type,* and *(involved) body parts*. Each dimension consisted of multiple items as shown in Table 1. The dimensions were based on the taxonomy we had used in [14], we only left out the view-point dimension, as our interactive storytelling scenario always kept the user in a first-person perspective and there were no changes in the view-point for this reason. Furthermore, we changed the wording of the *gesture type* dimension to be closer to McNeill [12].

Fig. 2 displays the overall taxonomy distribution for the 251 performed gestures. The frequency of static and dynamic gestures was quite similar. Users tended to perform few deictic gestures, but more metaphoric ones. They only seldom chose two hand gestures, but roughly an equal number of one hand and full body gestures. The gestures we categorized as iconic according to McNeill [12] in fact were very concrete, which means that most of them were directly miming the meant in-game action, e.g. *approach group* was often expressed by actually walking a step forward.

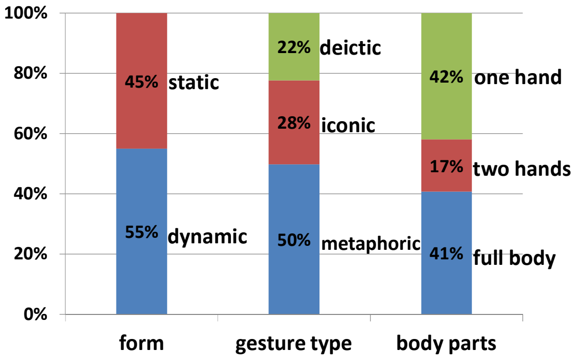


Fig. 2. Overall taxonomy distribution

Gesture Set

We selected suitable gesture candidates for each in-game action as follows:

1. For each action a we identified a set $M(a)$ containing all proposed gestures.
2. The gestures in $M(a)$ were then grouped into subsets of identical gestures $M_i(a)$ with $i \in 1..n_a$ and n_a being the total number of identified subsets for action a.
3. The first candidate $c_1(a)$ was determined by the largest subset $M_i(a)$, i.e.:

$$c_1(a) = MAX_{i\in 1..n_a}(M_i(a)). \tag{1}$$

4. The second candidate $c_2(a)$ was determined by the second largest subset $M_i(a)$:

$$c_2(a) = MAX_{i\in 1..n_a,\, M_i(a)\neq c_1(a)}(M_i(a)). \tag{2}$$

[4] `http://www.lat-mpi.eu/tools/elan/`

There can been more than one first or second candidate if there are multiple largest subsets, but this was not the case for our data. Further, we only took the second candidate into account if the size of the corresponding subset was at least half the size of the subset of the first candidate. This was done to ensure that the considered candidates always had been performed by a relevant percentage of users.

Table 2. Gesture candidates for each action

In-game action	Gesture candidates	Occurrences	Form	Gesture type	Body parts
yes	head nod	68%	dynamic	metaphoric	full body
no	head shake	68%	dynamic	metaphoric	full body
sit at bar and wait	sit down	56%	static	iconic	full body
approach group	step forward	56%	dynamic	iconic	full body
ask for directions	arms out	34%	static	metaphoric	two hands
leave bar	turn away	45%	dynamic	iconic	full body
	step backward	27%	dynamic	iconic	full body
ask about supervisor	arms out	50%	static	metaphoric	two hands
ask guard to talk to supervisor	point at one after the other	38%	static	deictic	one hand
	point to front	21%	static	deictic	one hand
approach supervisor	step forward	56%	dynamic	iconic	full body
ask permission	arms out	23%	static	metaphoric	two hands
	tip on shoulder	19%	dynamic	iconic	one hand

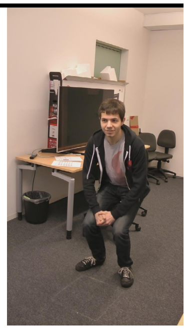
Table 2 summarizes the gesture candidates for all ten in-game actions. The third column includes the percentage of how often this candidate was performed among all gestures proposed for this action, and the last three columns depict the candidate's taxonomy. A second candidate was taken into account only in three cases (*leave bar*, *ask guard to talk to supervisor*, and *ask permission*).

The gesture candidates are further exemplified by images of users performing them in Table 3. For the actions *yes* and *no*, most users chose a *head nod* or *head shake* as gestures. The action *sit at bar and wait* was in most times represented by actually adopting to a sitting position (=*sit down*). Similarly, we found gesture candidates that represented the action quite directly for *approach group*, *leave bar*, and *approach supervisor*, in which the users did a *step backward* or a *step forward*. For the action *leave bar* a second gesture candidate was *turn away* which meant the user actually turned around as if going away. *Ask permission* was additionally expressed by the gesture *tip on shoulder* that was chosen because the supervisor – that participants should ask for permission to enter a park – stood there with the back to them (cf. Fig. 1 right-hand side: the virtual character at the back), so they assumed they first needed to get his attention. For the ask actions we often got the gesture *arms out* that always included moving the arms to an outward position with open hands, often accompanied by raising the shoulders. The only action for which the gesture candidates were pointing gestures was *ask guard to talk to supervisor*. Participants either chose to point in the direction of the supervisor (*point to front*), or to point at the guard first and only afterwards to the supervisor (*point at one after the other*).

Table 3. User images of the gesture candidates

head nod

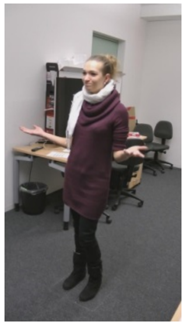
head shake



sit down

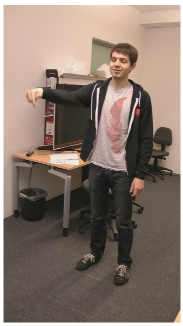
step forward (from left to right) / *backward* (from right to left)

turn away



arms out

point to front (left image) / *to one after the other* (both images)



tip on shoulder

User Ratings

Fig. 3 depicts the average user ratings for the easiness to invent the gestures for the 10 in-game actions on a scale from 0 (very hard) to 6 (very easy). Error bars represent the standard error. The actions are ordered according to their user rating.

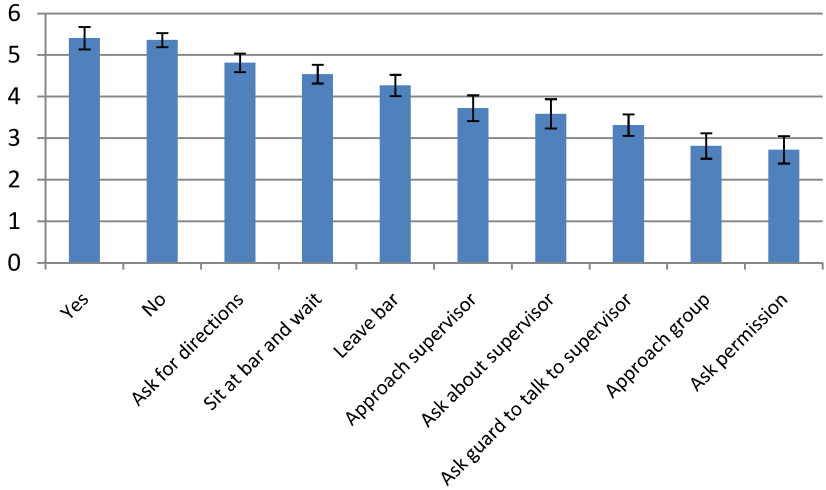


Fig. 3. User difficulty ratings of the 10 actions

A one-way repeated measures ANOVA revealed that the ratings differed significantly between the different actions with $F(9, 21) = 15.90$, $p < 0.01$, $\eta^2 = 0.43$. Post-hoc tests with Bonferroni correction showed that the more complex conversational actions were perceived as more difficult to invent a gesture for them. Accordingly, all actions that include asking character(s) something were rated as the most difficult ones. In particular, all those actions were rated significantly lower ($p < 0.01$) than the actions *yes* and *no. Approach supervisor* and *leave bar,* were also rated significantly higher ($p < 0.05$) than the ask actions, except for *ask permission. Approach group* was only significantly higher rated ($p < 0.05$) than *ask about supervisor*. We found no significant difference between *sit at bar and wait* and the ask actions, and there was also no significant difference between the ratings of the different ask actions.

Agreement Scores

To further investigate the level of agreement among the participants, we calculated an agreement score based on the process defined and used by Wobbrock et al. [18]. For an action *a*, the agreement score $AS(a)$ is defined as:

$$AS(a) = \sum_{i \in 1..n_a} \left(\frac{|M_i(a)|}{|M(a)|}\right)^2 \quad (3)$$

An agreement score $AS(a)$ for an action a is represented by a number in the range $[1\ /\ |M(a)|\ , 1]$, with a higher value corresponding to a higher agreement, and 1 representing a perfect agreement (all participants chose the same gesture for this action).

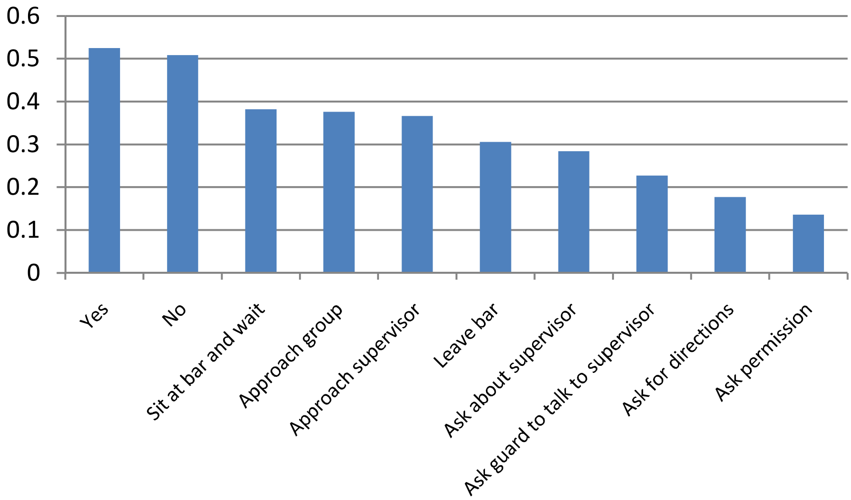


Fig. 4. Agreement scores for the 10 actions

The overall agreement we got for our action set was 0.329 (SD 0.129). Fig. 4 depicts the agreement scores of the different actions. They are ordered from the action with the highest agreement to the one with the lowest agreement. Similar to the user ratings, more complex actions (i.e. all "ask …" actions) caused lower agreement between the users and we got a large number of different gestures for those.

In fact, the results reveal that the level of agreement between the participants strongly correlates to the easiness to invent gestures (Pearson's $r = 0.812, p < 0.01$). When the participants thought it was easy to find a gesture for an in-game action, more participants chose the same gestures, and in opposite, when the participants thought it was difficult to invent a gesture for an action, we got a higher variation of gestures as well.

Timings

Table 4 depicts the times it took the users to perform one stroke of the gesture candidates. One stroke means e.g. for the head nod gesture that the user moves the head from the resting position upwards, then downwards under the resting position, and upwards to the resting position again. In other words, one stroke consists of the minimal gesture that can be found when dividing the gesture into equal sub gestures. For static gestures, one stroke consists only of a hold phase in which the user holds the relevant posture until moving back to a resting position. The table enlists the average times as well as the standard deviation, the minimum, and maximum among the different gesture performances.

Table 4. Times for one stroke of the gesture candidates

in-game action	gesture candidates	mean time	SD	MIN	MAX
yes	head nod	0.822	0.374	0.420	1.560
no	head shake	0.687	0.233	0.474	1.420
sit at bar and wait	sit down	0.661	0.635	0.120	1.970
approach group	step forward	1.508	0.702	0.422	2.770
ask for directions	arms out	0.707	0.392	0.295	1.375
leave bar	turn away	1.738	0.908	0.557	3.400
	step backward	2.041	0.593	1.295	3.000
ask about supervisor	arms out	0.589	0.341	0.200	1.280
ask guard to talk to supervisor	point at one after the other	1.013	0.257	0.540	1.410
	point to front	0.560	0.558	0.130	1.625
approach supervisor	step forward	1.444	0.659	0.517	2.470
ask permission	arms out	0.759	0.578	0.190	1.770
	tip on shoulder	0.356	0.066	0.257	0.410

4 Integration of the Gesture Set

After collecting the data for our gesture set, we aimed to enable the gestural interaction in our application. In the following, we describe the recognition framework we used to achieve this task, the implementation of the gesture recognizers, and how we integrated them into our interactive storytelling scenario.

4.1 Recognition Framework

For implementing the recognition of the gesture candidates in our application, we used our framework for full body interaction (FUBI) [7]. The recognition framework uses the Kinect for Windows SDK[5] for getting the depth data provided by the Kinect sensor, and for applying full body user tracking on that data. In this way, we get positions and orientations for 20 different user joints. The joint data is analyzed in the recognition framework for detecting gestures that are defined via XML files. Those XML files first can contain three types of basic gesture recognizers:

1. *Joint orientation* recognizers are defined by a minimum and/or a maximum angle for a specific joint.
2. *Joint relation* recognizers look at the position of a joint in relation to a second one, e.g. whether and how much a joint is above a second one or how large the distance between those two is.
3. *Linear movement* recognizers are defined by a specific direction and a minimum and/or maximum speed.

[5] `http://www.microsoft.com/en-us/kinectforwindows/`

In addition, those three types of basic recognizers can be combined to sequences in so-called *combination* recognizers. A *combination* recognizer consists of several states that contain sets of the above mentioned basic recognizers. For each state there is a minimum and maximum duration defined that those recognizers have to be fulfilled for the recognition process to get into and stay in this state. For the transition to the next state there is also a maximum duration defined. Fig. 5 depicts the XML definition for a *combination* recognizer that is meant to recognize a head nod gesture.

```xml
<!--HeadNod-->
<JointOrientationRecognizer name="HeadDown">
  <Joint name="head"/>
  <MaxDegrees x="-13"/>
</JointOrientationRecognizer>
<JointOrientationRecognizer name="HeadUp">
  <Joint name="head"/>
  <MinDegrees x="-5"/>
</JointOrientationRecognizer>
<CombinationRecognizer name="HeadNod">
  <State maxDuration="1" timeForTransition="0.4">
    <Recognizer name="HeadUp"/>
  </State>
  <State maxDuration="1" timeForTransition="0.4">
    <Recognizer name="HeadDown"/>
  </State>
  <State maxDuration="1" timeForTransition="0.4">
    <Recognizer name="HeadUp"/>
  </State>
  <State>
    <Recognizer name="HeadDown"/>
  </State>
</CombinationRecognizer>
```

Fig. 5. XML definition of the recognizer for a head nod

4.2 Integration of Gestures in the Application

For implementing the recognizers for the gesture candidates, one first has to determine by which sequence of basic recognizers a gesture candidate can be described and how the recognizers' parameters need to be adjusted. This is done by studying sample videos of the gesture in more detail and approximating the parameters. For determining the time constraints of a combination recognizer, the measured timings as depicted in Table 4 can be used. However, it has to be mentioned that in most cases, the timings cannot be used directly, but should rather serve as a basis for understanding a gesture and its temporal variance between the users.

To implement the gesture candidate *head nod* for the action *yes*, we used the XML definition as shown in Fig. 6. The definition includes two different *joint orientation* recognizers that observe the x angle of the head. Those *joint orientation* recognizers were combined in a sequence of four states. Note that all states do not have a minimum duration, so it would be sufficient if the recognizers of one state are fulfilled only for one frame of the tracking data stream. However, each state has a maximum duration of 1 second and a time for transition to the next state of 0.4 seconds. That means a head nod could include a head tilted upwards for at most 1 second, and then a movement downwards that lasts at most 0.4 seconds follows. The tilted downwards head could be hold for up to 1 second. After that follows an upwards movement for again at most 0.4 seconds, and the whole sequence needs to be repeated a second time. Therefore, a head nod is recognized as soon as two single nods are detected. This was sufficient for our purpose as we were not interested in distinguishing different kinds or numbers of nods.

The rest of the gesture candidates were implemented in a similar way. We used *combinations* of joint orientation recognizers for the gesture candidates *head shake*, *sit down*, and *turn away* to check the orientations of the joints included in the gesture. For the candidate *arms out* we used a recognizer that combines two *joint relation* recognizers in one state. The first *joint relation* recognizer observed the left hand and shoulder and waited for the hand to be in a height similar to the shoulder (y difference smaller than 30 cm) and that the hand was at least 35 cm left of the shoulder (using the x coordinate). The second recognizer looked for the same properties with the right hand and shoulder. The gesture candidate *point to front* as well used a *joint relation* recognizer for checking the hand position. In addition it ensured that the right elbow was not too far away from the line from the shoulder to the hand (at max 12 cm) and that the hand was not moving with more than 0.5 m/s. The last condition was implemented using a *linear movement* recognizer. The candidate *pointing at one after the other* was defined in the same way but with an additional state that allowed hand movement in between the two pointing states. We also implemented the second gesture candidate for the action *ask permission* that was the gesture *tip on shoulder*. This was basically realized the same way as point to front, but in addition, it waited for a sequence of *linear movements* in upward and downward direction of the hand. The gesture candidates *step forward* and *step forward* were implemented as *linear movement* recognizers looking at the torso joint and waiting for a movement in z or –z direction after a short phase of standing still.

The integration of the gestures in the application was done similarly to how we conducted the study. Instead of only displaying text boxes with the currently available actions, we additionally displayed symbols that visualized how the gestures for these actions should be performed. The symbols consisted of either a single image that is displayed constantly or multiple images that are displayed as an animation sequence. The Unity3D integration the FUBI framework only needs a recognizer definition that is named the same way as the image (sequence) that is used as a symbol for it. As soon as the symbol is displayed on screen, the recognition framework automatically checks the corresponding recognizer for users tracked with the depth sensor, and - if the recognition has been successful - it triggers an event related to the symbol. The event for each symbol can be defined in the same way as it is done for default interface buttons in Unity3D. Fig. 6 depicts a scene of the first critical incident, now displaying four symbols of the new gesture set that can currently be selected by the user. Although we implemented the recognizers for all gesture candidates, we had to decide which of the candidates we would actually use in the three cases in which two candidates were determined. As they seemed to fit a bit better to the parallel and surrounding gestures and because of their partly more reliable recognizers, we chose the gesture *turn away* for the action *leave bar*, *point to front* for *ask guard to talk to supervisor*, and *tip on shoulder* for *ask permission*.

Fig. 6. Gesture symbols in the first critical incident

We preliminarily tested the integration of the new gesture set in an informal study with the same setup as in the previous study, but this time a Kinect sensor was placed centered below the screen to enable the interaction. Three users took part in the study, two males and one female. They had already participated in the study for the creation of the gesture set, thus they were used to the setting. They again ran through the story script, but this time the gestural interaction was enabled, so they always had to perform one of the requested gestures to continue. We had to omit the part in which the two gestures *head nod* and *head shake* appeared due to (later solved) technical issues with the Kinect face tracking SDK. For the rest of the gestures, the detection already worked quite robust. Out of 15 performed gestures, 11 were directly recognized by the first attempt of the user, while four of them had to be executed a second time to be recognized by the system. Thus, we had four false negatives and 15 true positives, which can be summarized to a preliminary recognition rate of 79 %. In addition, most of the false negatives were due to a wrong gesture performance by the participant. This was caused by the fact that the gestures were not explained to the participants, but they had to guess how to perform them by looking at the symbols displayed on screen as depicted in Fig. 6. The system never recognized a wrong gesture, so there were no false positives. Overall, we got positive feedback about the integration, and the participants liked to interact by using the new gesture set.

5 Discussion

As far as the taxonomy distribution of our gesture set is concerned, we got quite different results in comparison to our previous work that investigated user-defined gestures for navigating a humanoid robot [14]. We got much less deictic gestures as in the previous work, but more metaphoric ones. This is due to the target of the gestural interaction. While deictic gestures seem to be more suitable for navigational actions, our action set also included conversational actions that need to be described with metaphoric gestures due to their increased semantic complexity. We also had a closer look at the taxonomy distribution of each action itself that revealed that the users never used iconic gestures for the conversational actions (all *ask actions* plus *yes* and *no*) except for the gesture *tip on shoulder* of the action *ask permission*. All other actions – which can be categorized as navigational actions – included all three types of gestures, except for *leave bar* for which we only observed iconic and deictic gestures, but no metaphoric ones. For further increasing the information content of their gestures, the participants more often used other body parts than their hands in opposite to the robot navigation task. However, they used less dynamic gestures, which indicates that full body gestures often provide enough information in a static version. As full body gestures were overall frequently chosen quite often, it can be said that this kind of gestures is worth to be used in interactive storytelling scenarios in general. It seemed that full body gestures are especially intuitive for triggering the in-game actions that occur in this kind of scenarios.

We proposed to select the gesture candidates according to how many users chose a gesture for one in-game action. However, this does not always have to be the best choice. For example, it makes no sense to give the user the choice between two actions represented by the same gesture at the same point in time. In this case it is better to select a less often chosen gesture candidate for at least one of the actions. There are also other cases in which it is helpful to select a different gesture, e.g. if the recognition software is not able to detect the gesture in a robust way. A more specific reason for doing this is also given in our application. As we aim at intercultural training, we want the users, at a later point in our scenario, to be confronted with gestures that are unfamiliar to them, as this can occur when travelling to different countries. For this purpose it might also be worth to conduct the study with participants of different cultural background to get a different gesture set.

Another challenge we faced was the problem of potentially too complex in-game actions, and especially the difficulty to represent verbal actions with non-verbal gestures. This can be seen in the relatively low agreement scores and user ratings we got for all actions that involve asking virtual characters about something. At the state of the study, our scenario never included multiple ask actions in parallel, so we had no problems with their ambiguous gestures, but as soon as this occurs, it might be necessary to refine the actions for making them easier to be related to gestural input. Otherwise, it might also be necessary to include different kinds of interaction, e.g. using a graphical user interface or other modalities such as speech input.

As described in section 4, it was feasible for us to implement a prototypic recognition logic for the gathered gesture set within our recognition framework [7] and to integrate the gestures in our storytelling scenario. The recorded videos of the gesture performances were very useful for this task as well as the measured timings for the gestures' stroke phases. Nevertheless, the creation of a gesture recognizer based on this data is still a challenging task that has to be done in a careful way to realize the gestures at least close to as they were intended by the users.

6 Conclusion and Future Work

In this paper, we presented how we produced a user-defined gesture set for an interactive storytelling scenario. We described the full process from conducting the study, over analysis of the data, to the implementation of the gesture candidates. During this process we obtained a taxonomy of full body gestures for our interaction set, user ratings and agreement scores for each in-game action, the time performances of all gesture candidates, and we finally integrated the gesture candidates in our applications using our open source full body interaction framework FUBI [7] .

A first validation for FUBI according to accuracy and usability was already done with a different interactive storytelling scenario [8] that included different kinds of iconic gestures. We plan to conduct a similar study with the new scenario in order to provide a more complete validation, also with more abstract metaphorical gestures. At the moment, we investigated the already implemented first part of our scenario. As soon as more of the application is ready, we want to continue the study with the additional in-game actions. In the meantime, it became clear that our scenario will have interactions that sometimes include multiple conversational actions in parallel that we will probably not be able to represent with unambiguous gestures. Therefore, we plan to include an additional interaction with a graphical user interface for those cases in our application.

Acknowledgments. This work was funded by the European Commission within FP7 under grant agreement eCUTE (FP7-ICT-257666). We would like to thank our project partners for the collaboration, and especially Nick Degens, Hélio Mascarenhas, Samuel Mascarenhas, and André Silva for their work on the storyboard, agent architecture, and asset creation that formed the background for our study.

References

1. Álvarez, N., Peinado, F.: Exploring body language as narrative interface. In: Oyarzun, D., Peinado, F., Young, R.M., Elizalde, A., Méndez, G. (eds.) ICIDS 2012. LNCS, vol. 7648, pp. 196–201. Springer, Heidelberg (2012)
2. Dias, J., Mascarenhas, S., Paiva, A.: Fatima modular: Towards an agent architecture with a generic appraisal framework. In: Proc. of the Int. Workshop on Standards for Emotion Modeling (2011)
3. Efron, D.: Gesture and Environment. King's Crown Press, Morningside Heights (1941)
4. Hofstede, G.J.: Role playing with synthetic cultures: the evasive rules of the game. Experimental Interactive Learning in Industrial Management: New approaches to Learning, Studying and Teaching, 49 (2005)
5. Kendon, A.: How gestures can become like words. In: Cross-cultural Perspectives in Nonverbal Communication, Hogrefe, pp. 131–141 (1988)
6. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors 12(2), 1437–1454 (2012)
7. Kistler, F., Endrass, B., Damian, I., Dang, C., André, E.: Natural interaction with culturally adaptive virtual characters. Journal on Multimodal User Interfaces 6, 39–47 (2012)

8. Kistler, F., Sollfrank, D., Bee, N., André, E.: Full body gestures enhancing a game book for interactive story telling. In: André, E. (ed.) ICIDS 2011. LNCS, vol. 7069, pp. 207–218. Springer, Heidelberg (2011)
9. Kratz, S., Rohs, M.: A $3 gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors. In: Proc. IUI 2010, pp. 341–344. ACM, New York (2010)
10. Kurdyukova, E., André, E., Leichtenstern, K.: Introducing multiple interaction devices to interactive storytelling: Experiences from practice. In: Iurgel, I.A., Zagalo, N., Petta, P. (eds.) ICIDS 2009. LNCS, vol. 5915, pp. 134–139. Springer, Heidelberg (2009)
11. Kurdyukova, E., Redlin, M., André, E.: Studying user-defined iPad gestures for interaction in multi-display environment. In: Proc. IUI 2012, pp. 1–6 (2012)
12. McNeill, D.: Head and Mind: What Gestures Reveal About Thought. University of Chicago Press (1992)
13. Myers, C.S., Rabiner, L.R.: A comparative study of several dynamic time-warping algorithms for connected-word recognition. The Bell System Technical Journal 60(7), 1389–1409 (1981)
14. Obaid, M., Häring, M., Kistler, F., Bühling, R., André, E.: User-defined body gestures for navigational control of a humanoid robot. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) ICSR 2012. LNCS, vol. 7621, pp. 367–377. Springer, Heidelberg (2012)
15. Rubine, D.: Specifying gestures by example. SIGGRAPH Comput. Graph. 25(4), 329–337 (1991)
16. Sezgin, T.M., Davis, R.: HMM-based efficient sketch recognition. In: Proc. IUI 2005, pp. 281–283. ACM, New York (2005)
17. Suma, E., Lange, B., Rizzo, A., Krum, D., Bolas, M.: FAAST: The flexible action and articulated skeleton toolkit. In: Proc. VR 2011, pp. 247–248 (2011)
18. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092. ACM, New York (2009)
19. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proc. UIST 2007, pp. 159–168. ACM, New York (2007)
20. Zinnen, A., Schiele, B.: A new approach to enable gesture recognition in continuous data streams. In: Proc. ISWC 2008, pp. 33–40. IEEE Computer Society, Washington, DC (2008)

User-Defined Gestures for Augmented Reality

Thammathip Piumsomboon[1,2], Adrian Clark[1], Mark Billinghurst[1], and Andy Cockburn[2]

[1] HIT Lab NZ, University of Canterbury, Christchurch, New Zealand
[2] Department of Computer Science and Software Engineering, University of Canterbury, Christchurch, New Zealand
Thammathip.Piumsomboon@pg.Canterbury.ac.nz, {Adrian.Clark,Mark.Billinghurst,Andy}@Canterbury.ac.nz

Abstract. Recently there has been an increase in research towards using hand gestures for interaction in the field of Augmented Reality (AR). These works have primarily focused on researcher designed gestures, while little is known about user preference and behavior for gestures in AR. In this paper, we present our guessability study for hand gestures in AR in which 800 gestures were elicited for 40 selected tasks from 20 participants. Using the agreement found among gestures, a user-defined gesture set was created to guide designers to achieve consistent user-centered gestures in AR. Wobbrock's surface taxonomy has been extended to cover dimensionalities in AR and with it, characteristics of collected gestures have been derived. Common motifs which arose from the empirical findings were applied to obtain a better understanding of users' thought and behavior. This work aims to lead to consistent user-centered designed gestures in AR.

Keywords: Augmented reality, gestures, guessability.

1 Introduction

By overlaying virtual content onto the real world, Augmented Reality (AR) allows users to perform tasks in the real and virtual environment at the same time [1]. Natural hand gestures provide an intuitive interaction method which bridges both worlds. While prior research has demonstrated the use of hand gesture input in AR, there is no consensus on how this combination of technologies can best serve users. In studies involving multimodal AR interfaces, hand gestures were primarily implemented as an add-on to speech input [2] [3]. In cases of unimodal gesture interfaces, only a limited number of gestures have been used and the gestures were designed by researchers for optimal recognition rather than for naturalness, meaning that they were often arbitrary and unintuitive [4] [5] [6]. Recent research has integrated hand tracking with physics engines to provide realistic interaction with virtual content [7] [8], but this provides limited support for gesture recognition and does not take into account the wide range of expressive hand gestures that could potentially be used for input commands.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 282–299, 2013.

To develop truly natural gesture based interfaces for AR applications, there are a number of unanswered questions that must be addressed. For example, for a given task is there a suitable and easy to perform gesture? Is there a common set of gestures among users that would eliminate the need for arbitrary mapping of commands by designers? Is there a taxonomy that can be used to classify gestures in AR? Similar shortcomings were encountered in the fields of surface computing and motion gestures, where Wobbrock et al. [9] and Ruiz et al. [10] addressed absences of design insight by conducting guessability studies [11].

In this study, we focus explicitly on hand gestures for unimodal input in AR. We follow Wobbrock's approach and employ a guessability method, first showing a 3D animation of the task and then asking participants for their preferred gesture to perform the task. Users were also asked to subjectively rate their chosen gestures, based on the perceived "goodness" and ease to perform. By analyzing the results of this study, we were able to create a comprehensive set of user-defined gestures for a range of tasks performed in AR.

This work makes a number of contributions; (1) The first set of user-defined gestures captured from an AR interface, (2) Classification of these gestures based on a gesture taxonomy for AR which was extended from Wobbrock's surface gesture taxonomy [9], (3) Agreement scores of gestures for selected tasks and subjective rating of the gestures, (4) Qualitative findings from the design process, and (5) Discussion of the implications of this work for AR, gesture interfaces, and gesture recognition.

2 Related Work

The topic of hand gesture classification as based on human discourse was excellently covered by the work of Wobbrock et al. [9]. As our work extends this approach to gesture interaction in AR, we focus on related work in bare hand and glove-based unimodal hand gestures interfaces, multimodal interfaces coupled with speech, and recent advancements in AR relevant to interaction. In addition, we briefly discuss previous research that utilized elicitation techniques.

2.1 Hand Gesture Interfaces in AR

Lee et al. [12] designed gloves with conductive fabric on the fingertips and the palm for gesture recognition and vibration motors for haptic feedback. The gloves were tracked using markers placed around the wrist area. A small set of gestures were used to allow selection, gripping, cutting and copying actions.

Lee and Hollerer [5] created Handy AR, a system capable of bare hand interaction using a standard web camera. The supported gestures were limited to an opened/closed hand for object selection and hand rotation for object inspection. Their follow up work allowed objects to be relocated using markerless tracking [13].

Fernandes and Fernandez [6] trained statistical models with hand images to allow bare hand detection. Virtual objects could be translated using the hand in a palm upwards orientation, while rotation and scaling along the marker plane was achieved using two handed pointing.

The main shortcoming of all these interfaces was that they only recognize a small number of gestures, and this gesture set was designed by the researchers for easy recognition. No support was provided for users to define their own gestures which were more comfortable or had contextual meaning.

2.2 Hand Gesture and Speech Interfaces in AR

SenseShapes [14] aimed to find spatial correlation between gestures and deictic terms such as "that", "here", and "there" in an object selection task. The user's hands were tracked using data gloves, and object selection was facilitated by a virtual cone projected out from the users' fingers. The region of interest was estimated based on speech, gaze projection and the pointing projection.

Heidemann et al. [2] demonstrated an AR interface which identified objects on a tabletop. Skin color segmentation was used to identify the user's index finger, allowing users to select virtual objects by pointing and make menu selections. Speech could also be used to issue information queries and interact with the 2D menu.

Kolsch et al. [3] created a mobile AR system that supported interaction by hand gesture, speech, trackball and head pose. Gesture recognition was implemented using HandVu, a computer vision-based gesture recognition library. They categorized tasks by dimensionality. For example taking a snapshot was defined as 0D, adjusting the focus region depth was 1D, using a pencil tool was 2D, and orienting virtual objects was 3D. Some actions such as relocating/resizing/orienting could be performed multimodally by speech, gesture or trackball, while other actions such as take/save/discard snapshot could only be performed by speech.

The work most closely related to ours was that of, Lee [15], who conducted a Wizard of Oz study of an AR multimodal interface to measure the types of gestures people would like to use in a virtual object manipulation task. In this study pointing, translation and rotation gestures were captured. She later developed a multimodal gesture and speech interface for a design related task, however speech was used as a primary input as in typical multimodal systems therefore gestures were only mapped to limited number of spatial related tasks for example pointing, grabbing and moving.

2.3 Recent Advancements in AR Technology

Advancements of markerless tracking algorithms and consumer hardware have enabled greater possibilities for gesture based AR interfaces. Modern vision-based tracking algorithms can robustly register the environment without markers, allowing for higher mobility [16]. Furthermore, an introduction of consumer depth sensors such as the Microsoft Kinect has made real-time 3D processing accessible and have introduced a new interaction paradigm in AR through real-time physically-based natural interaction [7, 8].

2.4 Previous Elicitation Studies

Wobbrock et al. describe prior studies involving elicitation of input from users [9]. The technique is common in participatory design [17] and has been applied in a variety of research areas such as unistroke gestures [11], surface computing [9] and motion gesture for mobile interaction [10]. In AR, a Wizard of Oz study [15] for gestures and speech was conducted and aimed to capture the type of speech and gesture input that users would like to use in an object manipulation task. It was found that the majority of gestures used hand pointing due to reliance on speech for command inputs. In this research, our focus is to explore the potential of hand gestures as the unimodal input.

3 Developing a User-Defined Gesture Set

To elicit user-defined gestures, we first presented the effect of the task being carried out by showing a 3D animation in AR, and then asked the participants to describe the gestures they would use. Participants designed and performed gestures for forty tasks across six categories, which included gestures for three types of menu. Participants were asked to follow a think-aloud protocol while designing the gestures, and also to rate the gestures for goodness and ease to perform. They were asked to ignore the issue of recognition difficulty to allow freedom during the design process, and to allow us to observe their unrestricted behavior. At the end of the experiment, brief interviews were conducted and preferences of the three types of proposed gesture menus were collected.

3.1 Task Selections

In order for the gesture set to be applicable across a broad range of AR applications [18], we surveyed common operations in previous research e.g. [3, 5, 6, 12, 19], which resulted in forty tasks that included three types of gesture menu, which are horizontal [20], vertical [19], and object-centric that we proposed. These tasks were then grouped into six categories based on context, such that identical gestures could be used across these categories, as shown in Table 1.

3.2 Participants

Twenty participants were recruited for the study, comprising of twelve males and eight females, ranging in age from 18 to 38 with mean of 26 ($\sigma = 5.23$). The participants which were selected had minimal knowledge of AR to avoid the influence of previous experience with gestures interaction. Nineteen of the participants were right handed, and one was left handed. All participants used PCs regularly, with an average daily usage of 7.25 hours ($\sigma = 4.0$). Fifteen owned touchscreen devices, with an average daily usage of 3.6 hours ($\sigma = 4.17$). Eleven had experience with gesture-in-the-air interfaces such as those used by the Nintendo Wii or Microsoft Kinect gaming devices.

3.3 Apparatus

The experimental interaction space, shown in Figure 1 (Left), was the area on and above a 120 x 80cm table. Each participant was seated in front of the table, and a Sony HMZ-T1 head mounted display (HMD) at 1280 x 720 resolutions was used as the display device. A high definition (HD) Logitech c525 web camera was mounted on the front of the HMZ-T1 as a viewing camera, providing a video stream at the display resolution. This HMD and camera combination offered a wide field of view, with a 16:9 aspect ratio, providing a good view of the interaction space and complete sight of both hands while gesturing.

An Asus Xtion Pro Live depth sensor was placed 100 cm above the tabletop facing down onto the surface to provide reconstruction and occlusion between the user's hands and virtual content. An RGB camera was placed in front of and facing the user to record the users' gestures. A PC was used for the AR simulation and to record the video and audio stream from the user's perspective. A planar image marker was placed in the center of the table, and the OPIRA natural feature registration library [21] was used for registration and tracking of this marker. The 3D graphics, animation and occlusion were handled as described by Piumsomboon et al. [22].

Table 1. The list of forty tasks in six categories

Category		Tasks	Category		Tasks
Transforms	Move	1. Short distance	Editing		21. Insert
		2. Long distance			22. Delete
		3. Roll (X-axis)			23. Undo
	Rotate	4. Pitch (Y-axis)			24. Redo
		5. Yaw (Z-axis)			25. Group
		6. Uniform scale			26.Ungroup
	Scale	7. X-axis			27. Accept
		8. Y-axis			28. Reject
		9. Z-axis			29. Copy
Simulation		10. Play/Resume			30. Cut
		11. Pause			31. Paste
		12. Stop/Reset	Menu	Horizontal	32. Open
		13.Increase speed		(HM)	33. Close
		14.Decrease speed			34. Select
Browsing		15. Previous		Vertical (VM)	35. Open
		16. Next			36. Close
Selection		17.Single selection			37. Select
		18.Multiple selection		Object-centric	38. Open
		19.Box selection		(OM)	39. Close
		20.All selection			40. Select

3.4 Procedure

After an introduction to AR and description of how to operate the interface, the researcher described the experiment in detail and showed the list of tasks to the participant. The forty tasks were divided into six categories, as shown in Table 1, and the participant was told they could choose to carry out the categories in any order, providing that there was no conflict between gestures within the same category. For each task, a 3D animation showing the effect of the task was displayed, for example, in the "Move – long distance" task, participants would see a virtual toy block move across the table. Within the same task category, the participant could view each task as many times as she/he needed. Once the participant understood the function of the task, she/he was asked to design the gesture they felt best suited the task in a think-aloud manner. Participants were free to perform one or two-handed gestures as they saw fit for the task (See Figure 1, Right).

Once the participant had designed a consistent set of gestures for all tasks within the same category, they were asked to perform each gesture three times. After performing each gesture, they were asked to rate the gesture on a 7-point Likert scale in term of goodness and ease of use. At the end of the experiment, a final interview was conducted, where participants were asked to rank the three types of menu presented (horizontal, vertical, and object-centric as shown in Figure 5) in terms of preference and the justification for their ranking. Each session took approximately one to one and a half hours to complete.

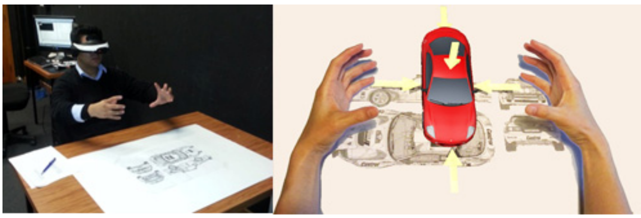


Fig. 1. (Left) A participant performs a gesture in front of the image marker. (Right) The participant sees an AR animation of a shrinking car, and performs their gesture for a uniform scale task.

4 Result

A total of 800 gestures were generated from the 20 participants performing the 40 tasks. The data collected for each user included video and audio recorded from the camera facing towards the user and the user's viewpoint camera, the user's subjective rating for each gesture, and transcripts taken from the think-aloud protocol and interview.

4.1 Taxonomy of Gestures in AR

We adapted Wobbrock's surface taxonomy [9] to better cover the AR gesture design space by taking their four-dimensional taxonomy, (*form*, *nature*, *binding*, and *flow)* and extending it with two more dimensions; *symmetry* and *locale*. Each dimension is comprised of multiple categories, as shown in Table 2.

The scope of the *form* dimension was kept unimanual, and in the case of a two-handed gesture, applied separately to each hand. In Wobbrock's original taxonomy, *form* contained six categories including *one-point touch* and *one-point path*, however, these two categories were discarded as they were not relevant to AR gestures that occur in three dimensional space.

The *nature* dimension was divided into *symbolic*, *physical*, *metaphorical* and *abstract* categories. Examples of symbolic gestures are thumbs-up and thumbs-down for *accept* and *reject*. *Physical* gestures were classified as those that would act physically on the virtual object as if it was a real object for instance grabbing a virtual block and relocating it for a *move* task. *Metaphorical* gestures express actions through existing metaphor e.g. pointing an index finger forward and spinning it clockwise to indicate *play* or *increase speed* as if one was playing a roll film. Any arbitrary gestures were considered *abstract*, such as a double-tap on the surface to *deselect* all objects.

The *binding* dimension considered relative location where gestures were performed. The *object-centric* category covered *transform* tasks such as *move*, *rotate*, and *scale*, as these are defined with respect to the objects being manipulated. Opening and closing horizontal or vertical menus were classified in the *world-dependent* category as they are located relative to the physical workspace. Gestures in the *World-independent* category could be performed anywhere, regardless of the relative position to the world, such as an open hand facing away from one's body to indicate *stop* during a simulation. Gestures performed across multiple spaces, such as *insert* where selection is *object-centric* and placement is *world-dependent,* fell into the *mixed dependencies* category.

In the *flow* dimension, gestures were categorized as *discrete* when the action is taken only when the gesture is completed, for example an index finger must be spun clockwise in a full circle to perform the *play* command. The gestures were considered *continuous* if the simulation must respond during the operation, such as manipulating an object using the *transform* gestures.

The first additional dimension we developed, *symmetry*, allowed classification of gestures depending on whether they were one handed (*unimanual*) or two handed (*bimanual*). The *unimanual* category was further split into *dominant* and *nondominant*, as some participants preferred to use their nondominant hand to perform gestures that required little or no movement, leaving their dominant hand for gestures requiring finer motor control. An example of this would be to use the dominant hand to execute a *selection*, and then use the non-dominant hand to perform a scissor pose for a *cut* operation. The *bimanual* category also subdivided, *symmetric* gestures representing two-handed gestures where both hands executed the same *form*, such as scaling, where both hands perform a pinch moving toward or away from each other. Two handed gestures, where the *form*s of the hands are different, fall into the *asymmetric bimanual* category. An example of this is the *copy (1)* gesture where one hand is used to select the target object (*static pose*) while the other hand drags the copy into place (*static pose and path*).

The other dimension we introduce is *locale*. If a gesture required physical contact with the real surface, they are considered *on-the-surface* as opposed to *in-the-air*. Gestures that require both are considered *mixed locales*. For example, an index finger tapped *on-the-surface* at a virtual button projected on the tabletop to perform horizontal menu selection task, as opposed to an index finger pushed *in-the-air* at a floating button to execute vertical menu selection. An example of a *mixed locales* gesture is, the *box selection (1)*, where one hand indicated the area of the bottom surface of the box by dragging an index finger diagonally along the table's surface, while another hand lifted off the surface into the air to indicate the height of the box (See Figure 5).

Table 2. Taxonomy of gestures in AR extended from taxonomy of surface gestures

Taxonomy of Gestures in AR		
Form	static pose	Hand pose is held in one location.
	dynamic pose	Hand pose changes in one location.
	static pose and path	Hand pose is held as hand relocates.
	dynamic pose and path	Hand pose changes as hand relocates.
Nature	Symbolic	Gesture visually depicts a symbol.
	physical	Gesture acts physically on objects.
	metaphorical	Gesture is metaphorical.
	abstract	Gesture mapping is arbitrary.
Binding	object-centric	Gesturing space is relative to the object.
	world-dependent	Gesturing space is relative to the physical world.
	world-independent	Gesture anywhere regardless of position in the world.
	mixed dependencies	Gesture involves multiple spaces.
Flow	Discrete	Response occurs after the gesture completion.
	continuous	Response occurs during the gesture.
Symmetry	dominant unimanual	Gesture performed by dominant hand.
	nondominant unimanual	Gesture performed by nondominant hand.
	symmetric bimanual	Gesture using both hands with the same form.
	asymmetric bimanual	Gesture using both hands with different form.
Locale	on-the-surface	Gesture involves a contact with real physical surface.
	in-the-air	Gesture occurs in the air with no physical contact.
	mixed locales	Gesture involves both locales.

4.2 Findings from Classification

Classification was performed on the 800 gestures as shown in Figure. 2. Within the six dimensional taxonomy, the most common characteristics of gestures were *static pose and path*, *physical*, *object-centric*, *discrete*, *dominant unimanual*, and *in-the-air*.

Within the *form* dimension, there was a slightly higher number of *static poses* (3%) performed with a *non-dominant* hand and lower for *static poses with path* gestures (2.5%) over a *dominant* hand. This slight discrepancy was contributed by some participants preferring to use their *dominant* hand for gestures with movement while using a *non-dominant* for a static pose.

In the *nature* dimension, overall the gestures were dominantly *physical* (39%) and *metaphorical* (34.5%). The gestures chosen to perform *transform*, *selection*, and *menu* tasks were predominantly *physical*, with the percentage of 76.1%, 50%, and 57.8% respectively. The *browsing* and *editing* task gestures were mainly *metaphorical* (100% and 40.9% respectively), while the *simulation* task, gestures were split across *symbolic* (37%), *metaphorical* (34%), and *abstract* (29%) categories. For the *binding* dimension, the majority of gestures for the *transform* and *selection* tasks were *object-centric* (100% and 75% respectively). *Simulation* (93%) and *browsing* (100%) task gestures were mainly *world-independent* (93% and 100%), while *editing* tasks gestures were *world-independent* (39.5%) and *object-centric* (32.3%). *Menu* tasks gestures were *object-centric* (50%) and *world-dependent* (45.6%).

For the remaining dimensions including *flow*, *symmetry*, and *locale*, the gestures chosen across all tasks were primarily *discrete* (77.5%), *dominant unimanual* (67.8%) and *in-the-air* (78%).

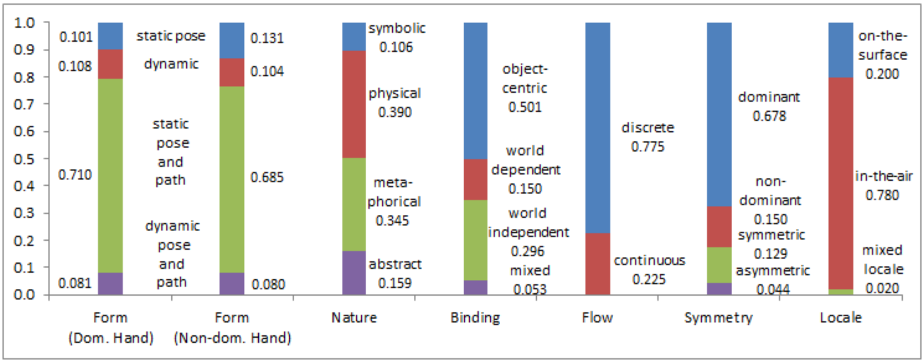


Fig. 2. The proportion of gestures in each category in the six dimensional taxonomy. *Form* has been calculated for each hand separately.

4.3 A User-Defined Gesture Set

As defined in prior work by Wobbrock et al. [9] and Ruiz et al. [10], the user defined gesture set, known as the "consensus set", is constructed based on the largest groups of identical gestures that are performed for the given task. In our study, each gesture valued at one point; therefore there were 20 points within each task and a total of 800 points for all tasks.

We found that participants used minor variations of similar hand poses, for example a swiping gesture with the index finger or the same swipe with the index and middle fingers, and therefore chose to loosen the constraints from "gestures must be identical within each group" to "gestures must be similar within each group". We defined "similar gestures" as *static pose and path* gestures that were identical or having consistent directionality although the gesture had been performed with different *static* hand poses.

We had classified the major variants of observed hand poses into 11 poses with the codes, *H01* to *H11*, as illustrated in Figure 4. For tasks where these variants existed, the variant poses could be used interchangeably, as indicated by the description under each user-defined gesture's illustration (Figure 5).

Exercising the "similar gesture" constraint, we were able to reduce the original 800 gestures into 320 unique gestures. The top 44 highly scored gestures were selected to make the consensus set, while the remaining 276 lowest scored gestures were discarded, defined by Wobbrock et al. [9] as the discarded set. The selected gestures of the consensus set represented 495 (61.89%) of the 800 recorded gestures (495 of 800 points). The consensus set of gestures comprised the overall task gestures in the following percentage *transform* (19.38%), *menu* (17.75%), *editing* (11.75%), *browsing* (5.00%), *selection* (4.63%), and *simulation* (3.38%), which sum up to 61.89%.

Level of Agreement. To compute the degree of consensus among the designed gestures, an agreement score *A* was calculated using Equation 1 [11]:

$$A = \sum_{P_s} \left(\frac{|P_s|}{|P_t|} \right)^2 \tag{1}$$

where P_t is the total number of gestures within the task, *t*, P_s is a subset of P_t containing similar gestures, and the range of *A* is [0, 1].

Consider the *rotate-pitch (y-axis)* task that contained five gestures with scores of 8, 6, 4, 1, and 1 points. The calculation for A_{pitch} is as follows:

$$A_{pitch} = \left(\frac{|8|}{|20|} \right)^2 + \left(\frac{|6|}{|20|} \right)^2 + \left(\frac{|4|}{|20|} \right)^2 + \left(\frac{|1|}{|20|} \right)^2 + \left(\frac{|1|}{|20|} \right)^2 = 0.295 \tag{2}$$

The agreement scores for all forty tasks are shown in Figure 3. While there is low agreement in the gestures set for tasks such as *all select*, *undo*, *redo* and *play*, there were notable groups of gestures that stood out with higher scores.

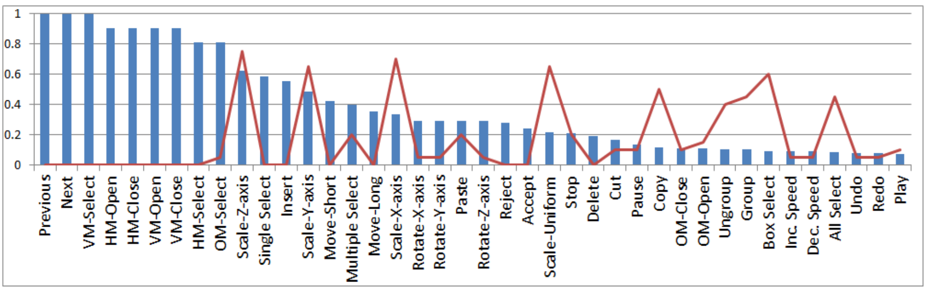


Fig. 3. Agreement scores for forty tasks in descending order (bars) and ratio of two-handed gestures elicited in each task (line)

User-Defined Gesture Set and Its Characteristics. As mentioned in Section 3.4, we allowed users to assign the same gesture to different tasks as long as the tasks were not in the same category. In addition to this, there were some tasks where there were two or more gestures commonly assigned by the participants. This non one-to-one mapping resulted in a consensus set of 44 gestures for a total of 40 tasks, which resulted in improved guessability [11].

When mapping multiple gestures to a single task, there was one task which had three gestures assigned to it (*uniform-scaling*), seven tasks had two gestures (*x, y, z scaling, box select, stop, delete, and copy*), and 23 tasks only had gesture. On the contrary, for the nine remaining tasks, two gestures were assigned to four tasks (*short, long move, insert, and paste*), one gesture assigned to three tasks (*play, increase speed, and redo*), and one gesture assigned to two tasks (*decrease speed and undo*).

When creating the consensus set, we only found one conflict between gestures within the same category. This was between the *pause* and *stop* tasks, where the gesture of an open-hand facing away from the body was proposed for both with scores of 4 and 7 points respectively. To resolve this, we simply assigned the gesture to the task with the higher score, in this case *stop*.

Play and *increase speed* as well as *insert* and *paste* were the exceptions where a single gesture was assigned to two tasks within the same category with no conflict. For *play* and *increase speed*, the participants intention was to use the number of spin cycles of the index finger to indicate the speed of the simulation i.e. a single clockwise spin to indicate *play*, two clockwise spin to indicate *twice* the speed and three spins for *quadruple* speed. For *insert* and *paste*, the participants felt the two tasks served a similar purpose; *insert* allowed a user to select the object from menu and placed it in the scene, whereas *paste* allowed a user to place an object from the clipboard into the scene. In the follow up interviews, participants suggested a simple resolution to this would be to provide unique selection spaces for the *insert* menu and *paste* clipboard.

With the minor ambiguities resolved, we were able to construct a consistent set of user-defined gestures which contained 44 gestures, where 34 gestures were unimanual and 10 were bimanual. The complete gesture set is illustrated in Figure 5.

The Subjective Rating on Goodness and Ease. After the participants had finished designing gestures for a task category, they were asked to subjectively rate their gestures for goodness and ease to perform on a 7-point Likert scale. By comparing these subjective ratings between the consensus set (user-defined set) and the discarded set, we found that the average score for gestures that users believed were a good match for the tasks was 6.02 ($\sigma = 1.00$) for the consensus set and 5.50 ($\sigma = 1.22$) for the discarded set, and the average score for the ease of performance was 6.17 ($\sigma = 1.03$) for the consensus set and 5.83 ($\sigma = 1.21$) for the discarded set. The consensus set was rated significantly higher than the discarded set for both goodness ($F_{1,\ 798} = 43.896$, $p<.0001$) and ease of performance ($F_{1,\ 798}=18.132$, $p<.0001$). Hence, we could conclude that, on average, gestures in the user-defined set were better than those in the discarded set in terms of goodness and ease of performance.

4.4 Findings from the Design Process

Participants were asked to think-aloud when designing their gestures, and a follow-up interview was conducted after the experiment was complete. Analysis of the resulting empirical data showed recurring thought processes. We present seven motifs which describe the mutual design patterns encountered in designing gestures for AR, which we describe as *reversible and reusable*, *size does matter*, *influence from existing UI*, *the obvious and the obscure*, *feedback backfired*, *menu for AR, axes and boxes,* and *variation of hand poses*.

Reversible and Reusable. The consensus set included reversible and reusable gestures. We defined reversible gestures as those when performed in an opposite direction yielded opposite effects e.g. *rotation*, *scaling*, *increase/decrease speed* etc. We defined reusable gestures as those which were used commonly for tasks which were different, but participants felt had common attributes e.g. increase speed/ redo, decrease speed/undo and insert/paste. In the experiment there were several dichotomous tasks that are separate tasks which perform the exact opposite operation. Participants used reversible gestures for both tasks where the opposite effect was presented in the single animation, such as *rotation* and *scaling*, as well as tasks where the opposite effects were shown separately, such as *increase/decrease speed*, *previous/next*, *undo/redo*, *group/ungroup*, and *open/close menus*. All two-handed dichotomous tasks were *symmetric bimanual* with the gestures performed on both hands being the same *form*.

Size Does Matter. We found that the virtual object's size influenced the design decision of some participants, especially with regards to the number of hands that they would use to manipulate the object for example the majority of gestures performed for *scale* tasks were bimanual. This was due to scaling involving shrinking and enlarging the target object within and beyond the palm size. Some comments are as follows:

"Instinctively, I would use two hands to adapt to the size of the model but it's cool if I can use just the two fingers (one-handed) for something as large." – P04

"Depending on the size of the piece, I can use two hands when it's big but in the case of small piece, it's enough to use the two fingers (thumb and index)." – P12

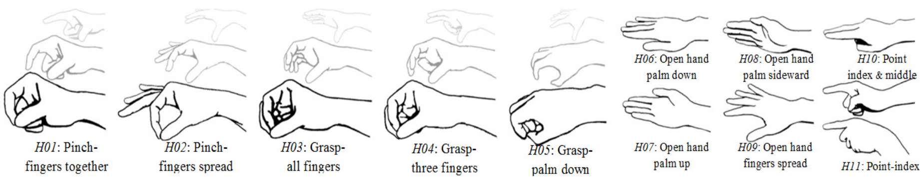


Fig. 4. Variants of hand poses observed among gestures where the codes, H01-H11, were assigned for ease of reference

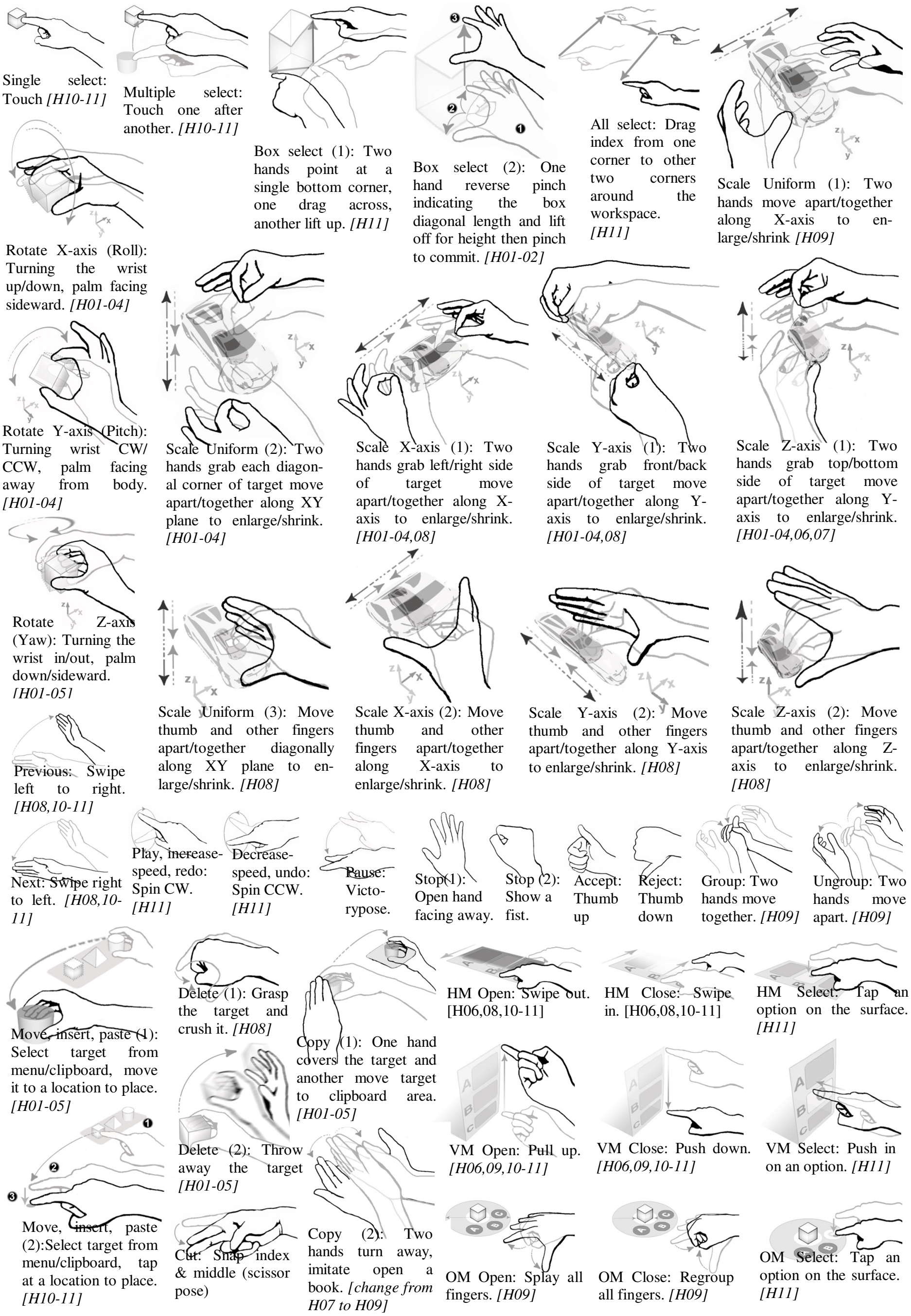


Fig. 5. The user-defined gesture set for AR. The number shown in the parenthesis indicates multiple gestures in the same task. The codes in the square bracket indicate the hand pose variants (Figure 4) that can be used for the same gesture.

Influence from Existing UI. When participants found it difficult to come up with a gesture for a particular task, they would often resort to using metaphors from familiar UI. For example when designing a gesture for the *delete* task, several participants imagined having a recycle bin that they could move the target object to. For other arbitrary tasks, users would often resort to *double-tapping*. Some examples of how participants explained these actions were:

"I would select and double-click… I'm thinking too much like Microsoft. It's just the thing that I'm used to." – P10

"The way I do it on my phone is that I would scale like this and then tap it once." – P14

The Obvious and the Obscure. Gesturing in 3D space allows for higher expressiveness, which in turn led to use of commonly understood gestures in the real-world. For example, there was a high level of agreement on the *symbolic* gestures thumbs up/down for *accept/reject* with scores of 9 and 10 respectively (out of 20). This was also the case for *metaphoric* gestures such as a scissor gesture for the *cut* task with the score of 7. User's liked the idea of using these gestures from the real world, resulting in higher than average goodness and ease scores, with averages of 6.87/6.75 (σ =.35/.71) for thumbs up, 6.5/6.5(σ =.71/.85) for thumbs down and 6.5/6.67(σ =.84/.82) for scissor pose.

The majority of participants found it challenging to come up with metaphors to design gestures for 3D tasks that they referred to as "abstract", such as *box selection*. In this task, users' had to design a gesture to define the width, depth and height of a 3D bounding box around target objects for selection. There was little agreement upon a common gesture, with a low agreement score of 0.095. In cases where the agreement score is below 0.1, we recommend further rigorous studies and usability tests to select the best gesture for the task. One participant expressed an opinion which was shared by many others:

"I don't think that it's unsuitable (the proposed gesture) but it's just very arbitrary and there is not a lot of intrinsic logic to it. If somebody told me that this is how to do it then I would figure it out but it's not obvious. It's just an arbitrary way of selecting a 3D area." - P11

Feedback Backfired. Our experimental design included the use of a 3D camera to support hand occlusion, which gave users some concept of the relative position between the virtual contents and their hands, however some participants found it to be obtrusive. We present ideas on how to improve this component of the experience in Section 5.1. One example of this criticism was as follows:

"Your hand gets in the way of the object so it can be hard to see how you're scaling it." – P11

Menus for AR. There was no significant difference in menu ranking. Some participants favored the horizontal menu because it was simple, familiar, easy to use/understand, supported on-the-surface gestures for touch sensing and did not interfere with virtual content. Others disliked the horizontal menu and felt it did not take advantage of 3D space with some options being further away and hard to reach.

The majority of participants found the vertical menu novel, and some found it to be appealing, easy to understand and that it made a good use of space with the distance to all options was evenly distributed. However, some found it harder to operate as they needed to lift their hands higher for options at the top if the buttons were arranged vertically.

Finally, some participants liked the object-centric menu because it was unique and object-specific so they knew exactly which object they were dealing with. However, some participants thought that it was unnatural and harder to operate in a crowded workspace. Furthermore, the open/close gestures for the object-centric menu were not as obvious, as indicated by the low agreement score of 0.11, as opposed to horizontal and vertical that scored 0.905.

Axes and Boxes. The *rotation* and *scaling* tasks, allowed for three possible coordinate systems, local, global, and user-centric, which corresponded to the *object-centric*, *world-dependent*, and *world-independent* categories in the *binding* dimension. In practice, we found that the transformations were mostly *object-centric*; the participant would perform gestures based on the direction of the transformation presented on the object. This was expected because people would naturally perform these tasks physically and adapted their bodies and gestures to suit the operation.

To perform a *rotation*, participants would grasp the object with at least two contact points and would move their hand or turn their wrist accordingly. For *scaling* on 3 axes, participants would grasp or use open-hands to align with the sides of object and increased or decreased the distance between them to enlarge or shrink in the same direction as the transformation. *Uniform scaling* was less obvious, for example some participants preferred using open hands moving along a single axis in front of them, as shown in Figure 5 *uniform scale (1)*. Others preferred grasping the objects' opposing diagonal corners and moving along a diagonal line across the local plane parallel to the table surface as shown in Figure 5 *uniform scale (2)*. Some user's expressed concern about how to perform the task for a round object, and suggested that bounding volumes must be provided for these models for manipulation.

Variation of Hand Poses. Variants of a single hand pose were often used across multiple participants, and sometimes even by a single participant. We clustered common hand poses into eleven poses, as shown in Figure 4. Multiple hand poses can be used interchangeably for each gesture in a given task.

5 Discussion

In this section, we discuss the implications of our results for the fields of AR, gesture interfaces, and gesture recognition.

5.1 Implications for Augmented Reality

While our experiment was conducted in a tabletop AR setting, the majority of the user-defined gestures are equally suitable to be performed in the air. Only four gestures were *on-the-surface*, *select all* and *open/close/select horizontal menu*, with three *mixed locale*, *box select (1)* and *insert/paste (2)*. This opens up our gesture set to other AR configurations, including wearable interfaces.

For our experiment, we implemented hand occlusion to give better understanding of the relative positions of the users' hands and virtual content. However we found that this could hinder user experience when virtual objects are smaller than the user's hand, occluding the object completely. We recommend that the hands should be treated as translucent rather than opaque, or occluded objects are rendered as outlines to provide some visual feedback of the objects' location.

As discussed in *axes and boxes* motif, a clear indicator of axes and bounding boxes should be provided during object manipulation tasks. Due to an absence of haptic feedback, visual feedback should be provided to inform users of the contact points between hands and objects.

5.2 Implications for Gesture Interfaces

We found most of the gestures elicited were *physical* (39%). Wobbrock et al. reached a similar outcome for surface gestures and suggested using a physics engine for handling these gestures. This approach was implemented by Hilliges et al. [7] and Benko et al. [8], who introduced "physically-based interaction", however only basic manipulations were demonstrated, with limited precision and control over the virtual contents. We believe that better control can be achieved by manipulation of the dynamical constraints imposed by the engine. Many gestures can make use of the collision detection component without the dynamics for tasks such as object selection, scaling etc.

In the *size does matter* motif we described how object size influences the number of hands used for manipulation. Since the resulting user-defined gesture set contains both one-handed and two-handed gestures for tasks such as scaling, we suggest taking an advantage of this fact to provide different levels of control. For example, in scaling tasks, by combining a snap-to feature for different granularities, unimanual scaling could offer snap-to in millimeter steps and bimanual in centimeter steps, as users tend to use one hand for an object smaller than their palm's size and two when it is larger.

As mentioned in *the obvious and the obscure* motif, care must be taken when choosing gestures for tasks with low agreement scores. We recommend follow up studies to determine usability by comparing these gestures, designer-refined gestures, menu options and even alternative modalities in case of multimodal interface.

5.3 Implications for Gesture Recognition

High degree of freedom hand pose recognition is achievable, however it is computationally expensive. In the *variation of hand poses* motif, we found a limited number of common poses (Figure 4), reducing the search space. Furthermore, the majority of the resulting gestures were *static pose and path*, which are simpler to recognize than *dynamic pose and path* gestures.

6 Conclusion and Future Work

We have presented an experiment and the results of a guessability study for natural hand gestures in AR. Using the agreement found among the elicited gestures, 44 user-defined gestures were selected as a "consensus set". Although gestures were found for all 40 tasks, agreement scores varied, suggesting that some gestures are more universally accepted than others. We are conducting a further study to validate our gestures, where a different group of participants will be shown the elicited gestures from both consensus and discarded sets and their preferences determined for each task to confirm our result.

References

1. Azuma, R.: A Survey of Augmented Reality. Presence 6, 355–385 (1997)
2. Heidemann, G., Bax, I., Bekel, H.: Multimodal interaction in an augmented reality scenario. In: ICMI 2004 - Sixth International Conference on Multimodal Interfaces, October 14-15, pp. 53–60. Association for Computing Machinery (2004)
3. Kolsch, M., Bane, R., Hollerer, T., Turk, M.: Multimodal interaction with a wearable augmented reality system. IEEE Computer Graphics and Applications 26, 62–71 (2006)
4. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., Feiner, S.: Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In: ICMI 2003: Fifth International Conference on Multimodal Interfaces, November 5-7, pp. 12–19. Association for Computing Machinery (2003)
5. Lee, T., Hollerer, T.: Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. In: 11th IEEE International Symposium on Wearable Computers, ISWC 2007, October 11-13, 2007, pp. 83–90. IEEE Computer Society (2007)
6. Fernandes, B., Fernandez, J.: Bare hand interaction in tabletop augmented reality. In: SIGGRAPH 2009: Posters, August 3-7. Association for Computing Machinery (2009)
7. Hilliges, O., Kim, D., Izadi, S., Weiss, M., Wilson, A.: HoloDesk: direct 3D interactions with a situated see-through display. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 2421–2430. ACM, Austin (2012)
8. Benko, H., Jota, R., Wilson, A.: MirageTable: freehand interaction on a projected augmented reality tabletop. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 199–208. ACM, Austin (2012)
9. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1083–1092. ACM, Boston (2009)
10. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 197–206. ACM, Vancouver (2011)
11. Wobbrock, J.O., Aung, H.H., Rothrock, B., Myers, B.A.: Maximizing the guessability of symbolic input. In: CHI 20 Extended Abstracts on Human Factors in Computing Systems, pp. 1869–1872. ACM, Portland (2005)
12. Lee, J.Y., Rhee, G.W., Seo, D.W.: Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. International Journal of Advanced Manufacturing Technology 51, 1069–1082 (2010)

13. Lee, T., Hollerer, T.: Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality. IEEE Transactions on Visualization and Computer Graphics 15, 355–368 (2009)
14. Olwal, A., Benko, H., Feiner, S.: SenseShapes: using statistical geometry for object selection in a multimodal augmented reality. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 300–301 (2003)
15. Lee, M.: Multimodal Speech-Gesture Interaction with 3D Objects in Augmented Reality Environments. In: Department of Computer Science and Software Engineering. Doctor of Philosophy, University of Canterbury, Christchurch (2010)
16. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, October 26-29, pp. 127–136. IEEE Computer Society (2011)
17. Schuler, D., Namioka, A.: Participatory Design: Principles and Practices. Lawrence Erlbaum, Hillsdale (1993)
18. van Krevelen, D.W.F., Poelman, R.: A Survey of Augmented Reality Technologies, Applications and Limitations. The International Journal of Virtual Reality 9, 1–20 (2010)
19. Broll, W., Lindt, I., Ohlenburg, J., Wittkamper, M., Yuan, C., Novotny, T., Fatah, G., Mottram, C., Strothmann, A.: ARTHUR: A Collaborative Augmented Environment for Architectural Design and Urban Planning. Journal of Virtual Reality and Broadcasting 1 (2004)
20. Lee, G.A., Billinghurst, M., Kim, G.J.: Occlusion based interaction methods for tangible augmented reality environments. In: Proceedings VRCAI 2004 - ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry, June 16-18, pp. 419–426. Association for Computing Machinery (2004)
21. Clark, A., Green, R.: Optical-Flow Perspective Invariant Registration. In: Proc. International Conference on Digital Image Computing: Techniques and Applications, pp. 117–123 (2011)
22. Piumsomboon, T., Clark, A., Umakatsu, A., Billinghurst, M.: Poster: Physically-based natural hand and tangible AR interaction for face-to-face collaboration on a tabletop. In: 2012 IEEE Symposium on 3D User Interfaces (3DUI), pp. 155–156 (2012)

Gesture-Based Interaction in Domotic Environments: State of the Art and HCI Framework Inspired by the Diversity

Ana Carla de Carvalho Correia[1], Leonardo Cunha de Miranda[1], and Heiko Hornung[2]

[1] Department of Informatics and Applied Mathematics, Federal University of Rio Grande do Norte (UFRN), Natal, RN, 59078-970, Brazil
[2] Institute of Computing, University of Campinas (UNICAMP), Campinas, SP, 13083-852, Brazil
anacarla@ppgsc.ufrn.br, leonardo@dimap.ufrn.br, heix@gmx.com

Abstract. Applications for the control and automation of residential environments (domotics) are an emerging area of study within Human-Computer Interaction (HCI). One of the related challenges is to design gestural interaction with these applications. This paper explores socio-technical aspects of gestural interaction in intelligent domotic environments. An analysis of literature in the area revealed that some HCI-related aspects are treated in a restricted manner that neglects socio-technical dimensions. We propose a framework for discussing related challenges in an integrated manner, considering the dimensions people, gestural mode of interaction, and domotics. Some of these challenges are addressed by literature outside the area of domotics. Many open research questions remain, e.g. how to design gestural vocabularies that minimize ambiguity and consider cultural and social aspects. The proposed framework might contribute to answering these questions thus to designing meaningful interaction that is intuitive and easy to learn.

Keywords: Gesture-Based Interaction, Home Automation, Smart Home, Domotics, Socio-Technical Framework.

1 Introduction

Home automation technology has emerged with the aim to facilitate activities in the household or at home, and to provide a more comfortable life for residents. Example tasks include programming the TV set, opening/closing window blinds, or controlling a home entertainment system. However, there are no user interface standards for these devices, often resulting in greater complexity of use. The same technology that simplifies life by enabling a greater number of features in a single device can also complicate our daily lives, making it more difficult to learn and use this technology. For example, changing the sound volume with some controls is done by up/down buttons, with others by left/right buttons.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 300–317, 2013.

This is the paradox of technology, and the challenge of the area of Human-Computer Interaction (HCI) is to minimize these effects via interfaces that are better suited to the growing diversity of users with access to digital technology. Intelligent domotics can offer many benefits to the residents of a house in order to decrease the complexity of using technology for this purpose, providing greater autonomy, comfort and safety. However, the field of residential applications poses some challenges to designers, since this usage context refers to an intimate setting that involves multiple people with different behaviors and different levels of tolerance regarding the effects of technology. Thus, further HCI-related studies are required in order to understand the effects of home automation and promote its benefits, considering socio-technical aspects and the diversity of people regarding culture, gender, social backgrounds, psychological states, physical capabilities, etc.

As a result of making technology available to everyone and integrating multiple device types, challenges arise related to the forms and modes how people interact with these systems. Because of these challenges and advances in hardware and software, the use of gestural interaction has been explored in literature as a viable alternative. Domotics frequently provides solutions for elderly and people with special needs. Without intending to enter the discussion whether these solutions really empower these groups of people [3], we understand that any user, regardless of capabilities or preferences is also a potential user of home automation interfaces, including interfaces with more natural interaction based on gestures.

Literature in domotics frequently studies the development of applications which control lighting, temperature or television in the home, often focusing on technological aspects of gesture recognition. There are few works in HCI dedicated to the theme of human aspects of this form of interaction, e.g. [42]. Thus, studies are required to identify limitations of gestural interfaces for the domotic context. This work presents the state of the art of gestural interaction in domotics. Additionally, we devised a framework for identifying and discussing topics and challenges for research and development of solutions, considering socio-technical aspects.

The paper is organized as follow: Section 2 contextualizes domotics and gestural interaction; Section 3 presents the state of the art of gestural interfaces for residential environments; Section 4 presents a triadic framework of multimodal interaction and identifies challenges of gestural interaction in domotics; Section 5 discusses these challenges on the optic of HCI; Section 6 concludes the paper.

2 Domotics and Gestural Interaction

The words “home” and “house” are often used interchangeably in literature. In 1985, Dovey [10] discussed the differences of these two concepts. His conception of the subject resonates well with HCI-related perspectives on domotic environments, such as those of Saizmaa and Kim [42]. According to Dovey [10], the “house” is an object, and “home” constitutes an emotional and significant relationship between people and their houses, i.e. the house is the local where the experiences of the home take place. For a more accurate conception of the phenomenon living”, Dovey proposes to examine the house with respect to the concepts “order”, “identity” and “compliance”.

The terms home automation, domotics, home computing, smart home and intelligent domotics have been employed in scientific papers in the area. However, there is still no consensus in literature regarding the use of these terms. Saizman and Kim [42] presented five scenarios of smart homes and analyzed common ideas between them. The conclusion was that applications use automated and intelligent computing in the context of the home. However, we believe that this description defines the "domotics" incompletely, since only the isolated meanings of each word are considered, and since it does not distinguish between what is, in fact, automated and what is intelligent.

For Aldrich [1, pp. 17], smart homes can be defined as a residence equipped with computing and information technology which anticipates and responds to the needs of the occupants, working to promote their comfort, security and entertainment through the management of technology within the home and connections to the world beyond. In this view, "smart homes" infer the needs of residents and (semi-)automatically execute them. On the other hand, "automated homes" require explicit commands from users to perform some action. In order to not to restrict ourselves to only either smart or automated homes, we chose the term "intelligent domotics" in our research, since its definition cover both cases. Our own preliminary definition of "intelligent domotics", which is based on the current state of the art, is: intelligent domotics comprises the use of automated and smart applications in the home, aiming at improving aspects such as security, comfort, and health of residents.

For Cook and Das [9], smart environments are able to acquire and apply "knowledge" about the environment and its inhabitants in order to improve the inhabitants' experience with the environment. Sadri [41] includes to the description of this environment the concepts "interconnection", "adaptation", "dynamism", "intelligence", and "integration". In this view, the traditional means of a system's input and output disappear. For environments controlled by technology, Sadri stipulates that the way of interaction should be the most intuitive and closest to the daily lives of residents. The use of gestural interaction, for being frequently used in everyday social life, is considered intuitive in human communication. Thus, joining gestures and intelligent domotics seems a topic that should be further explored.

In HCI, gestural interfaces can be studied within the context of Natural User Interfaces (NUIs). Studies in the area of NUIs are concerned with questions such as how the five senses of the human being can serve as a form of interaction with electronic devices. The basic idea is to approximate user experience to everyday contexts and dialogues without the need for complex learning. However, to Norman [33] NUIs are not "natural". Norman states that the gestural vocabulary of applications with interfaces based on gestures is artificial from the time of its definition. Developers typically define gestures arbitrarily, and over time these settings can become "natural", i.e. customary, for a group of people, but probably not for a user population with a great cultural diversity. To illustrate that many gestures are not intuitive or natural, Norman and Nielsen [32] cite the example of the zoom multi-touch interface, claiming that the "pinch to zoom" gesture is not natural: when reading a book and "zooming in", we bow our heads closer to the book without thinking. We concur with Norman's critique. When addressing the "naturalness" of interaction, i.e. an interaction that is "intuitive" and "easy to learn", it is indispensable to consider social and cultural aspects of a target audience when defining a gestural vocabulary. Since the adjective "natural" is now widespread in HCI literature of the area, we continue to use it in this text, however in quotes and as a synonym to "meaningful", i.e. when we write "natural gestures" we use it in the sense of "meaningful gestures that are intuitive and easy to learn".

3 State of the Art of Gestural Interfaces for Intelligent Domotics

In this section we present the state of the art of gestural interfaces and synthesize the main findings. In concordance with the taxonomy in [19] we grouped the analyzed literature into two distinct forms of gestural interaction, perceptual and non-perceptual, whereas perceptual input allows the recognition of gestures without the need for any physical contact with an input device or any physical objects.

3.1 Perceptual Technologies

While the Kinect has been a success in the games area, Panger [36] investigated the possibilities outside the living room. The author studied the problem of people who want to flip through a recipe book or select a tune to listen while cooking, even with sticky or oily fingers, or hands occupied with kitchen utensils. He proposed an application based on Kinect's depth camera that captures the user's joints for the recognition of movements that consist only in left, right, front and back. Another application that uses the Kinect is the Ambient Wall [21], a smart home system that can display the current status of the house through a projection on a wall, allowing the user to control the TV, check the room temperature, etc. Hands-Up [34] uses the Kinect device with a projector to project images onto the ceiling of the room. This projection location was chosen for usually being the least-used surface inside a house. Additionally, when people get tired of their jobs they often lie down on the bed or sofa and stare at the ceiling of the room. The Hands-Up application interface consists of a circular main menu, in which users can control various devices in the home.

The Kinect has also been used for applications that provide security to the user. Rahman et al. [38] mention a number of functions contained on a car dashboard that are controlled by touch interfaces, which increases the risk and distraction of drivers on the roads. To alleviate this problem, the authors developed and evaluated a purely gestural interface to control secondary functions of a car, that does not use a graphical interface, but audible and haptic feedback. Although this solution has been developed to control the sound system of a car, we deem it relevant for your presentation because it could be adjusted to the residential environment.

The need to always have a remote control on hand to interact with the devices was the main motivation of Solanki and Desai [43] to develop Handmote, an application which recognizes gestural movements to interact with various devices that use a remote control. Their Arduino-based solution recognizes images of the user's hand and processes them in real-time, sending infrared signals to the respective appliance. Example gestures that are converted to infrared signals for a TV set include signalizing a cross for muting the TV, or turning the hand clockwise or counter-clockwise to change the volume level or TV channel.

Irie et al. [15] discussed a three-dimensional measurement of smart classrooms using a distributed camera system to improve the recognition of three-dimensional movements of the hands and fingers. Their solution allows controlling appliances, TV sets and room lighting through gestures, even when multiple users are present. The Light Widgets [11] application was developed to enable an interaction that is "transparent" and low-cost and that might be accessed on different surfaces, e.g. on walls, floors or tables. Based on the configuration of a surface as the locus of interaction, Light Widgets "reacts" when a user approximates his or her hand to the surface. Users are identified by their skin color. Yamamoto et al. [46] stated that various methods of gesture recognition using the recognition of the user skin color have limitations being sensitive to changes in illumination and certain colors of clothes. Furthermore, using single fixed cameras in narrow spaces, the gesture recognition is restricted to only one person. Thus, the authors use multiple cameras in the corners of the ceiling pointing downwards, to view the entire bodies of users and their faces. A distinguishing feature of this system is its ability to simultaneously process body movements, gestures and face recognition.

For Kim and Kim [20] a major concern is the recognition of gestures as a segment of a few significant gestures from a continuous sequence of movements, i.e., the question of how to detect the start and end points of an intentional gesture. This is a complex process because the gestures have two properties: ambiguity of recognition – due to the difficulty to determine when a gesture starts and ends in a continuous sequence of movements, and segmentation – since multiple instances of the same gesture vary in shape, length and trajectory, even for the same person. To solve these problems Kim and Kim [20] proposed a sequential identification scheme that performs gesture segmentation and recognition simultaneously.

Henze et al. [13] analyzed static and dynamic gestures as forms of interaction with a music application. Static gestures refer to the user's pose or spatial configuration, and dynamic gestures to his or her movement in a certain time interval. These authors performed a 3-step evaluation with twelve users with different profiles, five male and seven female. The results indicated that dynamic gestures are easier to remember, more intuitive and simpler for controlling a music application. Kleindienst et al. [23] discuss the HomeTalk platform that assists users in some domestic services via multimodal interaction. The core of the platform is a residential gateway that acts as a center of family communication. Through direct interaction with a home appliance it is possible to automate different services, and monitor their progress on a PDA. This application provides a greater level of security to residents by providing information about different locations in the home as well as by controlling the food cooking temperature and time, thereby avoiding possible fires.

As an attempt to design more intuitive interfaces in domotic environments, the system developed by Hosoya et al. [14] uses a technique of real-time self-imaging on a translucent in order to improve feedback to the performer of the gestural interaction. The system developed by the authors visualizes the objects in a local or remote room on a screen and superimposes a translucent image of the user. That way, a user can "touch" an object without making real contact – the user's translucent mirror image touches the object on the screen –, and manipulate or interact with objects such as the TV set, or a lamp. Objects in remote rooms need to be tagged with infrared tags.

3.2 Non-perceptual Technologies

After having reviewed "perceptual technologies" for gestural interaction, we now present solutions that use non-perceptual technologies, i.e. solutions that enable gestural interaction via gloves, rings, wands or other physical artifacts.

Bonino et al. [5] mentioned that many domotic applications support interaction with devices by fixed touch panels, or by applications on desktop computers, tablets or smartphones. However the use of these technologies has limitations with respect to user interaction, e.g. regarding multi-purpose devices. Furthermore, in the case of mobile devices, there are situations where their use is not possible, e.g. during a shower. In order to circumvent these limitations, the authors chose to use a wristwatch-based solution they call dWatch [5]. Additionally to five other watch functions (time, alarm, temperature, motion, and list of favorite functions), gestural interaction is specifically responsible for controlling household appliances.

Rahman et al. [37] used a residential application to test the trajectory recognition of a user's gestural movements. Their glove-based system enables residential users to interact with the environment. This application consists of infrared cameras, infrared LEDs, and gloves, with the rationale to increase accuracy and to enable gesture recognition in the dark and at relatively low costs. To initiate interaction, the user presses a switch contained in the glove and then "draws" in the air. The system was built to control the lighting of the house, movies and music through movements that resemble some characteristic of the target object's interaction. For example, to control sound, the user "draws" the letter "S", to start media playback the sign ">".

Jing et al. [16] proposed a new physical interaction device called the Magic Ring (MR) which is intended to serve as a means of interaction with different electronics in a residential environment. A comparative evaluating of the use of a traditional remote control and the MR was performed. The results suggested that the use of MR has a smaller learning curve and provides the user with less fatigue than a traditional remote control. For Miranda et al. [28] the remote control in its current form is unsuitable for applications of interactive Digital Television (iDTV). For this reason, the authors proposed the use of Adjustable Interactive Rings (AIRs) to better interact with these applications. With a focus on diverse user capabilities and different contexts of use, the solution consists of three AIRs with distinct functionalities. According to its functionality, each ring has a different color, a single button and a Braille label.

XWand [45] is a multimodal application that enables input of speech and gestures to control various devices in the home. XWand is shaped like a wand that, when pointed at a device, can control it through speech or wand movement. To turn on a lamp, for example, the user has to point the wand at the lamp and say "connect". The system emits an auditory feedback when recognizing a target object. However, the interaction using the Magic Wand itself does not provide feedback to the user.

Carrino et al. [7] also described a multimodal approach based on deictic gestures, symbolic gestures and isolated words with the Wiimote control. The conceptual elements used in this study for three types of entries are: camera, accelerometer and microphone. The camera is attached to the arm or the hand of the user and dedicated to the recognition of deictic gestures using the method Parallel Tracking and Multiple Mapping (PTAMM). The accelerometer is used for recognition of symbolic gestures. The application provides auditory or haptic feedback, or gives feedback through the environment itself, e.g. the feedback of the successful execution of the command "turn on the lights" are turned on lights. The authors conducted a questionnaire-based usability evaluation of a prototype with ten participants using a Likert scale with questions of effectiveness, efficiency, experience and satisfaction. Another application that uses the Wiimote was developed by Neßelrath et al. [29] for gestural interaction with three appliances: kitchen hood, room lighting and TV. The main concern of the authors was to find a small set of commands for the application. They found that one way to decrease the set of commands is by gesture control in context, i.e., one gesture can activate functions of various applications. Another study that also is concerned with the gestural vocabulary is that of Kühnel et al. [24]. The application employs a smartphone and accelerometers as a means of gesture recognition, and can be used for controlling the TV, lamps, window blinds, as well as for interacting with an Electronic Program Guide (EPG).

Jung et al. [17] focused on the support of daily tasks of the elderly and people with special needs, especially those that require the use of wheelchairs. In order to increase comfort in posture and mobility, the authors developed an "intelligent bed", an "intelligent wheelchair" and a robot for transferring a user between bed and wheelchair. The authors developed interfaces based on hand gestures and voice to control equipment and a health monitoring system that was used not only to assess the state of health, but also as a means to improve comfort by controlling the environment temperature and the transfer of the user between bed and wheelchair.

The work related to health is usually related to monitoring and health care of the residents of the house. The motivation for the creation of the "Gesture Pendant" [12] was the need to reduce the number and complexity of the vocabulary gestures used to interact with the appliances in the home. The Gesture Pendant has a camera surrounded by LEDs, which recognizes gestures also in dark environments. A user can interact through pre-defined control gestures, user-defined gestures or voice commands. Another feature of this application is the monitoring of the user for diagnosis, therapy or emergency services, such as reminding the user to take medication or notify family members.

3.3 A Synthesis of the Main Findings of the Literature Review

Regarding HCI-related topics, the following were addressed in the works presented in the previous two subsections: accessibility, usability, personalization, privacy, ambiguity of gestures, gestural anthropology (i.e. the relation between gestures and culture), and gender-related issues.

Although many solutions address usability in some form, it can be noted that the "perceptual" solutions seem to put less focus on the other topics. This might be explained by the focus the "perceptual" solutions put on the quality of gesture recognition and related issues, i.e. before addressing accessibility, privacy or personalization, the underlying system of gesture recognition has to work at a satisfying level. On the other hand, when addressing the topics listed in Table 1, in the case of "non-perceptual" solutions this is not necessarily done considering gestural interaction per se but other components of the system. When addressing accessibility, this is often related to components of the system that provide visual feedback or related to specific solutions for people with specific special needs.

Accessibility considering gestural interaction would mean to investigate how e.g. people with motor impairments would be able to interact with the system. Usability is arguably the most completely considered area, however, in the case of non-perceptual solutions, usability tests usually focus on graphical interfaces or different modes of feedback. Few works address personalization of gestural interaction, which is a complex problem, e.g. due to the complexity of guaranteeing fast and accurate recognition of personalized gestures. Privacy is often concerned with camera positions or data storage of user images, and not with issues regarding the privacy of performing gestures. Ambiguity in the reviewed solutions is only discussed regarding the relation between already habitually used gestures and gestures used for interacting with the system, however without regarding other contextual factors, such as characteristics of the target audience. Only one solution treated issues related to anthropology, e.g. questions such as cultural aspects that influence the understanding or appropriateness of gestures within a certain context. Regarding gender-related issues, none of the solutions of Table 1 reported that these were a concern during the earlier stages of development, and only three studies reported at least the number of male or female participants [12,17,24], however without stating whether gender had any influence on testing or subsequent design cycles.

Table 1. HCI-related topics addressed by the works

	Perceptual												Non-Perceptual											
	[36]	Ambient Wall [21]	Hands-Up [34]	[38]	Handmote [43]	[15]	Light Widgets [11]	[46]	[20]	[13]	HomeTalk [23]	[14]	dWatch [5]	[37]	[16]	AIRs [28]	XWand [45]	Magic Wand [35]	[7]	[29]	[24]	SmartKom [44]	[17]	gesture Pendant[12]
Accessibility			X													X	X		X	X		X	X	X
Usability	X	X	X	X		X		X		X	X	X	X	X	X	X	X		X	X	X	X		X
Personalization																				X				
Privacy													X		X		X	X	X				X	X
Ambiguity					X				X				X	X				X	X		X			
Anthropology																					X			
Gender															X						X			X

We believe that some of these limitations are a result of considering the topics of Table 1 in an isolated manner regarding technology, gestural interaction, and individual/social aspects. For current solutions in the area of domotics this might be adequate to some extent, since the problems they address are of a relatively low complexity, e.g. functionalities in the areas of comfort or security such as turning on the lights. However, regarding the area of health, or more complex problems in the areas of comfort and security, or even new areas that are not yet addressed in domotics, we think a more comprehensive and particularly a socio-technical perspective is required, grounded on methods and frameworks of contemporary HCI. As a first step towards this direction, we thus propose in the next section a framework for conceptualizing research and design questions in an integrated manner.

4 Socio-technical Aspects of Gestural Interaction: Framework and Challenges

After analyzing the state of the art, we can observe that researchers and developers are more concerned with trying to offer “comfort” to users through some complex computational solutions that can control e.g. the lighting, the room temperature, TVs or home entertainment systems. However, besides of technical aspects of controlling appliances in the home, also aspects from the social sphere need to be considered.

In order to identify and discuss these HCI-related aspects and challenges of multimodal interaction, we propose a socio-technical framework of multimodal interaction in the context of intelligent domotics. The framework consists of the main dimensions technology, modes of interaction, and people. The concentric organization of these three dimensions symbolizes their interdependency in a triadic relationship. The main idea of this framework is to identify and discuss the challenges of different forms of interaction with technology considering socio-technical aspects in an integrated manner, i.e. we acknowledge that a challenge should not be considered isolatedly under a single perspective, but under a perspective that combines multimodal, technological, and social aspects and considers their interdependencies. It should be noted that in the context of this paper, we only consider the mode of “gestural interaction”, and only technology in the context of domotics (Fig. 1). Possible users of the framework include researchers and developers, who can catalog socio-technical research challenges and discuss challenges that permeate the implementation of their solutions.

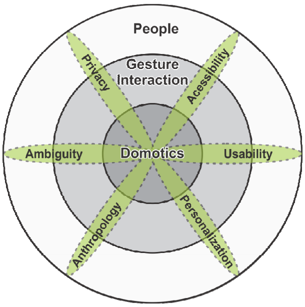


Fig. 1. Framework of socio-technical aspects of gestural interaction

The aspects discussed in this section do not comprise an exhaustive list, but are the ones that emerged from the literature review presented in the previous section, namely accessibility, usability, personalization, privacy, ambiguity, and anthropology. Each aspect is represented by a dashed ellipse in Fig. 1 and has intersections with the three dimensions domotics, gestural interaction, and people.

A noticeable characteristic of the state of the art presented in the previous section is the high share of concepts or techniques based mainly on solutions that use graphical interfaces or non-perceptual technologies. These solutions, e.g. guidelines or methods, were developed and used for traditional GUI or Web applications and often do not meet requirements for interaction with domotics, i.e. an interaction that should be more “natural” in the sense of being meaningful and intuitive. In the following, we will discuss the challenges regarding the six aspects of gestural interaction in the domain of intelligent domotics presented in Fig. 1. We will analyze which challenges can be addressed by adapting work from related areas (e.g. gestural interaction in other contexts, or general literature in HCI), and which research issues remain.

Accessibility – Accessibility has generally been perceived as a necessary attribute of quality of software and hardware systems. Consequently, we believe it is essential to provide accessibility in residential solutions. Some studies proposed multimodal interfaces for gesture and voice, aiming at the inclusion of a greater diversity of users, e.g. [12,23,38,44,45]. However, none of these studies was concerned with analyzing e.g. the accessibility of the gestural vocabulary with respect to people with mental or physical impairments.

Computing solutions from various contexts have visual interfaces, whether in tablets, phones, computers or through projections. Often, researchers focus on accessibility in these conditions. However, for the context of intelligent domotics, considering technologies that allow gestural interaction and that have no visual interfaces, accessibility is poorly explored. Although there are already consolidated accessibility guidelines aimed at a content that is perceivable, operable, and understandable by a wide range of users, as well as compatible with a wide variety of assistive technologies, not all principles of accessibility are “compatible” with the context of home automation applications. Examples include some guidelines for Web applications or guidelines such as “making all features also available to keyboard use” or “provide alternative text to non-textual content”. Changing the way people interact with the environment brings out new aspects of accessibility that require more research. There are several questions about how we can develop gestural applications that are more accessible to the diversity of the audience, and the challenge becomes even more complex by the lack of development methodologies for gestural applications, as well as evaluation methods for this particular type of interaction.

Kane et al. [18] reported a study on accessibility in gestural interfaces applied to touchscreen surfaces. They found that, given a gestural vocabulary, blind participants chose significantly different gestures than sighted participants. Blind participants showed strong preferences for gestures that were in the corners and edges of the screen, as well as for multitouch gestures. Kane et al. also discovered differences regarding the performance of gesturing, e.g. gestures produced by the blind participants were bigger, slower, and showed a greater variation in size than those produced by sighted participants. An important result of the study is that according to some blind participants, they did not know how to perform some of the gestures used in the defined gestural vocabulary, including letters, numbers and other symbols. However, there has been little research about the differences or peculiarities of gestures regarding so perceptual aspects and the residential context.

Another aspect not yet studied in literature is the accessibility depending on the gender of the users, since physical and psychological conditions differ for each gender. The structure of our framework considers the mutual dependencies of the mentioned challenges: accessibility may be interplaying with usability, personalization of gestures, privacy and the gestural ambiguity of the solution.

Usability – Most of the works presented in the previous section that discuss usability are concerned with efficiency and learnability, which are only two of Nielsen's [30] five main topics: learnability, efficiency, memorability, errors and satisfaction. One of the benefits of graphical interfaces is to aid the memorization of commands for interaction, because the information is organized graphically in windows and is represented by text, icons or other visual elements. The language of commands based on a menu structure has the cognitive advantage that commands can be recognized instead of being required to recalled. For solutions that use GUIs Lenman et al. [25] proposes the use of "marking menus". Learning the command set is performed gradually through "pie-menus" that indicate the direction of the movement that the user needs to perform. Another progressive form of learning gestures is through multimodal voice interaction, where the application supports the user by describing the movements required to perform an action. As the user learns to perform gestures, the application no longer provides the respective instructions.

For gestural interfaces, gestural ergonomics is also important, since the interface should not be physically stressing [31]. The comfort when interacting with the system is important, which is not achieved when a user has to "wear" technology, like a glove. With respect to the use of perceptual technology, discomfort and fatigue might arise when the user's main means of interaction are arms and hands interacting without a supporting surface. Nielsen et al. [31] describes ergonomic principles for building a good interface with gestures, e.g. relax muscles, avoid repetition or staying in a static position, avoid internal and external force on the joints.

Although usability guidelines are widely used in traditional solutions, new principles or the adaptation of existing ones are required for gestural interfaces [33], especially those without a GUI. Fundamental principles of interaction design that are independent of technology [33] might provide starting points for this investigation, e.g. visibility (related to affordances), feedback, consistency, non-destructive operations, discovery, scalability, and reliability.

Personalization – Many problems involving customization are related to the huge amount of information that needs to be managed simultaneously. To support versatility of gestural commands for different types of solutions it is essential that applications are customizable, as well as easy and fast to train. Achieving these properties leads to the problem of forming and recognizing gestures freely. In order for the application to learn new commands, users have to train it repeating the same command several times. This repetition might generate discomfort for the user. Liu et al. [27] aim to decrease the number of required repetitions by using discrete Hidden Markov Models (HMMs). However, the authors point out that there are several technical challenges for interaction based on gestures. Unlike many pattern recognition problems such as speech or writing recognition, gesture recognition lacks a commonly accepted standard or "vocabulary". Therefore, it is sometimes desirable and often necessary for users to create their own gestures. With customized gestures, it is difficult to gather a large set of training samples that is required to establish statistical methods. Furthermore, the gesture-based spontaneous interaction requires immediate engagement, i.e. the overhead of creating recognition instruments should be minimal. More importantly, the application platforms for specific custom gesture recognition are generally very limited in terms of cost and system characteristics, such as battery capacity or the buttons presented in [27].

A clear difference between customization of perceptual solutions of gestural interaction and "traditional" means of interaction is the difficulty of recognizing who interacts with the system, e.g. in a home, all residents are possible users of the application. Recognizing the user is indispensable for enabling customization. Thus challenges include how to recognize the interacting person, and how to achieve this unobtrusively, i.e. without e.g. requiring the person to utter his or her name or to look at a particular location for facial recognition. With regard to the integrated consideration of personalization within our framework, it is worth noting, that the technical feature "personalization", applied to the context of gestural interaction with domotics also has a strong social component, i.e. personalization should be consistent with the ideals of a natural, meaningful of interaction that respects the privacy of the users within the home. Furthermore, personalization may have a positive effect on usability and accessibility.

Ambiguity – In order to actually enable a "natural" interaction, the ambiguity of gestures, which is very present in the real world, needs to be reduced for interacting with the virtual world [33]. Gestures need to be cohesive and consistent in their meanings. For instance, if used as a command for interacting with a system, the movements used to express a farewell in the real world should to be used for the same purpose and with the same meaning, i.e. as a gesture of "farewell". Following this principle, the gestural vocabulary would become more intuitive, easing the learning curve of users. Some solutions acknowledge this principle [13,29], but do not take into account whether the gestures are ambiguous regarding the target population, since ambiguity is intrinsically related to the cultural aspects of the population. As an example of the problem of not considering the interdependency of gestural ambiguity with cultural aspects, the application in [39] uses only deictic gestures (pointing gestures). Gestures for some commands for this application resemble a firearm, which is probably not desired in a home or in war- or conflict-ridden regions of the world.

During social interactions, people use a large vocabulary of gestures to communicate in daily life. The gestures used vary according to contextual and cultural aspects [22] and are closely linked to other aspects of communication. A challenge that arises is that gestures for interaction with domotics must be sufficiently "natural", i.e. resemble to some degree the gestures used in everyday life. At the same time, these gestures must be recognized as intentional commands to the system, i.e. they must be distinguished from gestures of inter-person communication. This problem has also been called "immersion syndrome" [2], i.e. in a scenario in which all gestures are captured and can be interpreted by the system, gestures may or may not be intended for interaction with the system, and people can no longer interact simultaneously with the system and other people using gestures.

To clarify the idea of gestural ambiguity, we can draw an analogy to sign language. When the gestures of a domotics solution are established without a previous study, gestural ambiguity might occur on two levels, i.e. different gestures/signs might be required for asking a person or commanding the system to turn on the lights or draw the curtain, or the gesture for the command “draw the curtain” might have a different meaning in sign language. There are no studies on “gestural affordances”, i.e. the problem of gesture discoverability. Another question related to gestural affordances is if there exist any universal gestures. A positive answer to this question might reduce ambiguity and cultural dependencies. From these last considerations, it also becomes clear, that ambiguity has a strong relation to anthropology.

Anthropology – Symbolic gestures, the meanings of which are unique within the same culture, are an example of a classification used to discriminate gestures depending on the anthropology of a certain population. An example is the “thumbs-up” gesture which signifies approval in Brazilian culture but can be an insult in some middle-eastern countries. Sign languages also fall into this category and vary significantly between countries.

Researchers are still trying to understand how the gestures are influenced by culture [26]. Due to this still largely unexplored area, applications often make use of deictic gestures [6,45], i.e. pointing gestures which have much less cultural dependency, but which are also limited since not every function in a domotic environment can be executed by “pointing at things”. Furthermore, regarding the “naturalness” or “intuitiveness” of interaction in domotic environments, the use of only deictic gestures also imposes a limitation. Kühnel et al. [24] described anthropology as a requirement for computational solutions for domotic environments, not only with respect to gestural interaction, but in a broad context. One of the authors’ concerns was the writing direction of the user. Although this detail might seem irrelevant for the definition of gestures, it might very well influence whether a certain gesture is considered appropriate in a certain cultural context.

Privacy – With the rapid advancement of technology, sensors and information storage devices are becoming increasingly integrated in the solutions. These devices provide various benefits such as accuracy of command recognition, and mobility in use. However, the context of the home requires a number of concerns about the privacy of those who utilize these technologies, because it refers to an intimate environment in which the lack of privacy can have negative effects on the social relationships among residents and result in failure of the domotic application.

Considering the difficulties in the development of applications that address the requirements of the involved stakeholders in a home environment, Choe et al. [8] aim to investigate the types of activities of which residents do not want to have stored records. To obtain these results, the authors analyzed the questionnaires of 475 respondents, with 71.6% female and 28.4% male participants. A total of 1533 different activities that respondents did not want to be stored was identified. The male respondents most frequently reported activities related to the category of intimacy and the use of media. Female respondents reported activities related to the category of self-appearance and oral expressions. Moreover, Choe et al. identified places that need more care regarding residents’ privacy, e.g. bedrooms and bathrooms.

5 Discussion

Given the diversity of the population, designing purely gestural interfaces for residential environments might not be the most appropriate approach, because the sheer amount of gestures that would have to be memorized and performed would be exhaustive for the population as a whole, and especially for the elderly and people with special needs. As presented in Section 3, many applications choose to use multimodal interaction. Among the works mentioned, many authors use a combination of speech with gestures, in order to provide greater accessibility of the system, to facilitate recognition of user commands, or to decrease the complexity of the vocabulary of gestural applications.

However, as stated earlier, home automation has consequences far beyond the way we interact with the appliances in the home. Besides the technical, several other aspects about building those applications have to be addressed and analyzed under an integrated socio-technical perspective. These aspects include, but might not be limited to, accessibility, usability, ambiguity, privacy, anthropology and gender of users, since this context is closely related to physical, social, psychological, emotional and even spiritual concerns of each resident. Addressing these issues also increases the “naturalness” or “intuitiveness” of smart home automation applications.

Currently, the development of intelligent home automation applications often focuses exclusively on technological aspects, not taking into account what is actually necessary and desirable for users. Although affective, psychosocial and other aspects that cause an impact on residents are being explored more actively in the area of HCI, many open questions need to be studied, especially in the area of domotics.

Bardzell et al. [4] intended to explore issues of feminist thought intertwined with human-computer interaction. Both feminism and HCI have made important contributions to social science in recent decades. However, despite of the potential, there has been no engagement between the two areas until recently. A series of surveys, focused mainly in perceptual and cognitive tasks, revealed gender differences that can have implications for interactive systems design. However, believes still seem to be prevalent that gender does not have much influence on technology usage. Thus, one of the research objectives of Rode [40] is to show the importance of treating gender in HCI and to emphasize that it permeates all aspects of daily life, including domestic life. Many studies ignore important social aspects, in which the issues of gender occur daily. This point of view on users’ gender is relevant for home automation applications, as in a home environment all people need to interact with the applications. Thus, applications must meet the needs of both genders and should be designed and developed for this.

Aiming to address more adequately the challenges presented above, we found it necessary to design a framework that specifically addresses the challenges contained in gestural interaction with domotics. Although used exclusively in the context of gestural interaction and domotics, the framework presented in this paper could be used for a similar discussion in other application contexts and regarding other or additional modes of interaction. Saizmaa and Kim [42] presented a framework that addresses some conceptual aspects similar to the framework presented in this paper. Saizmaa and Kim [42] identified important issues that are considered or omitted in the development of intelligent homes, and organized these issues into three dimensions, human, home and technology, highlighting the need to not only to see a house as a physical thing, with walls and ceiling, but also as a "home".

The holistic approach advocated by Saizmaa and Kim [42] elucidates the complexity of smart home automation, as well as draws attention to not limiting an analysis in this domain to technological aspects alone. Although similarities such as the problem domain of domotics exist between our framework and the framework of Saizmaa and Kim, there also exist significant differences. While Saizmaa and Kim discuss different aspects of interaction, they do not elaborate on the peculiarities of different modes of interaction. In this paper, we discussed gestural interaction, but in principle the discussion could be extended to other types or combinations of modes. Saizman and Kim discuss issues in the dyads home-technology, technology-human, and human-home, and assign research topics to each dyad. Maintainability, e.g., is assigned to the dyad home-technology, although we think that also personal, cultural, or social aspects might be relevant for this topic. Hence, our framework uses a triadic structure which enables to discuss topics in all three dimensions of people, domotics, and gestural interaction. Depending on the role of the framework user, it is thus also possible to investigate one or more topics in only one or two dimensions. For instance a sociologist in the research or development team might be interested in the "people dimension" of different aspects, while a developer might be interested in technical aspects of accessibility.

The contribution of this framework is to analyze an explicit context from its initial form of interaction until its possible social implications. When we addressed these challenges from the perspective of the framework, we did not take the traditional way of analyzing only parameters of a graphical interface. This framework enables the discussion on a triadic unfolding of the dimensions that make up the framework. Although only discussed for gestural interaction, it can be applied to other modes of interaction as well, for example, gestural, mobile, brain-machine sound, among other perceptual or non-perceptual interaction. Consequently the analyses of these challenges can provide insights to the particularities of each specific mode of interaction.

We understand that regardless of the utilized technology, one of the main questions is the definition of the gestural vocabulary of the application. This definition should not be driven by technology, but by a human-centered perspective that considers the "naturalness" and "intuitiveness" of gestural interaction. In order to formalize gestural interaction we consider it essential to create a grammar which is initially free of technological aspects and which considers both multimodal commands as well as the points already discussed in this paper.

6 Conclusion

This paper presented the state of the art of residential applications that use gestural interaction to communicate with various types of appliances in the home. An analysis of scientific literature in the area of gestural interaction with domotics revealed that aspects such as accessibility, usability, personalization, privacy, ambiguity, or anthropology of gestural interaction are often not considered, or only considered in a restricted, isolated way. We argued that aspects of gestural interaction with domotics need to be discussed in a more integrated, socio-technical way, and proposed a framework that permitted us to discuss these aspects within the triadic relationship consisting of the dimensions people, domotics, and gestural interaction. We identified open research questions and challenges. Literature outside the domain of domotics provides some pointers to these questions, e.g. literature about HCI-related aspects of gestural interaction in general, or literature from HCI and related areas. However, we conclude that there are no formalized design and evaluation methods in literature about gestural interaction with perceptual technologies, i.e. applications without GUIs and without physical artifacts of interaction. Furthermore, literature in the domain of gestural interaction with domotics that discusses aspects such as anthropology, accessibility, usability, gender, personalization and privacy is scarce.

Based on the results of this work, future work involves the human-centered conceptualization and formalization of multimodal interaction for this complex context of use.

Acknowledgments. This work was partially supported by Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), by the Brazilian National Council of Scientific and Technological Development (CNPq grant #141058/2010-2), and by the Physical Artifacts of Interaction Research Group (PAIRG) at Federal University of Rio Grande do Norte (UFRN), Brazil.

References

1. Aldrich, F.K.: Smart Homes: Past, Present and Future, pp. 17–39. Springer (2003)
2. Audel, T., Beaudouin-Lafon, M.: Charade: Remote Control of Objects Using Free-Hand Gestures. Communications of the ACM 36(7), 28–35 (1993)
3. Bannon, L.: Reimagining HCI: Toward a more Human-Centered Perspective. Interactions 18(4), 50–57 (2011)
4. Bardzell, S., Churchill, E., Bardzell, J., Forlizzi, J., Grinter, R., Tatar, D.: Feminism and Interaction Design. In: CHI 2011, pp. 1–4. ACM (2011)
5. Bonino, D., Corno, F., Russis, L.: dWatch: A Personal Wrist Watch for Smart Environments. Procedia Computer Science 10, 300–307 (2012)
6. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the Usability of Gesture Interfaces in Virtual Reality Environments. In: CLIHC 2005, pp. 100–108 (2005)
7. Carrino, S., Péclat, A., Mugellini, E., Khaled, O.A., Ingold, R.: Humans and Smart Environments: A Novel Multimodal Interaction Approach. In: ICMI 2011, Alicante, Spain, pp. 105–112 (2011)
8. Choe, E.K., Consolvo, S., Jung, J., Harrison, B., Kientz, J.A.: Living in a Glass House: A Survey of Private Moments in the Home. In: UbiComp 2011, pp. 41–44 (2011)
9. Cook, D.J., Das, S.K.: How Smart are Our Environments? An Updated Look at the State of the Art. Pervasive Mob. Comput. 3(2), 53–73 (2007)
10. Dovey, K.: Home and Homeless. In: Home Environments, pp. 36–61. Plenum Press (1985)
11. Fails, J.A., Olsen, D.: Light Widgets: Interacting in Every-Day Spaces. In: IUI 2002, pp. 63–69. ACM (2002)

12. Gandy, M., Starner, T., Auxier, J., Ashbrook, D.: The Gesture Pendant: A Self-Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring. In: ISWC 2000, pp. 1–8. IEEE (2000)
13. Henze, N., Löcken, A., Boll, S., Hesselmann, T., Pielot, M.: Free-Hand Gestures for Music Playback: Deriving Gestures with a User-Centred Process. In: MUM 2010, pp. 1–10. ACM (2010)
14. Hosoya, E., Kitabata, M., Sato, H., Harada, I., Nojima, H., Morisawa, F., Mutoh, S., Onozawa, A.: A Mirror Metaphor Interaction System: Touching Remote Real Objects in an Augmented Reality Environment. In: ISMAR 2003, Tokyo, Japan, pp. 1–2 (2003)
15. Irie, K., Wakamura, N., Umeda, K.: Construction of an Intelligent Room Based on Gesture Recognition: Operation of Electric Appliances with Hand Gestures. In: IROS 2004, Sendai, Japan, pp. 193–198 (2004)
16. Jing, L., Yamagishi, K., Wang, J., Zhou, Y., Huang, T., Cheng, Z.: A Unified Method for Multiple Home Appliances Control through Static Finger Gestures. In: SAINT 2011, pp. 82–90. IEEE (2011)
17. Jung, J., Do, J., Kim, Y., Suh, K., Kim, D., Bien, Z.Z.: Advanced Robotic Residence for the Elderly/the Handicapped: Realization and User Evaluation. In: ICORR 2005, Chicago, Illinois, pp. 492–495 (2005)
18. Kane, S.K., Wobbrock, J.O., Ladner, R.E.: Usable Gestures for Blind People: Understanding Preference and Performance. In: CHI 2011, pp. 413–422 (2011)
19. Karam, M., Schraefel, M.: A Taxonomy of Gestures in Human Computer Interactions. Technical Report ECSTR-IAM05-009, University of Southampton (2005)
20. Kim, D., Kim, D.: An Intelligent Smart Home Control Using Body Gestures. In: ICHIT 2006, pp. 439–446. IEEE (2006)
21. Kim, H., Jeong, K., Kim, S., Han, T.: Ambient Wall: Smart Wall Display Interface which can be Controlled by Simple Gesture for Smart Home. In: SA 2011, Hong Kong, China, pp. 1–2 (2011)
22. Kita, S.: Cross-Cultural Variation of Speech-Accompanying Gesture: A review. Language and Cognitive Processes 24(2), 145–167 (2009)
23. Kleindienst, J., Macek, T., Serédi, L., Šedivý, J.: Vision-Enhanced Multi-Modal Interactions in Domotic Environments. In: ERCIM 2004 (2004)
24. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S.: I'm Home: Defining and Evaluating a Gesture Set for Smart-Home Control. Int. J. Hum.-Comput. Stud. 69, 693–704 (2011)
25. Lenman, S., Bretzner, L., Thuresson, B.: Using Marking Menus to Develop Command Sets for Computer Vision Based Hand Gesture Interfaces. In: NordiCHI 2002, pp. 239–242 (2002)
26. Lim, C.J., Pan, Y., Lee, J.: Human Factors and Design Issues in Multimodal (Speech/Gesture) Interface. In: JDCTA 2008, pp. 67–77 (2008)
27. Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer-Based Personalized Gesture Recognition and its Applications. Pervasive Mob. Comput. 5(6), 657–675 (2009)
28. Miranda, L.C., Hornung, H.H., Baranauskas, M.C.C.: Adjustable Interactive Rings for iDTV. IEEE Trans. Consum. Electron. 56(3), 1988–1996 (2010)
29. Neßelrath, R., Lu, C., Schulz, C.H., Frey, J., Alexandersson, J.: A Gesture Based System for Context-Sensitive Interaction with Smart Homes. Ambient Assisted Living, 1–12 (2011)
30. Nielsen, J.: Usability Engineering. Morgan Kauffmann, San Diego (1993)

31. Nielsen, M., Moeslund, T., Störring, M., Granum, E.: A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for Human Computer Interaction. In: Internation Gesture Workshop, Genova, Italy (2003)
32. Norman, D.A., Nielsen, J.: Gestural Interfaces: A Step Backward in Usability. Interactions 17(5), 46–49 (2010)
33. Norman, D.A.: Natural User Interfaces are Not Natural. Interactions 17(3), 6–10 (2010)
34. Oh, O., Jung, Y., Cho, Y., Hahm, C., Sin, H., Lee, J.: Hands-up: Motion Recognition Using Kinect and a Ceiling to Improve the Convenience of Human Life. In: CHI EA 2012, Texas, USA, pp. 1655–1660 (2012)
35. Ouchi, K., Esaka, N., Tamura, Y., Hirahara, M., Doi, M.: Magic Wand: An Intuitive Gesture Remote Control for Home Appliances. In: AMT 2005. IEEE (2005)
36. Panger, G.: Kinect in the Kitchen: Testing Depth Camera Interactions in Practical Home Environments. In: CHI 2012, pp. 1985–1990. ACM (2012)
37. Rahman, A.M., Hossain, M.A., Parra, J., Saddik, A.E.: Motion-Path Based Gesture Interaction with Smart Home Services. In: MM 2009, pp. 761–764. ACM (2009)
38. Rahman, A.S.M.M., Saboune, J., Saddik, A.E.: Motion-Path Based in Car Gesture Control of the Multimedia Devices. In: DIVANet 2011, pp. 69–76. ACM (2011)
39. Richarz, J., Plötz, T., Fink, G.A.: Real-Time Detection and Interpretation of 3D Deictic Gestures for Interaction with an Intelligent Environment. In: Int. Conf. on Pattern Recognition, pp. 1–4 (2008)
40. Rode, J.A.: A Theoretical Agenda for Feminist HCI. Interacting with Computer 23(5), 393–400 (2011)
41. Sadri, F.: Ambient Intelligence: A Survey. ACM Computing Surveys 43(4), 1–66 (2011)
42. Saizmaa, T., Kim, H.C.C.: A Holistic Understanding of HCI Perspectives on Smart Home. In: NCM 2008, pp. 59–65. IEEE (2008)
43. Solanki, U.V., Desai, N.H.: Hand Gesture Based Remote Control for Home Appliances: Handmote. In: WICT 2011, pp. 419–423. IEEE (2011)
44. Wahlster, W., Reithinger, N., Blocher, A.: SmartKom: Multimodal Communication with a Life-Like Character. In: Eurospeech 2001, pp. 1547–1550. ISCA (2001)
45. Wilson, A., Shafer, S.: XWand: UI for Intelligent Spaces. In: CHI 2003, pp. 545–552. ACM (2003)
46. Yamamoto, Y., Yoda, I., Sakaue, K.: Arm-Pointing Gesture Interface Using Surrounded Stereo Cameras System. In: ICPR 2004, pp. 965–970. IEEE (2004)

Thumbs Up: 3D Gesture Input on Mobile Phones Using the Front Facing Camera

Paul Schmieder[1], John Hosking[2], Andrew Luxton-Reilly[1], and Beryl Plimmer[1]

[1] University of Auckland, Private Bag 92019, Auckland, New Zealand
[2] Australian National University, RSISE Bldg, Canberra, ACT 0200, Australia
psch068@aucklanduni.ac.nz, john.hosking@anu.edu.au,
{beryl,andrew}@cs.auckland.ac.nz

Abstract. We use the front facing camera in a smart phone to capture gesture input. Thumb gestures performed above the camera are recognized and used to invoke commands. In contrast to other input modalities the camera requires no device movements and no valuable screen space is used. To be viable, this type of interaction requires gestures which are comfortable and memorable for the user and real-time accurate recognition of those gestures. Given the performance constraints of phones and their cameras we needed to determine whether accurate and reliable recognition is possible and identify types of gestures that are recognizable and user appropriate. As a proof of concept, we conducted a user study testing three gestures for performance and user satisfaction. The results demonstrate that the 3D gestural input is successful and we provide detailed insights into successful recognition strategies for this novel interaction modality.

Keywords: Motion gestures, mobile interaction, image recognition.

1 Introduction

Typical mobile devices have limited input modalities available: touch screen, two or three buttons, sensors, cameras and microphones. Gesture input has the advantage of supporting continuous input thus reducing the demand for input actions – one successful example of gesture input is the swipe keyboard [15]. The disadvantage of most continuous input modalities is that they use the touch screen which obstructs the information on the display, or the gyroscope or accelerometer that require distracting device movements [10]. Recent smart phones include front facing cameras. Their location means that such cameras are easily accessible away from the screen while their field of view is visible to the user and moving a thumb in this field does not require distracting device movements (Fig. 1). In this paper we explore how *3D gestures* can be used to provide *continuous input* captured by the *front facing camera*.

Using *gesture input* is common for a variety of devices and situations. A main difference is the number of dimensions gestures are performed in. For mobile devices, 2D gestures are common on the touch screen [4], including pinch and zoom, and

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 318–336, 2013.

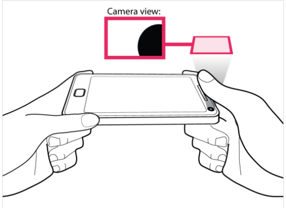


Fig. 1. A smartphone with a front facing camera and a thumb performing a gesture above it on the right

swipe. 2D touch gestures can also be combined with 3D motion gestures [9]. These use a variety of sensors: gyroscope [9], accelerometer [12] and cameras [21].

A *continuous input* mechanism is desirable, as it provides affordances suited for controlling continuous variables, such as volume. Ideally, a continuous input mechanism is application independent allowing fast and precise manipulation in both directions (e.g. volume up and down, fast for big changes or slow for precise adjustments) such as with the scroll wheels on MP3-Players like Apple's IPod Classic or Microsoft's Zune. However, to adjust volume most portable devices, including smartphones, use buttons. Alternatively continuous input can be captured by the touch screen but this means displaying an appropriate interface element occluding current information. Sensors in smart phones are also able to provide continuous input but usually require distracting and disruptive device movements. For example Hinckley and Song's techniques [9] combine motion and touch using the gyroscope and accelerometer most of which required motions disruptive to user tasks, such as tilting the device away or shaking it.

Gesture recognition is typically a computationally intensive process. Performing such recognition on mobile devices with limited processing power is challenging, particularly when it involves image capture. Wang et al [20] used a Motorola v710 to capture gesture input and reported a maximum rate of 15.2 frames per second (fps), but when image processing algorithms were applied to the input, the capture rate dropped to 12fps. Gu et al. [7] showed that many image processing algorithms reduced frame rates below one when implemented on a variety of different phones. To solve this problem, we developed novel recognition algorithms to reliably recognize the gestures in real-time on mobile devices with limited processing power.

Cameras in mobile devices have been widely used in research systems. This has focused on cameras on the back of the device tracking the devices' position [3] or hand movements [20, 21]. Kato [14] attached a spring with a marker at its end above the back facing camera, allowing continuous input by tracking the marker in three dimensions. However, due to the camera position and the attached spring, operating the marker was rather constrained. Huerst et al. [11] tracked fingers with the back facing camera using markers attached to the tips to capture gestures.

Here we investigate the use of a front facing camera on a mobile device for continuous gesture based input, a combination we believe is novel. Using a front facing camera has multiple advantages: first, the required hardware, camera and finger are already present. Second, the input is independent of application (except for those directly using the front facing camera) and therefore always accessible. Third, no distracting device movements are required. Fourth, by offering an interaction space away from the display, screen occlusion is reduced. Fifth, the camera can be combined with other sensors to enrich interaction.

There are three components required to support this interaction modality: the *gesture capture,* the *gesture recognition,* and the *gesture design*: all of which contribute to the user experience. To test the feasibility of using three-dimensional gestures for continuous input, we defined three gestures and developed appropriate recognizers with varying demands on usability and hardware. In the second phase we evaluated the gestures on a smart phone in different scenarios. Besides the gesture recognizer performance we were interested in the demands on participants in terms of interaction space and the comfort of the thumb movements.

The results yield three contributions. First, continuous input can be *successfully captured* by front facing phone cameras. Second, current smart phones can *accurately recognize* 3D gestures in real-time. Third, our evaluation provides insights into the *design implications* of our gestures on usability including the definition of the interaction space above front facing cameras where gestures can be comfortably performed.

2 Related Work

Our work builds on related research involving interaction techniques for mobile devices; camera based phone applications; gesture input; and computer vision input.

2.1 Interaction Techniques for Mobile Devices

Some research focuses on facilitating mode manipulations based on the input channels available in mobile devices. For example Hinckley and Song [9] explored gestures combining touch and motion using gyroscope, accelerometer and touch screen. They propose motions for continuous input such as tilt-to-zoom and discrete input such as hard-tap. Research is not limited to the sensor abilities available in current devices. Brewster and Hughes [5] investigated the use of pressure on touchscreens to differentiate between small and capital letters entered on the phone's keyboard.

2.2 Camera Based Interaction on Mobile Devices

Camera based techniques on mobile devices can be differentiated by camera location: integrated in the device or external. Vardy et al. [19] used a camera attached to the wrist to capture finger movements feeding the recognized gestures to a computer for interaction. More common is to use device integrated cameras which are predominantly on the rear. TinyMotion [20] uses a rear facing camera to measure cell phone movements. It computes phone tilts to control games such as Tetris and text input. TinyMotion requires a button press combined with a phone tilt to input a character [20]. Another application is a see-through tool for applications such as maps [2].The device's position over a surface is recognized and augmented with information.

Kato and Kato [14]developed a tangible controller for continuous analog input. They attached a spring on top of the rear-facing camera with a marker attached to its end. The user provides input by moving the marker in any direction.

2.3 Gesture Input

Gestures, including touch and motion gestures, are commonly used for mobile device interaction. Bragdon et al. [4] investigated the impact of distractions on touch interaction. Comparing gestures and their soft button equivalent, they found gestures are less impacted by distractions and perform as well as soft buttons without distractions.

Aiming to build a gesture set for smartphones Ruiz and Lank [18] conducted a study in which end-users proposed gestures for given scenarios. They found consensus among study participants regarding movement parameters and gesture-to-command mappings. The resulting taxonomy differentiates between *how* a gesture is mapped to a command and a gesture's physical characteristics. Using thumb input on touch screens is widely researched. For example, Karlson et al. looked at the interaction between hands and mobile devices including user preferences and mechanical limitations [13].

2.4 Computer Vision Based Input

Computer vision systems are widely used to collect visual information for gesture based input [6]. The choice of technique depends on factors such as desired information, environmental circumstances and available computational power. Commonly sought information is recognition of objects and/or gestures [16] and object tracking [22]. To gain this information environmental factors such as lighting, motion speed, blur and occlusions as well as their speed of change must be considered [17].

Tracking a large fast moving object dominating a scene is challenging. Fast optical flow algorithms and graph cut approaches are viable but computationally expensive [1]. With limited compute power, temporal difference images have been used successfully to detect coarse-scale motion [6]. To further deal with changing lighting and blur, other properties of the tracked object such as object shape and color [23].

3 Gestures

The novelty of using front facing cameras for gesture input on a mobile device means the feasibility of such a system is unknown. To address this, we conducted a user study testing three gestures of varying demands on hardware performance and user input capability. This study addresses three questions: 1) are current mobile devices powerful enough for camera based gesture recognition in real-time? 2) are 3D gestures performed above the camera an easy and precise way to capture input? 3) what is the interaction space above the camera where gestures can be comfortably performed? To answer these, gestures of differing complexity were tested for recognition accuracy and usability in different scenarios. For the development and testing we exclusively use a Samsung Omnia W, featuring a single core 1.4 GHz Scorpion processor and Windows Phone 7. As results will show, camera based gestural input can be successfully used in mobile devices for enriched natural interaction.

3.1 Design

We designed three gestures and an accompanying recognition algorithm to accurately identify each gesture. To allow for continuous input, each gesture can be performed such that it differentiates between two motion directions: for example to increase and decrease the volume, to zoom in and out of an image or to scroll a list up and down. As the motion direction depends on the phone's orientation, the orientation was used by the recognizers to decide what motion direction maps in to what input direction. Additionally, each gesture has different levels of acceleration: for example, one level may increase the volume by one each iteration, another by three and another by five. Each gesture has a neutral state where no changes occur.

Before the gestures are recognized each captured image is preprocessed. The first step scales the image to 40 x 30. We tested different resolutions for the gesture recognizers and found 40 x 30 to be the best tradeoff between performance and accuracy. Additionally, the image is converted from 32 bit RGB to 8 bit grayscale.

The thresholds used for the algorithms were obtained by testing the algorithms under different lighting conditions: bright: outside during a sunny day; neutral daylight: in the office with open blinds; artificial light: in the office with blinds down and light on, half dark: in the office with blinds down. Completely dark was not trialed as the algorithms require some light to track the thumb.

Tilt Gesture. The thumb is lying on the camera and its tilt determines the input (Fig. 2). The tilt direction and intensity determines the direction and acceleration of change in input. Tilting the thumb to the right maps to an increase in input and tilting to the left a decrease. The higher the tilt is the higher the level of acceleration. If no tilt is applied to the thumb covering the camera no changes in input are invoked.

The recognition algorithm analyzes the light captured by the camera to calculate the input. The first step is to split the captured image in half depending on the phone orientation: the split divides the image so that each half is covered by a half of the thumb. Second, the average brightness value is computed over the pixels in each half. The image half with the lower brightness is the side the thumb is tilted towards. Similar low brightness values mean that the thumb evenly covers both image halves so no input changes are triggered. Finally, based on the brightness values and thus the amount of tilt, a step function is used to determine the level of input acceleration: the brighter the image half the higher the acceleration (Fig. 2).

A brightness threshold determines if the tilt gesture is currently being applied. The gesture recognizer is only used if the average brightness over the complete image is below a predefined threshold. This threshold is set to 55% of an image brightness value taken when nothing covers the camera (e.g. when the application is started).

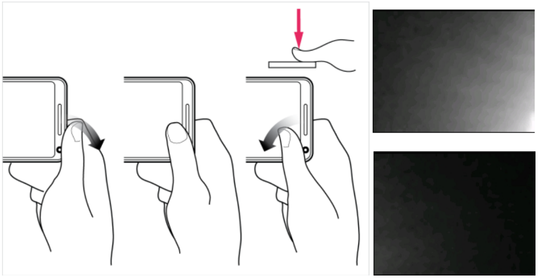


Fig. 2. The tilt gesture: Left: Tilting the finger to either side determines the direction of input and the amount of tilt determines the level of acceleration. Right: The captured images used to compute the tilt. (top) A strong tilt to the left and (bottom) a slight tilt to the right.

The chosen threshold provides the best offset between brightness ranges to determine the level of acceleration and activate/deactivate the gesture recognizer.

Distance Gesture. The distance between thumb and front facing camera is used to adjust the input (Fig. 3). A predefined distance marks a neutral point where no input manipulation happens. A smaller thumb-camera distance triggers a decreased input and a larger one an increase. The exact distance determines the level of acceleration. Input is based on the thumb's diameter in the captured image: the larger the diameter the closer the thumb. To calculate the distance, the thumb's position in the captured image has to be determined. Then, brightness values are computed over the image. Finally, the input value is determined using the thumb's distance and brightness.

To detect the thumb in the image a Sobel edge detector with a 3x3 kernel is applied. The resulting Sobel image (Fig. 3) contains objects defined by their borders. To identify the thumb three assumptions are made: 1) the thumb partly covers the center row of the Sobel image; 2) in the center row the thumb is the biggest object; 3) the thumb forms an uninterrupted blob starting at the image. Hence the recognition algorithm analyzes the center row of the image looking for the biggest group of dark pixels bounded by white pixels. If the biggest section is above half the image's width, the algorithm further tests whether the section is part of a blob connected to at least one image edge. If no object satisfies all three the criteria the algorithm discards the image and no input is invoked assuming the gesture is not currently being performed.

The size of the range between minimum and maximum thumb distance is divided into smaller spans, each accommodating a different input value. A span has to be sufficiently large to be easily found by the user in terms of distance of his/her thumb to the camera. For example, in a 40 pixel wide image the total range is 20 as the thumb must be at least half the image width wide. This leaves 20 pixels to be split into smaller spans. We found spans of one-eighth of the image width to work best. Brightness values determine if a thumb is covering the camera completely. This allows further differentiations when the thumb covers the captured image width completely but does not yet physically cover the camera.

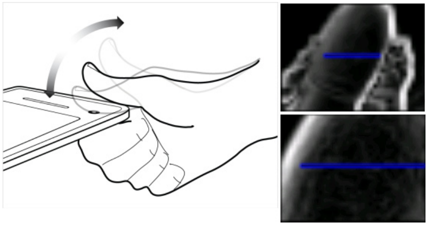


Fig. 3. The distance gesture: Left: the input depends on the distance between camera and thumb. A predefined distance is the neutral input where no changes are invoked. Right: The images after the Sobel operator has been applied with the thumb further away (top) and close (bottom) to the camera.

Circular Gesture. A circular motion is performed close above the camera to adjust the input (Fig. 4). Depending on the motion direction (clockwise or counter-clockwise) input is either increased or decreased. The speed of the gesture determines the level of acceleration.

The circle gesture recognizer first detects the thumb. Afterwards the thumb's center of mass is used to calculate the motion direction. Motion speed is based on the overlap of thumbs between two consecutive images.

To detect the thumb a flood fill algorithm is used based on the captured image's greyscale values. The thumb appears darker than the rest due to the movement close to the camera: cameras found in smart phones are not fast enough to adjust their blending settings to evenly illuminate the image. As the thumb changes illumination conditions quickly, the camera makes the background behind the thumb appear much brighter than normal. The flood fill algorithm finds the biggest connected thumb blob composed of pixels with an 8-bit greyscale value below 90. If the biggest blob covers more than 21% of the captured image it is assumed to be the thumb. Otherwise the thumb is assumed not to be in the image. If no thumb is detected in more than four successive images further computational steps are stopped and only resumed when a thumb is found in at least four successive images.

Motion direction is calculated using the cross products of three successive images. The weighted average location for each thumb blob is calculated to determine its center of mass (Fig. 4): the center of mass's x-coordinate is the average value over all the blob pixels x values. The same applies to the y-coordinate. Afterwards, the cross products between the four most recent thumb centers of mass are computed. If the sum of the three resulting vector magnitudes is positive the gesture motion is clockwise and otherwise counter clockwise.

Motion speed depends on movement and on the number of intersection points between two successive thumb blobs. To detect a stationary thumb the image is divided into four even quadrants. If the thumb remains in the same quadrant over four successive images, the motion counts as stopped. If not, the thumb is moving and the overlap of two thumbs in successive images is computed. A step function based on the overlap percentages is used to define thresholds for the different levels of acceleration depending on the number of acceleration levels. For example, given two acceleration levels, our empirical tests showed a threshold of 42% to be optimal: a higher overlap than 42% means a slower acceleration and a smaller overlap a higher acceleration. Using quadrants to detect a stationary thumb rather than overlaps is better as it compensates for small unintentional thumb movements.

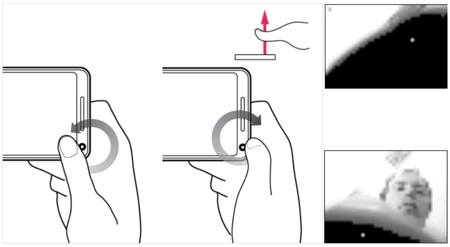


Fig. 4. The circular gesture: Left: The direction of the circular motion determines the input direction and the speed of the motion determines the level of acceleration. Holding the finger at a position invokes no change (i.e. neutral input). Right: The images with the thumb in black after the flood fill algorithm. The white dot in the black blob indicates the center of mass.

3.2 Performance and Robustness

The computation of a new input value consists of multiple steps of which capturing the image is the most time costly (Table 1). If no other computations are performed the Samsung Omnia W captures up to 16 images per second. In the following we comment on the robustness and accuracy of the recognizers in general and under different lighting conditions (see section 3.1).

The tilt and circle gestures are both unaffected by different lighting conditions and fast lighting changes. Only when it was completely dark did both recognizers stopped working. The main reason for the robustness is that in both cases the thumb dominates the captured images making the background appears bright and unfocused (Fig. 4).

The distance gesture worked perfectly under the tested lighting conditions except in the half dark condition. In this condition if there were multiple dark objects in the image background they were occasionally merged with the thumb: because of the insufficient lighting the Sobel operator did not detect all the edges. Fast lighting changes did not affect the algorithm.

Table 1. Performance measurements in milliseconds using a Samsung Omnia W phone

	Tilt Gesture	Distance Gesture	Circular Gesture
Pre-process Image	image captured (66.75msec) scaled & converted to greyscale (1.9msec)		
Process Image	-	4.93 (Sobel)	
Detect Thumb	-	1.12	2.88 (Flood Fill & mass center)
Compute Direction & Acceleration	0.97	0.51	1.76

4 Evaluation

We conducted an evaluation study to explore the interaction with a smartphone using continuous gestures captured by a smartphone's front facing camera. This study focuses on the requirements of suitable 3D gestures and the space they are performed in with the aim to inform further research regarding the basic requirements of this novel and yet unexplored interaction. Four pilot participants were used to determine algorithm parameters such as the number of acceleration levels and the rate at which changes were applied. Only one of the recognizers was activated at each point in time requiring a button press to change between them. We were specifically interested in the following questions:

- What defines the interaction space in which gestures can be comfortably performed above the camera?
- Are the evaluated gestures easy and intuitive and are their associated apps an appropriate match?
- Is the phone's hardware sufficiently powerful to guarantee real-time performance?
- Are the gestures recognized accurately in terms of direction and acceleration changes and reactiveness?

4.1 Participants

Participants had to have a smartphone as their mobile device to guarantee device familiarity. We recruited 24 participants, 8 females and 16 males, ranging in age from 19 to 41 (mean: 25.63 years, SD = 5.17). Five were left-handed. There was no compensation for participation. Since accurate tracking objects may depend on the visual properties of the tracked object, we wanted to ensure a variety of different thumb size, shape and color were represented in the evaluation. To test the robustness of our algorithms we intentionally recruited participants from different ethnic groups (10 Caucasians, 6 Indians, 5 Asians and 3 from the Middle East) and of different height as an approximation of thumb size (mean: 171cm, SD = 8.62).

4.2 Apparatus

To test the gestures, participants used a Samsung Omnia W I8350 featuring a single core 1.4 GHz Scorpion processor and Windows Phone 7. The front facing camera was located on the shorter side's corner (Fig. 1). The capture rate of the front facing camera is on average between 8-10 images per second on the Windows Phone

The gesture recognizers were integrated in a program running on the phone. Each recognizer was configured to support two directions and to differentiate between two acceleration levels: one and three. We decided to use two acceleration levels based on the results from the four pilot study participants. While the tilt and circle gestures would have worked intuitively enough with three acceleration levels, the distance gesture did not. When the thumb was moved with the lower acceleration changes were applied every 500ms to allow for accurate input. This low change rate was chosen as the pilot study participants frequently overshot the target with a faster change rate. For the faster acceleration level changes were applied after each image was analyzed and the new value computed (i.e. on average every 75msecs). Sessions were recorded using a video camera mounted above the participants.

4.3 Tasks

We designed three tasks each dedicated to a particular gesture and one for familiarization across all three gestures. Each task required participants to use all facets of the gesture so that the invoked input direction and level of acceleration were frequently changed. The choice of task for a particular gesture was based on similar mappings found in current environments as explained in the following. The order of gestures was varied between the participants forming six conditions.

Contact List – Circle Gesture. To scroll a list using a circular motion is familiar practice. MP3 players such as Apple's iPod Classic use a circular motion to browse their content. Scrolling a list requires participants to bridge bigger distances faster as well as precise manipulations to jump between neighboring list entries.

The scrolling mini app displayed a list of 78 contacts to simulate a standard smartphone Contact list (Fig. 5a). To scroll down/up, participants performed a clockwise/anti-clockwise motion. If the acceleration level was 1 the selector, indicated by a blue color, jumped from the currently selected contact to the next neighbor. An acceleration level of three jumped three contacts at once.

In the top right corner a name was displayed which was also in the contact list. Participants were asked to scroll to that name in the list as fast as possible. The correct name had to be selected for two seconds requiring participants to provide neutral input. After two seconds the name in the top right corner was replaced by another name from the contact list. In total, participants had to scroll to 10 different names. The choice of names made participants perform precise selections (e.g. to go from one name to a neighbor) and to bridge larger gaps (e.g. Paul to Billie).

Zooming – Distance Gesture. Participants controlled the size of a green square on the screen (Fig. 5b). To zoom in, participants moved their thumb close to the camera and to zoom out the thumb was moved further away. This gesture is familiar for zooming as going closer to an object makes it naturally appear bigger and going further away makes it smaller.

An acceleration level of one increased or decreased the square's size by 10 pixels and a level three by 30 pixels. The square can have any width between 0 and 300 pixels resulting in 30 distinct zoom-able sizes.

The participants' aim was to match their square's size with a red frame shown on the display. To indicate equal sizes, the red frame turned white. Participants had to keep the size constant for two seconds, requiring them to adjust the thumb camera position to neutral, before the frame changed size and turned red again. In total the red frame changed size 10 times.

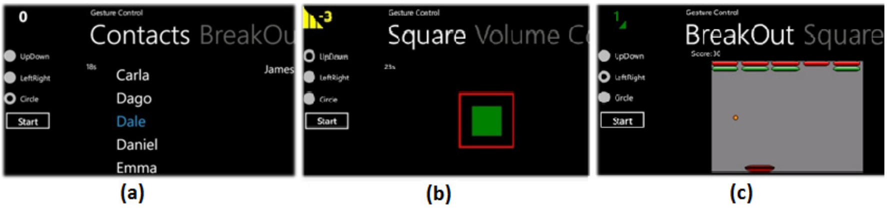


Fig. 5. The test program showing the mini apps: (a) the contact app, (b) the zooming app and (c) the Break-Out app. In the top left side an indicator is shown displaying the currently captured input ((a) neutral, (b) fast-left, (c) slow-right).

BreakOut - Tilting Gesture. Participants controlled the paddle in BreakOut (Fig. 5c) to bounce the ball upwards. The aim was to destroy all ten bricks lined in the top of the game area by hitting each with the ball. After a brick was hit, the ball was deflected down towards the paddle. The angle of deflection when the ball hit the paddle depended on where the ball hit the paddle. To destroy the bricks the user needed both small and large paddle position adjustments forcing them to use all facets of the tilt gesture. The game area was 300 pixels wide. An acceleration level of one meant a paddle move of 4 pixels while a level of three meant a 12 pixel move. In total the paddle had 75 distinct positions. Each destroyed brick scored 10 points making 100 points the highest score. The game was also used in previous research to study camera based input techniques [8].

4.4 Procedure

Participants were seated at a table with the smartphone lying immediately in front of them. They were given a tutorial guiding them through the study. The participants' first task was to complete the first section of the questionnaire.

To familiarize themselves with the Windows Phone participants were asked to pick it up and the facilitator explained the main elements on the screen. Once familiar, the participants started the test application. During this the phone was held in portrait orientation before changing to landscape to interact with the test application.

At the beginning of each gesture the facilitator explained the gesture to the participant. Afterwards the participant was told to navigate to the volume mini app and trial the gesture. Then the participant completed both a gesture specific task and the questionnaire for this gesture. This was repeated for each gesture.

After all tasks were finished the final section of the questionnaire was completed. The experiment concluded with a discussion of other possible gestures allowing continuous input. It took the participants on average 29 minutes to complete the study.

5 Results

The collected data includes the questionnaires, session video recordings and transcripts of the final discussion. The questionnaire comprised three sections. The first gauged participants' familiarity with gesture input. The second asked participants to rate the gestures regarding their ease and recognition accuracy using a five point Likert scale. The last contained a ranking of the gestures as well as questions gauging the interaction space above the camera including camera location.

5.1 Interaction

All 24 participants held the smartphone in landscape orientation with the camera on the right. Regardless of their handedness all participants chose to use their right thumb to perform the gestures. The matches between gesture and associated mini apps were never rated as negative thus confirming our design choices (Fig. 6).

Ideal Camera Position. The camera should ideally be centered on the horizontal axis in landscape mode on the phone's right side. When asked in the questionnaire to indicate the optimal camera position on the phone all participants chose the right side of the screen. Six participants chose the top corner, eight the center and nine the lower corner. Ten participants noted that the camera on the horizontal axis should be centered between screen and phone edge. The camera on the test phone was closer to the edge than to the screen which sometimes created problems with the tilt and circle gesture: a tilt to the right (of a big thumb) meant a tilt slightly over the phone's edge. While performing the circle gesture some participants drifted away from the camera's center. They stated that they used the space between screen and phone edge as a frame of reference with their circle motion touching both sides.

Keeping the Thumb in the Camera's Viewing Field. Each gesture used a different interaction space above the camera with different demands on the user. The tilt gesture was the easiest one to keep in the camera's viewing field as it essentially was performed as a two dimensional gesture with the thumb lying flat on the camera (Fig. 6). It received ratings above neutral with 83% of the participants strongly agreeing to the ease and 17% agreeing. In contrast, the ease of keeping the distance and circle gestures in the camera' viewing field was rated lower.

The dominant usability problem with the distance gesture was to keep it in the viewing field. To measure the distance between camera and phone, the recognition algorithm calculates the thumb's diameter. If the thumb is close but not centered above the camera, it is only partially visible thus its measured diameter is smaller than it actually is so the calculated distance is incorrect. As the thumb moves away from the camera it does so in an arc. Its tip moves towards the bottom of the phone as it moves up thus leaving the camera's viewing field on the lower edge. However, this usually occurs after the thumb position starts feeling uncomfortable due to the high finger stretch. In the questionnaire we asked participants what the maximum distance between thumb and camera is before holding the thumb becomes uncomfortable. Measured with a ruler, the average distance stated was 4.75 cm (SD = 1.12cm, Min = 2 cm, Max = 6 cm). The majority of the participants stated that it becomes uncomfortable just before thumb and thumb socket form a straight line.

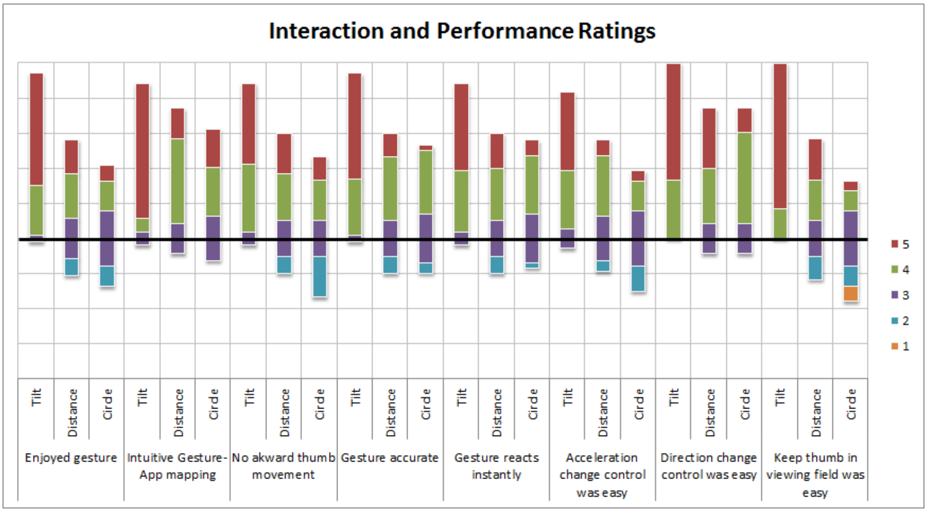


Fig. 6. The questionnaire ratings for the three gestures on a 5 point Likert scale centered around neutral. The ratings are: 1= strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.

For correct recognition, the circle gesture had to be performed at a certain height above the camera and the motion had to circle the camera. The gesture did not require participants to draw a perfectly round circle nor one perfectly centered above the camera. However, moving outside the camera's viewing field, as some participants did, made it impossible to track the thumb resulting in incorrect recognition. On occasion the thumb was too close to the camera so the captured image was almost totally dark leading to incorrect recognition. Participants on average found it harder to keep the thumb in the right area for the circle compared to the distance gesture with the distance gesture having 17% rating it below average and the circle gesture 29%. This was also reflected in the questionnaire where eight participants said it was difficult to keep the circular movement above the camera. There were no complaints about the distance gesture.

In summary, the tilt gesture is the easiest with only positive ratings. The distance gesture is the second highest rated followed by the circle.

Ease of Gesture Motions. The thumb movements for the distance and circle gestures were perceived as physically uncomfortable at times. Having to hold the thumb still at changing positions over an extended duration (Zooming task: Ave = 141 sec, SD = 36) was perceived as uncomfortable by 12.5% of the participants. For the circle gesture, 29% of the participants experienced discomfort when performing the gesture: three stated their thumb movement was restricted as they were holding the phone with the same hand. This was particularly evident when the thumb tip moved close to the thumb base forming a right angle between the thumb's first and second phalanges.

5.2 Phone Performance and Recognition Accuracy

All gesture recognizers and mini applications ran in real-time on the phone. In the questionnaire participants rated the overall accuracy of the gesture recognizer as well as the recognizers' reactiveness and the ease and correctness of changing the acceleration level and direction (Fig. 6).

The gestures and their associated mini applications ran smoothly apart from several occurrences when the camera froze for two seconds during interaction. This occurred randomly for different recognizers and appeared to be an artifact of the phone's API to the camera buffer. We speculate that this is unlikely to be an issue with a multicore processor phone where camera and interface run on different threads.

The recognizers' reactiveness was rated neutral and above for the tilt gesture and neutral and above by 21 participants for the distance gesture and by 23 for the circle gesture (Fig. 6). To allow for precise adjustments changes were applied every 500ms, which some participants perceived as too slow. For the distance gesture, changes were invoked instantly when the thumb distance entered a new range. For example, to zoom in with an acceleration level of one the thumb had to be roughly (depending on thumb size) between 3 and 4 cm away. Any other distance range was mapped to another input. If the thumb position was changed but not enough to change ranges, participants perceived this as slow to react. For the circle gesture the thumb had to rotate half a circle to trigger any change in input. A lower threshold meant that sometimes unintended changes were invoked as the circular motion was not performed.

Changing the level of acceleration was rated below neutral for the distance and circle gesture by two and five participants respectively (Fig. 6). For the distance gesture, to change the level of acceleration, the finger has to be in the distance range associated with the desired acceleration. To find the correct range one either recalls it from practice/experience or moves towards the desired range until the input value changes to the desired outcome. As participants had little training not all remembered the positions perfectly so had to use the second strategy which some perceived as suboptimal. For the circle gestures, correctly recognizing the acceleration required the circle to be performed above the center of the camera. If this condition was violated the recognizer could not accurately differentiate between the acceleration levels of one and three. Due to the problems of keeping the thumb in the required space (Fig. 6), the acceleration levels were occasionally recognized incorrectly.

Changing the input direction was always rated as neutral or higher for all gestures (Fig. 6). Differentiating between directions was easier than changing the acceleration levels because the motions were more different thus easier to remember and recognize. For example, for the circle gesture one had to change the motion direction regardless of acceleration. Additionally, to accurately recognize the correct direction it was sufficient for the camera to capture part of the circular motion thus not requiring a centered motion above the camera.

To sum-up, well designed gesture recognition algorithms can run in real-time on current generation smartphones. The gestures allow for accurate adjustments of acceleration and directions if the gesture requirements such as motion path and accuracy are satisfied.

5.3 Summary and New Gestures

Both, natural gestures and accurate recognizers are required for satisfactory interaction. If a gesture is too complicated to be intuitively performed, a recognizer may be unable to deal with the inaccurately performed motion. The three tested gestures varied in terms of motion complexity; this was reflected by how much participants enjoyed each gesture (Fig. 6) and the final gesture ranking. The tilt was the easiest to perform and was ranked first by 21 of 24 participants and second by the remaining participants. The distance gesture was more difficult to perform mainly because the desired camera-thumb distances were hard to remember with the little practice. Thus, the distance gesture was ranked second by over half the participants (1^{st}: 2, 2^{nd}: 13, 3^{rd}: 9 participants). The circle was the most difficult of the three, as it required the participants to first, perform the motion at a certain distance above the camera while, second, centering the motion. The gesture accurately recognized only if both criteria were satisfied. 15 participants ranked the circle gesture third, 8 second and 1 first.

When asked for other gestures, participants proposed one of three gestures. The first, proposed by four participants, is a circular motion but on the z-axis above the camera. Thus staying inside the camera's viewing field would become easier. Circle direction and speed determine acceleration level and direction. The second and third are variants of swipe motions. The second, proposed by four, consists of a horizontal swipe motion with the swipe direction determining the input direction. Once the swipe in a direction reached the outer border the finger is either lifted high above the camera or besides it to return to the other side and continue the gesture. The third proposed by seven has an additional horizontal swipe that makes the extra motion to return to the start position redundant.

6 Discussion

The aim of the presented work was to investigate the potential of a yet untested interaction to capture continuous input with the focus on performance and 3D interaction. Due to the new challenges we felt that such an evaluation was required before other evaluations such as comparing the proposed techniques with state-of-the-art interaction methods (e.g. touch screen based input) were conducted. Additionally, with the knowledge gained a more detailed study looking at creating user defined gesture set is possible similar to [18].

6.1 Capturing and Recognizing Gestures

Gesture recognition using front facing cameras has to run in real-time despite a mobile devices' limited hardware performance. To track objects like a thumb a number of tracking approaches may be *unsuitable* due to high performance requirements. We found the following methods provide sufficient information while allowing for real time interaction: working with the image's *brightness, edge detection* and *differential images*; the last is unsuitable if fast movements are involved due to fast lighting changes.

A captured image's brightness (or the excess of it) can be used to track moving objects near the camera. If a tracked object moves fast and covers at least quarter of the image, the background is overexposed making it too bright with almost no contours. Using a threshold to extract the close darker object from the background is easy and has low performance cost as shown by the circle recognizer.

To track an object at any distance that is not moving, fast edge detectors can be used. To identify the object based on its edges additional information is required such as its shape, origin and/or size. The last two were successfully used for the distance gesture where the finger always originates from at least one corner and covers at least a certain proportion of the image.

Capturing the gestures with the camera using these algorithms has some disadvantages. Constant capturing of images consumes more power draining the battery faster. While it does not matter whether the light is natural or artificial, sufficient lighting is still required: our tests showed that except for the Sobel operator, the algorithms worked in all conditions except the completely dark one.

6.2 Designing Gestures

Designing successful motion gestures means creating intuitive gestures without constraints. As the camera has a limited field of view gestures have to be fitted to this area: gestures close to the camera have to be directly above it while those further away can be further away from the camera's center. However, gestures cannot be too far away as this becomes uncomfortable to the point where holding the device with both hands becomes impossible.

Each state in a gesture motion must be easily identifiable and performable. If it is unclear what motion to perform to effect change or if the motion requires physically uncomfortable moves, the gesture may be unsuitable for continuous input. The distance gesture showed that having no explicit knowledge of the required motion to change to a given acceleration level may dissatisfy users.

The number of acceleration levels depends on the differentiability of the gesture motions. For example, with the distance gesture there was a total range of 6 cm available, which had to be divided in sub-ranges each dedicated to a unique input value. Trying to divide this range into more than five ranges (two directions with two acceleration levels each plus a neutral range) would be unrealistic as users' ability to discriminate those ranges is too difficult as shown in the evaluation. User feedback is important, a change in gesture must immediately cause a change of input while providing sufficient control for small changes.

6.3 Designing the Device

The camera position preference, regardless of the users' handedness, is on the right side of the phone. When asked for their preferred camera position, the participants' answers did not show any strong preference between the vertical position (top, center or bottom). However, there was a preference for the camera to be centered on the phone border.

Regardless of the camera position the gestures did not impair screen visibility. When asked, participants stated that they were able to see the entire screen during the interaction. An analysis of the video recording showed that items of importance such as the next name for the scrolling tasks situated on the screen's right border next to the camera were no covered by any gesture.

7 Conclusion and Future Work

We developed, implemented and evaluated three gestures and associated recognizers to allow for continuous input on portable devices. To capture input the front facing camera is used as its location offers an easily reachable input channel away from the screen. Using camera-based gestures avoids the need to move the device, which would potentially interrupt the interaction, and does not occupy valuable screen space.

Our work shows how 3D *gestures* are successfully *captured* and *recognized*. Different low-cost recognition strategies are implemented and trialed. The evaluation demonstrates the robustness of the recognition algorithms under different conditions; e.g. different lighting conditions and different colors and sizes of the tracked thumbs. The evaluation also provides insights into, first, the design of gestures and their requirements and, second, the interaction space above front facing cameras where gestures are comfortably performed.

Motivated by the results of our study we are exploring further gestures for discrete and continuous input. We are also looking at combining camera-based gestures with other input modalities such as data captured by gyroscope and accelerometer. Motivated by [13], we are also exploring making camera based interaction accessible for one handed interaction.

References

1. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. International Journal of Computer Vision 12(1), 43–77 (1994)
2. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and magic lenses: the see-through interface. In: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, pp. 73–80. ACM (1993)
3. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 2287–2296. ACM (2010)
4. Bragdon, A., Nelson, E., Li, Y., Hinckley, K.: Experimental analysis of touch-screen gesture designs in mobile environments. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 403–412. ACM (2011)
5. Brewster, S.A., Hughes, M.: Pressure-Based Text Entry for Mobile Devices. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–4. ACM (2009)
6. Freeman, W.T., Anderson, D.B., Beardsley, P., Dodge, C.N., Roth, M., Weissman, C.D., Yerazunis, W.S., Kage, H., Kyuma, K., Miyake, Y., Tanaka, K.I.: Computer vision for interactive computer graphics. IEEE Computer Graphics and Applications 18(3), 42–53 (1998)
7. Gu, J., Mukundan, R., Billinghurst, M.: Developing Mobile Phone AR Applications using J2ME. In: 23rd International Conference on Image and Vision Computing New Zealand, IVCNZ 2008, pp. 1–6. IEEE (2008)
8. Hansen, T.R., Eriksson, E., Lykke-Olesen, A.: Use Your Head: Exploring Face Tracking for Mobile Interaction. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 845–850. ACM (2006)
9. Hinckley, K., Song, H.: Sensor synaesthesia: touch in motion, and motion in touch. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 801–810. ACM (2011)
10. Holz, C., Baudisch, P.: Understanding Touch. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2501–2510. ACM (2011)
11. Hürst, W., Wezel, C.: Gesture-based interaction via finger tracking for mobile augmented reality. Multimedia Tools and Applications, p. 1–26 (2012)
12. Iwasaki, K., Miyaki, T., Rekimoto, J.: Expressive typing: a new way to sense typing pressure and its applications. In: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4369–4374. ACM (2009)
13. Karlson, A.K., Bederson, B.B., Contreras-Vidal, J.: Understanding one handed use of mobile devices. In: Handbook of Research on User Interface Design and Evaluation for Mobile Technology (2008)
14. Kato, H., Kato, T.: A camera-based tangible controller for cellular phones. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–2. ACM (2009)
15. Kushler, C.A., Marsden, R.J.: System and method for continuous stroke word-based text input. ForWord Input Inc., USA (2003)
16. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(3), 311–324 (2007)
17. Moeslund, T.B., Granum, E.: A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding 81(3), 231–268 (2001)

18. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 197–206. ACM (2011)
19. Vardy, A., Robinson, J., Li-Te, C.: The WristCam as input device. In: The Third International Symposium on Wearable Computers, Digest of Papers, pp. 199–202 (1999)
20. Wang, J., Zhai, S., Canny, J.: Camera phone based motion sensing: interaction techniques, applications and performance study. In: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, pp. 101–110. ACM (2006)
21. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., Shen Lucid, C.: touch: a see-through mobile device. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, pp. 269–278. ACM (2007)
22. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38(4), 13 (2006)
23. Yilmaz, A., Xin, L., Shah, M.: Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(11), 1531–1536 (2004)

User and System Cross-Learning of Gesture Commands on Pen-Based Devices

PeiYu Li, Manuel Bouillon, Eric Anquetil, and Grégoire Richard

IRISA/INSA Rennes, Campus de Beaulieu, 263 Avenue du Général Leclerc, 35042 Rennes, France
{Pei-Yu.Li,Manuel.Bouillon,Eric.Anquetil, Gregoire.Richard}@irisa.fr

Abstract. This paper presents a new design and evaluation of customizable gesture commands on pen-based devices. Our objective is to help users during the definition of gestures by detecting confusion among gestures. We also help the memorization gestures with the guide of a new type of menu "Customizable Gesture Menus". These menus are associated with an evolving gesture recognition engine that learns incrementally, starting from few data samples. Our research focuses on making user and recognition system learn at the same time, hence the term "cross-learning". Three experimentations are presented in details in this paper to support these ideas.

Keywords: Handwritten gesture recognition, Marking Menus, Customizable gesture interfaces.

1 Introduction

To facilitate the interaction between human and computers, a lot of new concepts appeared, like interactive tablets (using a finger or a pen). These tablets offer users a new way to interact: a user can draw a gesture to run a command. It is easy to remember all gestures for a simple application with less than ten commands. But when there are too many possible commands, it becomes difficult to remember them all directly. This is one of the challenges of the research on Pen-based Human Computer Interaction nowadays.

In this paper we study how to obtain natural cursive gestures and how to design a framework which helps users learn them efficiently. We can identify several ways to improve the learning process of the user to memorize all the gestures. The first point we have identified in our experiments, reported below, is that it is easier for users to memorize gestures if they have defined the set of gestures themselves. In fact, a "meaningful" gesture for a command can be memorized more easily than an arbitrary gesture without meaning. A meaningful gesture should have a semantic meaning or an ergonomic meaning. Several questions are interesting to address to go in this way: How to define a meaningful gesture? How to help users define and memorize their set of gesture commands? To offer the user the ability to define their own gestures, we

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 337–355, 2013.

use a gesture recognition system able to learn from scratch with very few samples. Moreover, this recognition system learns incrementally with the user interaction to improve its performance during its use. As user and recognition system learn at the same time during utilization, we study the notion of "cross-learning".

Besides, to help users learn their gestures, we would like to take advantage of Marking Menus [1, 2] which can organize gestures in semantic and graphical way. We adapted them to personalized gestures and present in this paper the new concept of Customizable Gesture Menus.

The last concept we illustrate in the paper is the necessity of helping users during the definition of their gestures. In fact, as the user is able to personalize his gestures, confusing gestures can be introduced by the user during the initialization phase. Thus we introduce an automatic mechanism to help the user during the definition of his gestures to avoid confusing gestures. This mechanism is based on a conflict detection system.

In this paper, we first analyze the state-of-the-art of help on learning gestures for users. Then the self-evolving recognition system is described briefly. From the third section, we present three different experiments. Each experiment justifies one objective of our research. The first experimentation explains why we chose to leave the user the freedom to personalize his gestures. The second one compares the help of learning in the form of menus and table. And in the end, the last one evaluates the new concept of help on gesture definition before the conclusion.

2 Context and State-of-the-Art of This Study

2.1 How to Help Users Learn Gestures?

Main approaches of gestures' learning help are based on Marking Menus [2] which propose two ways of utilization: the *novice mode* where the user has menus displayed to help him finalize his gesture and the *expert mode* where he only needs to draw the gesture needed and a recognizer will try to understand which command is invoked. First enhancement of the Marking Menus is their hierarchical version [1]. They allow multiple levels of menus so that commands can be organized by family and more commands' gestures can be presented.

New improvements have taken off afterwards. Zone and polygon menus proposed compound multi strokes to execute one command [3,4]. These hierarchical menus are cut between levels. Each stroke is a simple straight line. The display of these menus does not take a lot of space since each time only one level appears. For the same reason, the accuracy is also good so that more gestures can be included. But the final gesture of one command cannot be drawn in one stroke, which makes it slower.

All these approaches help users to memorize gestures by making them practice drawing. Evidently, the final form of gestures depends strongly on the menus' ergonomics. For user's familiarization and learning, this is a big constraint.

Another interesting example is Octopocus [5]. In fact, Octopocus went a step further: it permits natural fluid gestures. The interaction on novice mode is also interesting. Firstly it uses a colorful aspect to help the learning [13]. Secondly, in novice mode, it gave a continuous feedback. It highlights the currently recognized command, and weakens gradually the others. This approach helps users' learning and it has a good accuracy. The gestures in this case were predefined too.

2.2 Evolving Classification Systems

With the increasing use of touch sensitive screens, there is a growing demand for handwritten gesture recognition systems. Two strategies exist for building a handwritten gesture recognition system: using a writer independent training set or a writer dependent one.

The first require a lot of different writers to gather a heterogeneous training database to build a system independent of the user. This approach has limited performances, in particular if in the end user writes his gestures differently from the training users, his gestures won't be recognized well. The second strategy only uses data samples from the final user to train the system. This allows better performance, since the system is designed for a specific user, but requires him to spend a lot of time writing training gestures.

More recently, new approaches have been proposed to put up with those drawbacks. Paper [15] suggested turning a writer independent system into a writer dependent one. Doing so limits the need for end users to train the system and at the same time improves recognition performance. However, recognized classes are pre-specified, and end users can't add new classes. Paper [16] proposed a very simple template matching classifier (the 1$ classifier) with very few training requirements. It enables end users to easily build a classifier and allow them to add new gestures. On the other hand, such a simple system has limited performance compared to common statistical classifiers such as Support Vector Machines or Multi Layer Perception.

To go further in that direction, we use an evolving classification system that can start from very few training samples and learn incrementally during its use [9]. Interfacing of the gesture command application with the evolving recognition engine is illustrated in Figure 1.

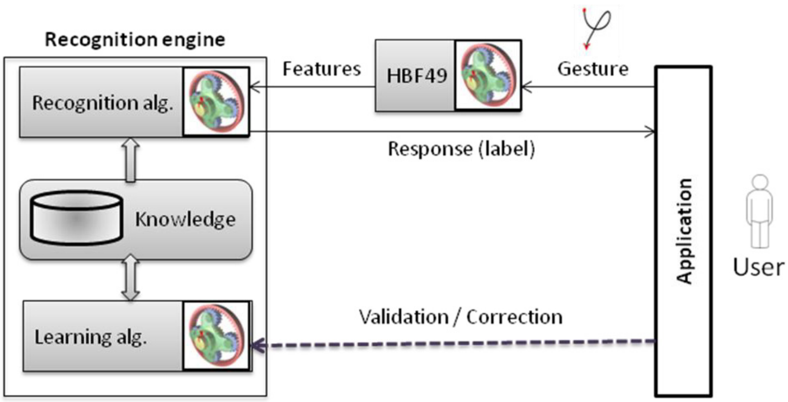


Fig. 1. Interfacing of the application with the evolving recognition engine

Evolving classification systems have appeared in the last decade to meet the need for recognizers that work in changing environments. They use incremental learning to adapt to the data flow and cope with class adding (or removal) at runtime. This work uses such an evolving recognizer, namely *Evolve* [10], to recognize gesture commands. *Evolve* is based on a fuzzy inference system that learn incrementally and cope with class adding. As *Evolve* is a statistical classifier, and not a matching classifier, it needs to extract features from gesture drawings, as illustrated in Figure 1. In our case of gesture commands, class templates are chosen by end users, which make it difficult to design a feature set. Because of that, we chose to use the Heterogeneous Baseline Feature set (HBF49) [6] which had been optimized on several well-known handwritten gesture databases to be the most general possible. Besides, using such an existing and available feature set enables easy classifier comparisons.

3 First Experimentation: Personalization?

3.1 Background

After studying many variations of Marking Menus, we chose one approach to analyze how it works: Continuous Marking Menus (CMM) [7]. This approach is based on Hierarchical Marking Menus, so gestures are organized by families and have multiple levels. It allows cursive mono-stroke gesture, and it offers a continuous feedback on novice mode, to guide users by darkening the chosen gesture and lightening others. The table of some gestures is presented in Figure 2. We proved that user learning improved with the help of these menus [19]. But after an opinion poll, surprisingly we received some negative feedback from testers. We discovered that while CMM allow continuous cursive gestures, the final form of gestures still depends strongly on the menus' ergonomic. For users, these gestures look simple, but it is difficult to link them to specific commands. The most important is that they did not like the idea that we forced them to use pre-defined gestures. That is how we came to think about the personalization of gesture commands on pen-based devices.

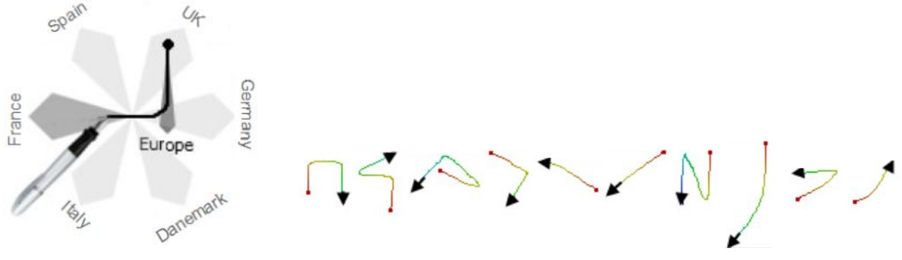


Fig. 2. CMM (on the left) and their generated gestures (on the right)

3.2 Experimental Protocol

Our hypothesis is that users learn better their own personalized gestures than predefined ones. We want to evaluate this hypothesis in the context of a real application. For that purpose, we made a picture editor in which users can manipulate pictures with some basic commands, like: “copy”, “select”, “zoom”, etc. There are twenty-one commands in total for this application, which are separated into seven different families. For gesture recognition in this picture editor, we use the *Evolve* recognizer with incremental learning presented in section 2.2, so we can also test the behavior of this recognizer in a real application.

Users were divided into two groups. Group 1 needed to choose by themselves twenty-one gestures for all commands, while group 2 needed to manipulate the application with twenty-one predefined gestures as shown in Figure 3. The experimentation was divided into five phases as follow.

- Initialization phase. Group 1 defined all gestures and practiced them twice. On the other hand, we showed a table of predefined gestures to group 2, and they repeated them 3 times. Both groups draw each gesture three times and this gave three samples for the classifier to learn.
- First learning phase. It was made of thirty-four questions, each of them corresponding to a command. Moreover, some commands were asked more often than others. This was to illustrate the utilization of real applications. Six of them were asked three times, four were asked twice and eight were asked once. The remaining three commands were not asked at all. During the learning phases, users had access to a help window with a table of all the gestures [Figure 3].
- First test phase. During the test phase the user would be asked to achieve once each of the twenty-one gestures in a random order. The help window would not be displayed if the user could not remember the gesture.
- Second learning phase. Same as the first learning phase.
- Second test phase. Same as the first test phase.

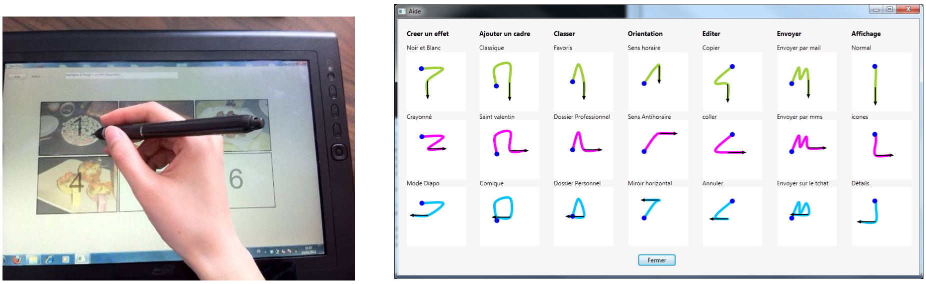


Fig. 3. View of application (on the left) and Table of predefined gestures (on the right). During learning phases, the help window was presented in the same way as this table.

When a gesture was made, it was sent to the recognizer. If the gesture was correctly recognized, the visual effect of the command applied so the user could be informed that he had successfully performed his gesture, and then another question would be asked. On the other hand, if the answer of the recognizer was different from the gesture that was asked in the question, we must know if the recognizer failed or if the user performed a wrong gesture. So, in this situation, a window would pop up to show the user the expected gesture and the gesture he drew. In this window, the user had to state whether or not he had performed the right gesture. If he drew the correct gesture, the visual effect applied and the test went on. However if he drew a wrong gesture, he would then have to try again. When a gesture was correctly recognized, the classifier used it to strengthen its learning for this class.

3.3 Results

The parameter we compared between two groups here is the capability of user & recognition system cross-learning.

Thirty persons participated in the experiment, with fifteen persons in group 1 and fifteen persons in group 2. The mean learning curves are presented in the left part of Figure 4. We can see in the graph that Group 1 got a better score of memorization during first test phase (>90%) than group 2 (41%). This advantage is kept for the second test phase where we got always more than 90% for group 1 and only 60% for group 2. We can see that group 2 made a big improvement from test 1 to test 2, but it could still not overtake the memorization rate of group 1. The statistics test on our results proved the same. Group 2 made significantly more progress on learning rate than group 1. However group 1 always made significantly better score than group 2 (p-value < 0.001). This validates our initial hypothesis that personalized gestures are easier to be remembered and accepted by users.

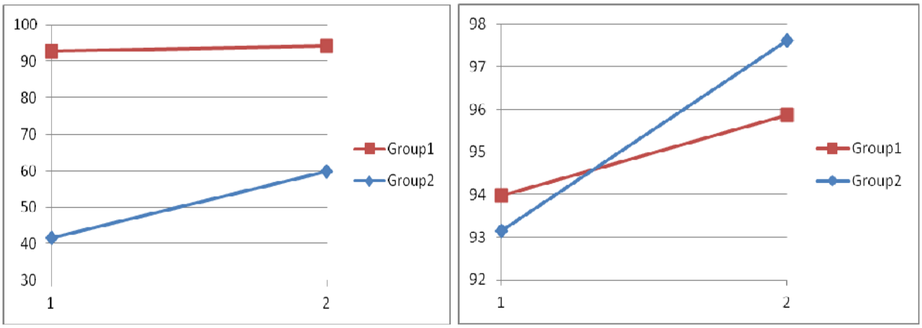


Fig. 4. Mean learning curves of users (on the left) and of recognition system (on the right) for the two groups. The abscissa presents the two test phases. The ordinate presents the learning rate.

As mentioned above, we would also like to see the learning of the recognition system (cf. right part of Figure 4). We noticed that both curves increased. Group 1 got a better score at the first test point (94%) compared to 93% of group 2. But group 2 took over group 1 at the second test point with an increase of nearly 5% while group 1 made an increase of 2%. This does not mean that the recognizer worked better on group 2's scenario. After a test of significance, we can only say that the recognizer increased significantly for both groups, but there is no significant difference between the improvement in group 2 and the one in group 1 (Sign. > 0.05). We proved that the recognizer really learnt incrementally during the experimentation so it obtained a better score in the end. We then concluded that by looking at the learning rates of both users and recognizer (cross-learning), the personalization of gestures permits a better memorization of gestures.

4 Second Experimentation: Help on Gesture Learning

We were not totally satisfied with this result. From a design point of view, the help list in the form of an additional window (a table of gestures) is neither practical nor visually appealing. On a tablet device, it takes the place of nearly the whole interface and hides the information of the application main window. Compared to this, Marking Menus are much more advantageous because they take less space and do not hide main information. Some of them can show dynamically commands that we are interested in [5]. So an adaptation of Marking Menus on personalized gestures is necessary for improvement: we designed *Customizable Gesture Menus* [18] which are Marking Menus adapted for personalized gestures.

4.1 Customizable Gesture Menus

Since the gestures are personalized, the construction of menus is much more complex. It means the construction should be dynamic, and should guarantee a good presentation of all gestures at the same time. These gestures are presented in a circle (similar to *Wave Menus*' form [8]. Strokes are smoothed by *Canonical Spline* to give a clearer appearance to gestures [11]. The place where each gesture is put on the circle is not random. It depends on the initial direction of user's gesture. For example, a gesture with an initial direction from left to right will be placed on the right part of the circle. The exact position is decided by the initial angle.

Once the application detects that user's stylus is moving very slowly or not moving at all, he will be considered as novice. And the menus will be displayed. Moreover, the menus are colorful. According to research of [13], a colorful aspect helps the learning. The center of the circle is an inactive (empty) zone. Without raising the stylus, user can go through all icons of gestures and the corresponding command labels will be displayed, as well as the gesture at bigger size. We decided to show a label just near the circle, unlike other menus which show the label at the end of gesture [2,3,5]. It makes it easier for the user to see the right command as he does not need to search too far to know if it is the right one, especially in case of long gestures. Once the user finds the command that he is looking for, he just needs to follow the gesture shape (his drawing can be approximate), and he will see a continuous feedback in the menu, where the stroke spline fills with white color during his movements. The rest of menus lighten little by little. As he finishes the stroke, the command will be validated and executed (cf. Figure 5).

An advantage of hierarchical marking menus is that they organize commands by family, so final gestures of commands share a semantic meaning among them. As mentioned before, *Customizable Gesture Menus* present commands one by one, so the commands from the same family may not be found side by side. For example, command "next page" and command "previous page" belong to family "page manipulation". And we associate a line from left to right and a line from right to left as gesture for these two commands. Since they do not have the same initial direction, they will not be placed side by side. So when user is looking his gesture for "next page", he will neither think about, nor learn at the same time, his gesture for "previous page". Since our menus are one-level menus, our solution to conserve the semantic meaning of gestures is when user's stylus moves on one command; the other commands of the family are highlighted too. In this way, even they are not side-by-side; they are still displayed at the same time (cf. Figure 5).

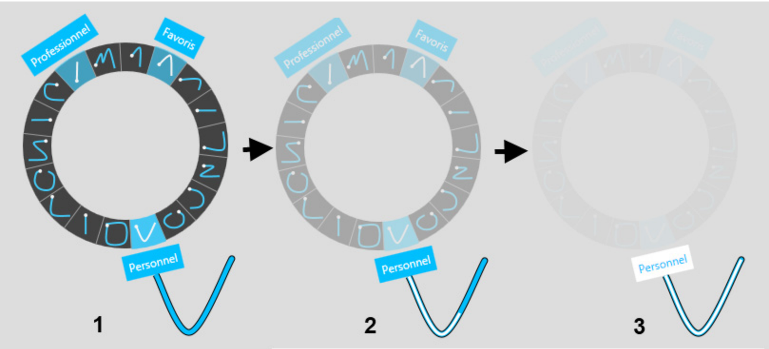


Fig. 5. Selection of gesture on Customizable Gesture Menus

4.2 Experimentation to evaluate Customizable Gesture Menus

We wanted to test if menus really bring an advantage compared with the help displayed as a table of all gestures. We made two groups, each using a different form of help, either the menus or table. Group 1 defined all their gestures and the help was presented by a simple table of gestures (cf. Figure 6). Group 2 defined also all their gestures but they were helped with *Customizable Gesture Menus* (cf. Figure 7).

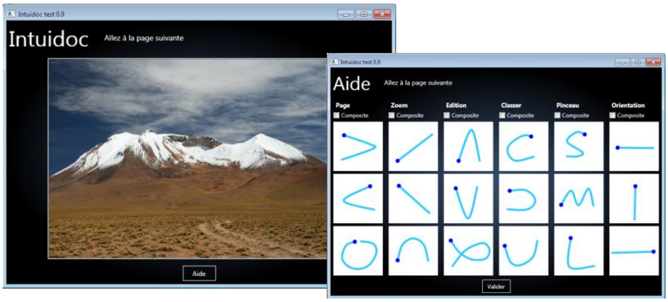


Fig. 6. View of the application during the learning phase with the table of gestures layered. The table is displayed directly in front of test interface when user asks for help.

Fig. 7. View during the learning phase with Customizable Gesture Menus

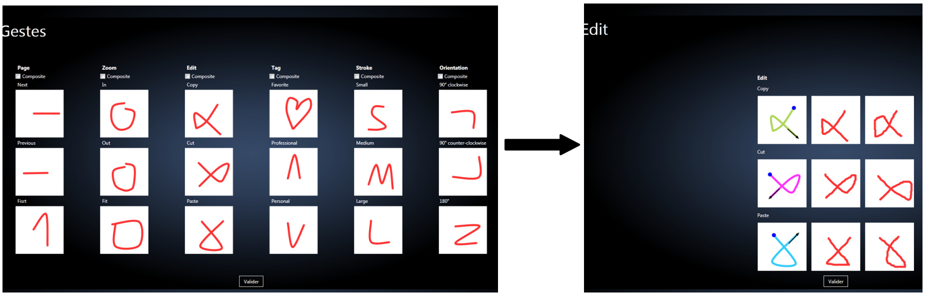


Fig. 8. Screen shots of initialization phase. Left part presents the interface when user defined all gestures. Right part presents the interface when user repeated twice each gesture for the initial learning of the recognizer.

We used eighteen commands (separated into six families). During learning phase, in each family, there was one command which was asked three times; one command which was asked once and another command which was not asked at all. This is to simulate the utilization of a real application where some commands are used more often than others.

The process of tests is the same for the two tests. Tests are constructed of four phases:

- Initialization phase. Users could just fill each case of eighteen commands with a stroke to define gestures. Then for each group, users needed to repeat their gestures once so that the recognizer got another sample to be initialized from (cf. Figure 8).
- First evaluation phase. The user would not have access to help (neither “Gesture menu” nor “table of gestures”). But once he made an error, a pop-up message would be displayed. It showed the user the expected gesture and the recognized gesture, and the user had to tell the system if it was his fault (he draw a wrong gesture) or the recognizer’s fault (he drew the right gesture but the recognizer did not recognize it). So in the first evaluation, user and recognizer could still continue to learn. Compared to the previous experimentation, we did not put a first learning phase before the first evaluation phase so that we know if the user is able to

memorize all gestures right after their definition. Because if all users were able to learn by heart all gestures since the beginning, neither menus nor table of gestures would be necessary during the utilization of application to help users. And the whole tests comparatives would make no sense.

- Learning phase. This phase includes several sequences of questions. Each question concerns one command. If the user knew the gesture asked in the question, he could draw it directly (cf. Figure 9). Otherwise, user of Group1 and Group2 could access help from a classic table of all gestures or with our *Customizable Gesture Menus* respectively (cf. Figure 7 and Figure 8). The user might draw unconstrained gestures, and the recognizer is not able to recognize them. In this case, constructive feedback is necessary to help user understand that he made an error [14]. In this phase, in case of error, the user would have the table of gestures or Customizable Gesture Menus respectively displayed in a pop-up. We showed him the recognized gesture at the same time. It was up to the user to tell us if it was him who drew a wrong gesture or if it was the recognizer who gave the wrong answer. If it was user's fault, the same question would be asked again for the user to practice. If it was the recognizer's fault, the misunderstood gesture would be used to improve the recognizer by incremental learning. And the user moved on to the next question. So in case of error, both recognizer and user could still learn via the pop-up windows. Then the user would move on to the next question.
- Second evaluation phase. The same as the first one, but no pop-up would be displayed in case of error because the learning for user and recognizer had stopped already.

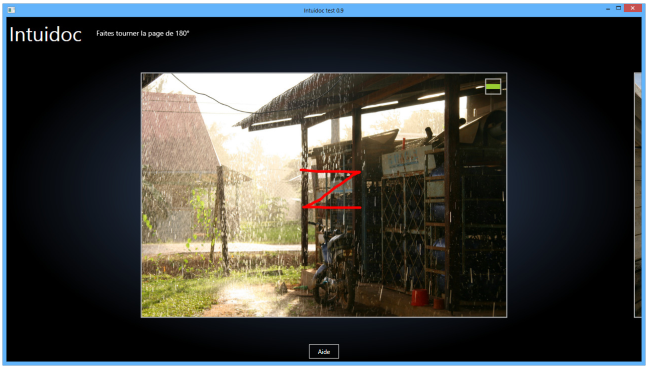


Fig. 9. Application interface during learning phase. Here is an example where user knew the gesture asked in question, so he just draw it directly without asking any help.

4.3 Results

For this experimentation, fifty-nine persons participated in the tests. They are from twenty-two-year-old to forty-two-year-old. For group 1, there were nineteen persons. For group 2, there were thirty nine persons. They are all familiar with computers.

The parameter we compared between the two groups is also the capability of user & system cross-learning.

To analyze users learning, we made learning curves by collecting the results of the two evaluation phases (cf. Figure 10 left part). Few users (13%) succeeded to have 100% of learning rate from the beginning (first evaluation phase). This proves that even if gestures are all personalized, few people are able to learn them all by heart directly. The learning curves of the users for each group are presented. First we can observe that the gradient of learning is quite the same for the two groups (6.4% of progress for group1 and 7% for group2). According to statistical tests, there is no significant difference between learning rates of group 1 and group 2 for either results collected from test phase 1 or results from test phase 2. Both group 1 and group 2 made a significant progress of learning (results from test phase 2 are better than results from test phase 1). But there is no significant difference between these two progresses (Sign. > 0.05). This experiment shows that *Customizable Gesture Menus* does not disturb user's learning of gestures in comparison with classical tabular presentation.

To analyze the recognition system learning (cf. Figure 10 right part), we can see that there is an improvement between the two evaluation phases of the recognition rate of the evolving recognition engine for both groups (6.9% of progress for group1 and 5.7% for group2) . According to statistical tests, the recognizer engine succeeded to learn better and better significantly for both groups, but there is no significant difference between the recognizer improvements for each of the two groups (Sign. > 0.05).

Customizable Gesture Menus allow a more compact presentation of all gestures that is a key point to deal with the relatively small size of tablet screen. This experiment demonstrates that both user and recognition engine can improve, at the same time, their capabilities of learning respectively: how to draw the gestures for the users and how to recognize the gestures for the recognition engine.

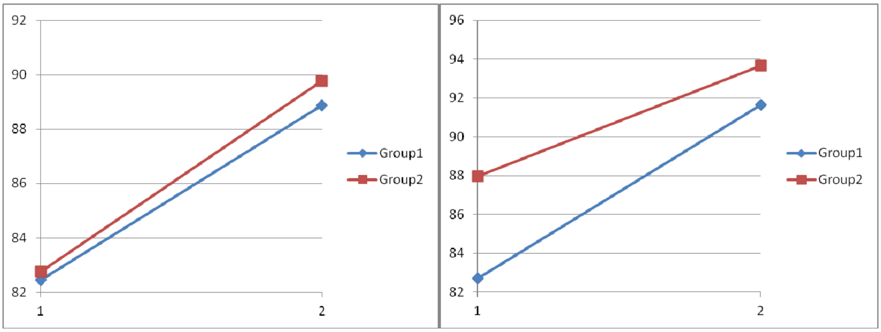


Fig. 10. On the left users' mean learning curves. On the right recognizer's mean learning curves. The abscissa presents the two test phases. The ordinate presents the learning rates.

5 Third Experimentation: Help on Gesture Definition

In the two previous experiments, we have presented a real cooperative interaction between the user and the evolving recognition engine, that are both able to learn from each other to improve their capabilities of respectively, reproducing and recognizing gestures, in the context of real applications. But by taking a precise look on obtained gestures from the second experiment, we noticed that even when we asked users to define a unique gesture for each command, there were still persons who defined too similar gestures that induce confusions. In this case, the *Evolve* recognizer can have strong difficulties to discriminate these too similar classes of gestures. Moreover we can assume that it could be also more difficult for the user to memorize these similar gestures. So we decided to design an automatic mechanism which detects potential conflicts between gestures defined by the user.

To introduce this third experiment, we firstly tested the conflict detection system on the second experiment to see how many users had defined confusing gestures. Then we added this conflict detection algorithm on the initialization phase of our application and made a third experimentation.

5.1 Conflict Detection with Recognition System

A gesture is said to be confusing when the classifier gives it high probabilities of belonging to multiple classes. In other words, a gesture is not confusing when the recognizer gives it a high probability of belonging to a single class, and low probabilities of belonging to other classes. In our application, we state gesture classes as confusing when the third gesture sample (of the initialization gestures) of this class is found confusing by the classifier after learning the two first gesture samples.

A fuzzy inference system, like *Evolve,* is composed of several fuzzy rules that are each associated with a cluster of the input space. If every rule participates in the recognition process of every class, each cluster is mainly associated with a single class. Any gesture that is between two or more cluster, and any clusters that are too close to each other, should be signaled as confusing. Such a confusing gesture will activate several rules to similar levels and will likely be given similar probabilities of belonging to multiple classes.

We use this property to compute a confidence degree as the difference between the activations of the most activated rule ("$rule_{first}$") and the second most activated rule ("$rule_{second}$") normalized by the activation of the most activated rule.

$$confidence = \frac{activation(rule_{first}) - activation(rule_{second})}{activation(rule_{first})} \tag{1}$$

This confidence degree varies between 0, when two rules are equally activates, and 1 when a single rule is activated. It allows us to flag gestures as confusing when confidence is below a defined confidence threshold (0.09 in our case). The choice of the threshold is explained in our paper [17].

When conflicts are detected, we suggest a change of gestures for the corresponding class. If the user doesn't want to change his gestures, then more gesture samples are asked to try to reduce the conflict between classes.

5.2 Utilization of Conflict Detection on Previous Experiment Data

As mentioned above, our first objective is to find out how many users draw potentially confusing gestures in the last experiment. To stay homogeneous, we took only users from group 2 where users were helped by *Customizable Gesture Menus*.

We have developed an application which present all gestures defined by users one by one. Figure 11 shows one specific user of group 2. The histograms present the confidence degree of each gesture. On the top of each histogram, there is a sample of each gesture. By fixing the threshold at 0.05, we detect 2 confusing gestures, and we can tell that apparently they look very similar, so the conflict detection algorithm works well.

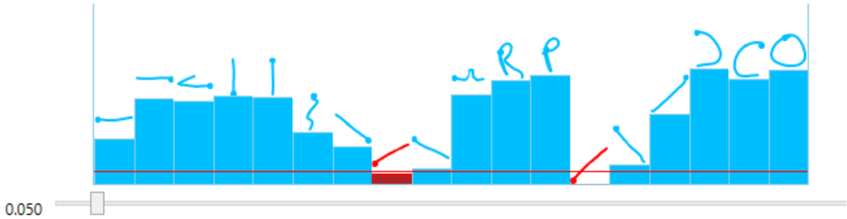


Fig. 11. Screen shot of the application showing all 18 gestures of one specific user. The confusing gestures are marked in red while others are in blue since they are all different.

We made this test on all forty users, and then separated them into two categories. In the green category, no user made confusing gestures for the recognizer. In the red category, every user made at least two potentially confusing gestures. Finally we got nine persons in green category and thirty-one persons in red category. This big difference proved that people really need help to avoid making confusing gestures during the definition of their gestures.

5.3 Experimentation 3 with Conflict Detection during Gesture Definition

Our second objective by using the conflict detection is to add it on the initialization phase. Our hypothesis is to see if it really helps on the gestures' definition and if it can bring any advantage on cross-learning results.

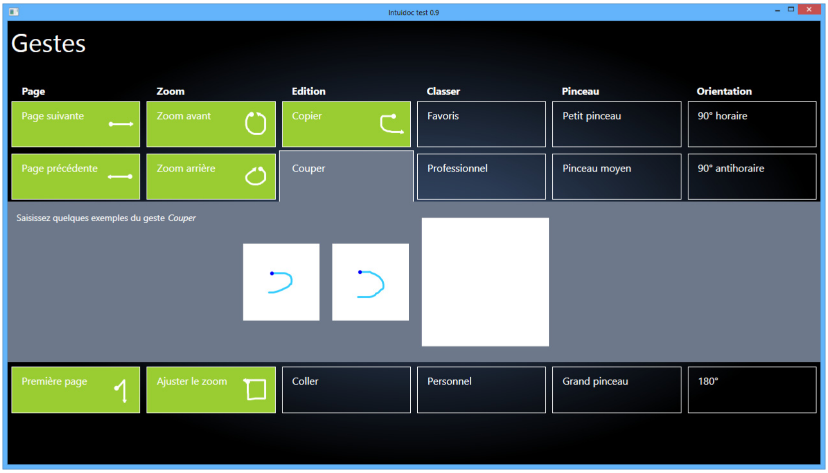


Fig. 12. New interface of initialization phase

The third experimentation is nearly the same as the second one. It followed also the proceedings: initialization phase - test phase 1 - learning phase - test phase 2. We just made some changes on the design of initialization phase. Instead of separating the definition of gestures and practice of gestures for recognition system, we combined them together (cf. Figure 12). By clicking on the label of each command, we reached the definition box of this command. We needed to give three samples of gesture and one sample would be shown by the side of command's label as a preview. In this way,

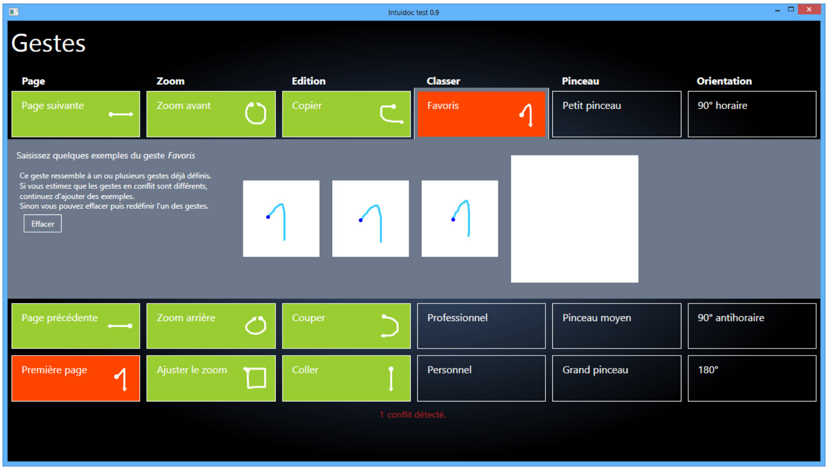


Fig. 13. Initialization phase with one detected conflict. In this example, we have the same gesture for the command "first page" (column 1, line 3) and the command "file in favorite folder" (column 4, line 1). These commands are marked in red while other commands already defined are in green because they are different from one another. The user can choose either to enter a fourth sample to one of two confusing commands or to change completely all samples of one command.

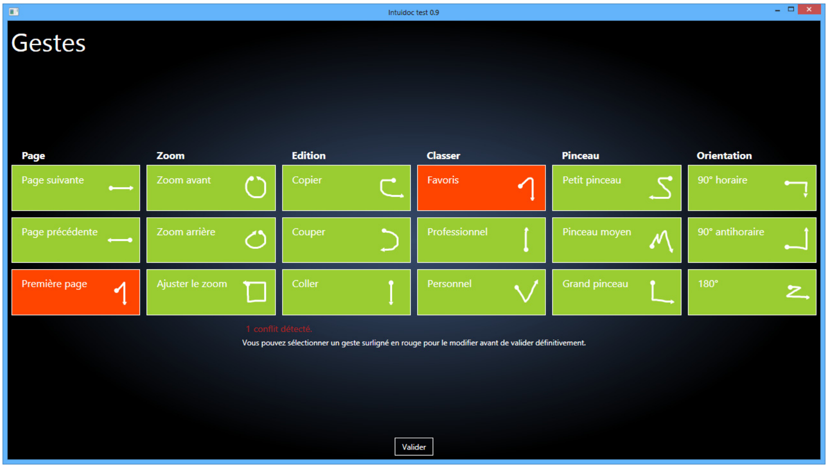


Fig. 14. Conflict can also be detected in the end once we click on "validate". If according to him, these gestures are different between them, he can click again on “validate” to force the learning of gestures by the recognition system.

the user can always have a general idea of gestures that he has already defined, and he can change gestures easily. When there was no conflict detected, all gestures were shown in green. This detection of conflict was checked each time three samples of one command were submitted.

When a conflict was detected, the two concerned gestures would be shown in red as shown in Figure 13. Moreover, a final check would be done once all gestures of commands were defined (cf. Figure 14). If there was any conflict, the user just needed to enter the definition box by clicking on command’s label to change the gesture. If from his opinion, the highlighted red gestures were not confusing at all, he could choose to add one sample or more to one command, or to force the recognizer to learn as it was by clicking on “validate”.

5.4 Results

Sixteen users participated in this new experiment. They are around twenty-two-year-old. They are all familiar with computers. In this experiment, users were alerted if they made similar gestures thanks to the conflict detection algorithm. There were always users who decided to force the learning of gestures for recognition system even there was still conflict (cf. Figure 15). But it was up to them to decide so.

We also obtained the learning rates of user (memorization rates) and recognition system (recognition rates). Users from this new experiment are called as Group 2. We compared them to results obtained by group 2 in the second experiment which is named as Group1 here. Group 1 is separated into Group1+ (users who made less than three confusing gestures) and Group1- (users who made at least 3 confusing gestures). We than got twenty seven persons in Group1+ and twelve users in Group1-. We can see that in average score (cf. Figure 16 left part), Group1+ and Group2 are both better

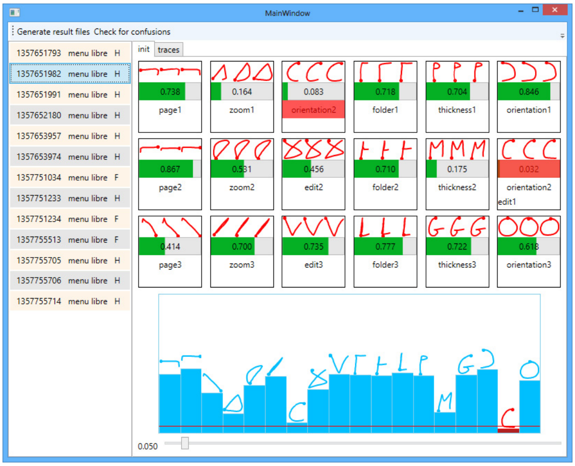


Fig. 15. Conflict detected for 2 gestures

than Group1-, either for test phase 1 or for test phase 2. We obtained similar results for mean recognizer learning rates too. But it is not enough to say that our hypothesis is true. As we see in the right part of Figure 16, we have a lot of variation within each group. So we need to make a signification test to justify our hypothesis.

The results of the pairwise tests to compare Group1+ to Group1– and Group2 to Group1– are shown in Table 1 and Table 2. We can see that Group1+ is significantly better than Group1- for almost all measure, while Group2 is significantly better than Group1- too. We did not obtain any significant difference between Group2 and Group1+. This is as we expected since Group2 was helped by the conflict detection system and it should be as good as Group1+ which made few confusing gestures. From these statistical tests, we can conclude that both Group2 and Group1+ obtained better performance than Group1- in term of cross-learning of the user and the recognizer.

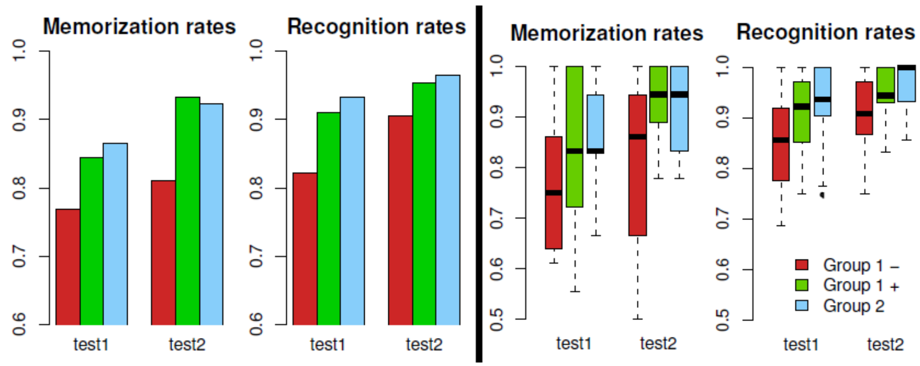


Fig. 16. At left, bar plots of users' mean memorization rates and mean recognition rates at two test phases of all groups. At right, Box plots of memorization rates, of recognition rates for all groups.

Table 1. Group1+ compared to Group1-

Rates	Statistical formulas	Significance
memo-test1	Nothing	Nothing
memo-test2	$\chi^2 = 5.0743$, df = 1, p-value = 0.02428	Group1+ > Group1- Significant
reco-test1	F(1, 13.133)=3.6463, p-value=0.07	Group1+ > Group1- Hardly Sign.
reco-test2	$\chi^2 = 3.8429$, df = 1, p-value = 0.04996	Group1+ > Group1- Significant

Table 2. Group2 compared to Group1-

Rates	Statistical formulas	Significance
memo-test1	$\chi^2 = 4.0617$, df = 1, p-value = 0.04387	Group2 > Group1- Significant
memo-test2	$\chi^2 = 3.2602$, df = 1, p-value = 0.07098	Group2 > Group1- Hardly Sign.
reco-test1	F(1,15.644)=5.2093, p-value = 0.0368	Group2 > Group1- Significant
reco-test2	$\chi^2 = 4.8478$, df = 1, p-value = 0.02768	Group2 > Group1- Significant

6 Conclusion

In this paper, we analyzed "user & system cross-learning" of gesture commands through several new concepts. We reported three experiments, summarized in Table 3, to support our approach for designing pen-based interfaces for complex application needing more than ten gestural commands: the first concept is to offer users the possibility to personalize their gestures; the second concept is to use *Customizable Gesture Menus* to help users memorize gestures; the third concept is to help users during the definition of

Table 3. Overview of 3 experiments

Exp	Goal	Help on gesture learning	Help on gesture definition	Result
1	Determine if it is important to leave user the freedom to personalize their gestures	Table	None	Personalized gestures are better than pre-defined gestures
2	Determine which form of help suits better user and tabletops to memorize gestures Evaluate the cross-learning of user & recognition engine	Table *vs* Customizable Gesture Menus	None	Menus are equal to table to help users learn gestures Both (the user and the recognition engine) can improve, in the same time, their capabilities of learning
3	Avoid confusing gestures	Customizable Gesture Menus	Conflict detection algorithm	With the help of the conflict detection system on gestures' definition, the user and the recognition engine learn better

their gestures prevent too similar gesture definition. On this last concept, more tests need to be done to consolidate the efficiency of the conflict detection algorithm (because all rates are not significantly better). Moreover the reported results show that using an evolving gesture recognition engine that learns incrementally, starting from few data samples, is a really promising strategy to induce a cross-learning of user and recognition engine. We summarize that personalized gestures can offer an easier way to interact with software on pen-based devices. The handover of software offering the gesture's personalization should be separated into two steps: the help on definition of gestures and the help on memorization during the utilization.

Besides these new concepts, we know that Marking Menus can offer more advantages on interactive devices compared to traditional way with either menu bar or tabular menu, like entering text or parameters [12, 13]. Integration of these extension concepts will be our future work.

References

1. Kurtenbach, G., Buxton, W.: The limits of expert performance using hierarchic marking menus. In: Proc. INTERACT 1993, pp. 482–487. ACM Press (1993)
2. Kurtenbach, G., Moran, T.P., Buxton, W.: Contextual animation of gestural commands. Computer Graphics Forum 13(5), 305–314 (1994)
3. Zhao, S., Agrawala, M., Hinckley, K.: Zone and Polygon Menus: Using Relative Position to Increase the Breadth of Multi-stroke Marking Menus. In: Proc. CHI 2006, pp. 1077–1086. ACM Press (2006)
4. Zhao, S., Balakrishnan, R.: Simple vs. Compound Mark Hierarchical Marking Menus. In: Proc. UIST 2004, pp. 33–42. ACM Press (2004)
5. Bau, O., Mackay, W.E.: OctoPocus: a dynamic guide for learning gesture-based command sets. In: Proc. UIST 2008, pp. 37–46. ACM Press (2008)
6. Delaye, A., Anquetil, E.: Hbf49 feature set: A first unified baseline for online symbol recognition. Pattern Recognition 46(1), 117–130 (2013)
7. Delaye, A., Sekkal, R., Anquetil, E.: Continuous Marking Menus for learning Cursive Pen-based Gestures. In: Proc. IUI 2011, pp. 319–322. ACM Press (2011)
8. Bailly, G., Lecolinet, E., Nigay, L.: Wave Menus: Improving the Novice Mode of Hierarchical Marking Menus. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 475–488. Springer, Heidelberg (2007)
9. Almaksour, A., Anquetil, E., Quiniou, S., Cheriet, M.: Evolving Fuzzy Classifier: Application to Incremental Learning of Handwritten Gesture recognition Systems. In: Proc. ICPR 2010 (2010)
10. Almaksour, A., Anquetil, E.: Improving premise structure in evolving Takagi-Sugeno neuro-fuzzy classifiers. Evolving Systems 2, 25–33 (2011), doi:10.1007/s12530-011-9027-0
11. Petzold, C.: Canonical Splines in WPF and Silverlight, `http://www.charlespetzold.com/blog/2009/01/Canonical-Splines-in-WPF-and-Silverlight.html`
12. Guimbretiére, F., Winograd, T.: FlowMenu: combining command, text, and data entry. In: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, San Diego, California, United States, November 06-08, pp. 213–216 (2000), doi:10.1145/354401.354778

13. Igarashi, T., Matsuoka, S., Kawachiya, S., Tanaka, H.: Interactive beautification: a technique for rapid geometric design. In: Proc. UIST 1997, pp. 105–114. ACM Press (1997)
14. Norman, D.A.: Natural user interfaces are not natural. ACM Magazine Interactions 17(3), 6–10 (2010)
15. LaViola, J., Zeleznik, R.: A Practical Approach to Writer-Dependent Symbol Recognition Using a Writer-Independent Recognizer. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(11), 1917–1926 (2007)
16. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In: Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2007), Newport, Rhode Island, October 7-10, pp. 159–168. ACM Press, New York (2007)
17. Bouillon, M., Li, P., Anquetil, E., Richard, G.: Using Confusion Reject to Improve User and System Cross-Learning of Gesture Commands. In: Proc. ICDAR 2013 (to be appeared, 2013)
18. Li, P., Anquetil, E.: Graphical Gesture Commands' learning with the help of Customizable Gesture Menus. In: Proc. IGS 2013 (to be appeared, 2013)
19. Li, P., Delaye, A., Anquetil, E.: Evaluation of Continuous Marking Menus for Learning Cursive Pen-based Commands. In: Proc. IGS 2011, pp. 217–220 (2011)

Adoption and Appropriation: A Design Process from HCI Research at a Brazilian Neurological Hospital

Junia Anacleto[1] and Sidney Fels[2]

[1] Department of Computer Science, Federal University of Sao Carlos, Sao Carlos, Brazil
[2] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
junia@dc.ufscar.br, ssfels@ece.ubc.ca

Abstract. Through our research on natural ICT solutions for integration into a non-ICT based workflow at a Brazilian chronic care hospital, we created a new design process and two additional HCI design criteria for maintaining natural work processes using information and communication technologies (ICT). For our HCI design we propose two design pathways: 1. iterating on *adoption* of designed technologies and 2. iterating on *appropriation* of these technologies. The degree of appropriation provides an indicator of how natural a design is since it allows for users' inventiveness to uncover latent affordances for use in new contexts. Thus, the use of an interface along with whether its potential is realized in new, user-oriented contexts, are critical elements for designing natural interfaces. We report our insights gained through observations and user-centered design for health professionals at a non-ICT based, large chronic-care hospital to support this perspective.

Keywords: human computer interaction, appropriation, adoption, design process, user centered design, health care, mental illness, non-disruptive workflow.

1 Introduction

We have partnered with the Brazilian Hospital CAIS Clemente Ferreira, that is a special chronic care hospital for individuals with neurological and brain disorders. A primary objective of the Hospital is for professional health staff to facilitate patients' transitioning from the hospital back into normal society. The hospital has 3 floors with 6 wings with 800 patients and 600 professionals distributed throughout. The hospital has four distinct professional roles: administration, health professionals, nurse assistants and maintenance. Currently the Hospital has essentially no Wi-Fi or cell phone coverage due to architectural issues (at least 70 cm thick concrete walls) and budget (no funds to install Wi-Fi service). They use voice and a paper-based system for their primary workflow mechanisms. Our primary research project is targeted towards the health professionals and not the patients. We believe that working with these professionals is an opportunity to design information technology based workflow that avoids a WIMP oriented strategy. Likewise, we can use their

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 356–363, 2013.

workflow as a starting point in deciding what *natural* is, given that their workflow has evolved over more than two decades of practice. Thus, we argue that introducing unnatural technologies into this workflow should be obvious if we see a substantial disruption.

We began our design for the hospital looking to gesture and speech based technologies as they promise to enable computers to understand people's naturally evolved and learned methods of interactions. Further, in some environments where the existing Windows, Icons, Menus and Pointers (WIMP) approach doesn't match many of the workflow activities, alternative approaches that look towards gesture and speech interaction may make more sense. In particular, in our hospital environment, most of the health care workers are nomadic. Thus, designing to free the professionals from workstation/WIMP based solutions appears to be a logical fit. However, during our user-centred design process to construct these WIMP alternative solutions, we identified two main challenges: 1. There is not a design process to follow for maintaining natural workflow and 2. It is unclear what natural *is* and how to measure it. This paper addresses both of these issues through our research on user interface design with health care professionals in a hospital in Brazil. We look for *natural ICT solutions*, independently of the adopted interaction paradigm. We begin with related work followed by our modified user-centred design process model in relation to some well-established WIMP based approaches. Next, we describe the research, leading to a discussion and analysis supporting our proposed design process. We provide conclusions and future work at the end.

2 Related Work

Recently, many studies, such as [2,3,11], have been carried out for finding ICT, and WIMP based solutions especially, within hospital settings with many different objectives, like: improving communication processes among hospital staff, providing more accurate diagnostic tools and treatments, assist in therapy processes, increase patient's medication adherence, and others. These efforts are categorized in the e-Health concept encompassing various initiatives to use these technologies to improve health processes. In general, e-Health has been defined as the use of internet and other technologies in the health sector to improve access, efficiency, effectiveness and quality of clinical processes used by health organizations, patients and consumers in an attempt to improve the health of patients [12]. Thus, while we are considering ICT *solutions* for health care workers, we also seek to enable hospital staff to *evolve* their practices to determine how best to adopt some of the success that may apply to them.

3 Natural ICT Design Process

Our proposed design process, shown in figure 1, includes both *adoption* and *appropriation* activities. We consider *adoption* as the process of users learning and accepting a designed solution for some of their activities. In contrast, we look as appropriation as the use of an artifact designed for one type problem, to solve a separate one. We arrived at this process by looking to define *natural interface design* to have two properties: 1. the introduction of technology should not interrupt existing workflow and 2. that the adopted technology must also be shown to be appropriated for use in other contexts. To accommodate these properties, in our HCI design process, we added an additional phase to the usual design-prototype-evaluate design process [10]. From our observations designing for the Hospital, we determined that it is necessary to support users to first pass through the adoption phase so that they perceiving the affordances of a developed solution or prototype. Then, include affordances and mechanisms for supporting the appropriation phase where they can envision the use of such a solution in a new context.

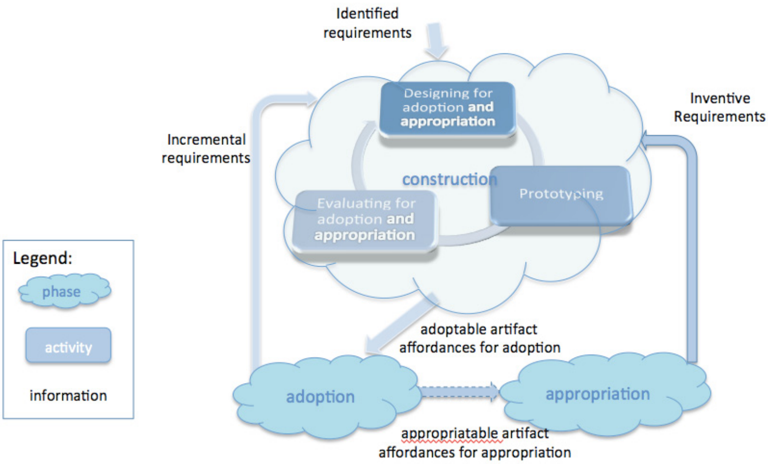


Fig. 1. Natural ICT Solution Design Process

Our development process starts with a *Construction* phase triggered by some initial requirements. We envision this phase including established HCI user-centered development (UCD) processes such as covered in [10]. Traditionally, the criteria for successful design in the construction phase for WIMP interfaces focuses on usability and usefulness to improve the chances that an artifact will be adopted by the target user community for a specified set of tasks. For a hospital context, researchers [5,6,7,9], have used a 3-phase design-prototype-evaluate approach. For example, we employed participatory design, ethnography, work place shadowing and other UCD design strategies during our work. Each iteration within the Construction phase leads to a release of an ICT artifact. Once the artifact is created, the next step is deployment and adoption by the target user group. This could be either a small test group or the expected target population. Once adopted, new requirements arise from the users to improve the system to solve the problem that the system was designed for. These normally come as incremental requirements to adjust the interaction design to allow for a gradual refinement to a well-designed system to solve user's problems that they can adopt easily. The systematic adoption evaluation activity in the construction phase attempts to discover these requirements to increase the effectiveness of adoption. In UCD, this process continues from the early stages of construction to the complete lifecycle of the product. Our observations suggest that while the process, up to this point, leads to establishing new methods to iteratively improve adoption of solutions to a particular problem, it does not include any method to facilitate appropriation. As described in section 4, in our research, this appropriation is critical as it is part of how the Hospital professionals have evolved their work practices. Thus, we add the appropriation phase as part of the development cycle for maintaining natural workflows. During the adoption phase, a user (or possibly users) becomes comfortable enough with the technology that he or she may appropriate the design for a new context to solve problems unforeseen during the Construction phase.

Once appropriation occurs, the user creates new requirements based on the new context of use and the appropriated solution. These are inventive requirements. The appropriation evaluation in the Construction phase can then establish these new, inventive requirements to feed into a new design. The design may be an entirely new product tailored to solve the problems in the new context in a new way. We identify some of the appropriation design approaches that facilitate this process in section 5. Alternatively, the designers may adjust the next version of the product to expand the contexts that it may be used in.

In summary, the Natural ICT Design process leads to designs that both support adoption and appropriation. For our work, we are applying our approach to gesture-based, mobile solutions with the additional constraint that it should impose minimal disruption in the existing workflow to be natural. However, we suggest that the Natural ICT Design approach is also appropriate for ICT solutions where the aim is to support the natural evolution of work practices. As we discuss in the next section, our approach progress as we designing solutions in the Hospital.

4 Our Design Process in the Hospital

We have used been using UCD design process as described in the previous section that currently has been ongoing from June 15, 2011 to now. During this time, we have had 9 half-day or full day meetings with healthcare staff, documented and validated their workflow, established 4 scenarios, and created 3 prototypes for testing, including a long term study that is ongoing. As well, we applied a survey and a one week workplace shadowing activity.

Our process began with a survey and initial discussions with the Director of the Hospital to initialize the Construction phase. Within the Construction phase we used both Participatory design (PD) and task-centred design for design methods as well as low and medium prototyping and various evaluation methods. A group of six trained health professionals (Figure 2) with extensive experience were part of the PD team: 1 physiotherapist (15+ years experience, undergraduate degree), 2 nurses (20+ years of experience, nursing degree), 1 occupational therapist (15+ years experience, undergraduate degree), 1 social assistant. Two professionals never used e-mail or any social network and never bought something via internet; two use iPods and demonstrated abilities with ICT; all professionals had contact with touch technology, mobile phones and web search; none of them had contact with free-form gesture based technology. Our choice of PD was intended to avoid our own bias towards the use of WIMP interaction for some of the scenarios that were identified. We continued with PD-prototyping-evaluation until we arrived at five scenarios leading to three medium-fidelity prototypes.

Fig. 2. PD activity with health professionals

The first prototype involved placing public displays in common workrooms that allowed the Hospital Director to put a message to call emergency meetings. The prototype worked by having the Director's assistant make a phone call to the rooms when a meeting is called (the current practice) to let people know there is a meeting; however, for rooms with our prototype, the person who answered the phone pressed a large red button displayed on the screen (using a mouse pinned on the wall) that displayed a large message that a meeting has been called that would stay on screen for 30m. We deployed this in one wing and used the current verbal relay method in another wing as a control. The second prototype had two variations due to the changing nature of the physiotherapist's scenario. It used a Kinect™ to sense body posture and provided limb, head and neck angles for use in the prototype systems. This device was used for prototyping both physiotherapists' scenarios: the game distraction system and the hand's free goniometer capture system. The third prototype was a simple ticketing system that used the public displays. Health care professionals could generate a request for help by typing in a message that would appear on the screen. Then, someone seeing that message, could help out and then flag the request as attended to. We had two different public visualizations of the usage of the system to see whether competitive or cooperative representations would help with enhancing the adhoc helping mechanisms that were identified as key elements of the success of the functioning of the hospital.

One important observation made during the design session with the physiotherapist (Feb 23, 2012) is that she disclosed that she had recently used the camera on her cell phone to take photos of one of her patients recovering from an injury. She was planning to some of the photos at a conference. However, she decided to print the photos and put them in a photo album documenting healing progression. One day, she showed the album to her colleagues that stimulated

Fig. 3. Photos using personal cellphone to document physiotherapy results

discussion and how they could use this approach for their patients (Figure 3). As discussed below, this is an example of appropriation of the technology, suggesting that the camera interface itself has natural properties.

5 Analyses and Discussion

Through the 21 months of design activities, we observed situations that led us to argue for two main points about designing natural ICT solutions: 1. the process must include design for appropriation and 2. it should minimize disruption to the existing workflow. To meet these criteria, we argue that the notion of a natural interface provides both Norman's sense of affordances [4] *and* Gibson's affordances [1]. As pointed out in [8], Norman affordances provide visual cues as to the function of a particular interface that match users common sense knowledge of use. This helps the process of adoption and also leads to less learning required. In contrast, Gibson affordances are the latent functions that an object can support. The typical example provided is that a chair has clear visual affordances that it can be used to sit on; a Norman affordance. However, there are other functions, such as standing on, that may not be visually obvious, but that the chair can support. These would be Gibson affordances as the chair *affords* making the user taller. Typical, WIMP interfaces focus on Norman affordances for their functions but tend not to afford other functions making it difficult for users to appropriate to new contexts.

In our design process, including Norman affordances supports the design for adoption and leads to easier to use interfaces. Though, if designing ICT solutions, the designer still needs to be concerned with how much disruption the artifact has with the current workflow. For the design for appropriation, however, Gibson affordances are important. As a user encounters new contexts in their activities, there is a natural inclination to appropriate objects designed for one purpose for use in the new one. The ability to do this depends upon the object having the potential to be used in the new context and the user to perceive this. By designing Gibson affordances into the artifact so that objects have the potential to be used for different purposes, this natural evolution is supported. Or conversely, part of the measure of naturalness, is whether the designed objects are appropriated for use in contexts other than what the designed Norman affordances are for. This insight leads to the two design pathways for natural ICT design, as shown in Figure 1.

Designing for appropriation remains a challenge. Dix's basic studies on appropriation [13], suggests that designers should accept that we do not understand completely what will happen in real use with ICT solutions. He argues this is especially true within the context of non ICT-based workflow. Nonetheless, based on our Natural Design process, it is necessary that the ICT solutions are designed so that they can be used in unexpected ways. Further, they need to address the natural environment dynamics so that designing for use is designing for change. To accommodate this, we suggest, the construction phase of design should embody the principles for appropriation listed in Table 1. We refine Dix's principles for natural ICT solutions as that is the direction we are pursuing at the Hospital, however we look to future work to determine if they generalize. Our studies and preliminary observations at the Hospital reinforce that appropriation is related to users being comfortable enough with technology to use it in their own ways. We encountered: one successful, one partially successful and one failed appropriation. Specifically, the one success was when the physiotherapist used her camera phone to begin documenting her patient's healing process, then printing the timeline and bringing it to share with her colleagues at the Hospital. This use led to others in the same wing trying the same idea (Figure 3). The camera interface on her cell phone was simple: start the camera application, point and shoot. However, the printing and organizing required significant time and effort to do. The physiotherapist had mastered use of this camera for her personal activities. What we observed is that she made the connection between the capabilities of accumulating a chronology of photos to watch the healing process and then made the leap to see this would be useful for the care of this patient. Our second prototype used a Kinect™ during physiotherapy as we determined that the physiotherapist identified a problem when she takes care of two children at the same time. After 3 iterations of design, a Kinect-based application was designed, approved by the therapist, and delivered. However, once she tried the actual first functional prototype, with the all the limitations of the Kinect™, she lost interest in the approach and changed her mind about the whole scenario. We believe, after discussion with her, continuing another scenario, and analyzing video, that there is a substantial gap between what she thought the technology would do and what it would actually do. We concluded that this design supports adoption but it did not lead to appropriation, as she was able to use it but did not see anything beyond the limits of the technology. Our design for a meeting calling scenario was neither successful for adoption nor appropriation. The deployed functional prototype was successfully executed with messages appearing on a public display, but the public displays were completely ignored. Hence, the construction cycle lead to technologies supported by the users during PD but ultimately, were rejected.

Table 1. Principles for Designing for Appropriation (derived from [13])

Principle	Description
Allowing interpretation	Including elements where users can add their own meanings
Providing visibility	It is often the *irrelevant* state and *internal process* that can be appropriated.
Exposing intentions	Deliberately exposing the intention behind the system.
	Appropriations may subvert the rules of the system
Supporting, not controlling	Designing a system so that the task can be done instead of a system to do the task.
	Providing the necessary functions so that the user can achieve the task (instead of driving the user through the steps).
Respecting plugability and configuration	Designing systems where the user can plug the parts together in different ways.
Encouraging sharing	Designing systems that support sharing appropriations
Learning from appropriation	Observing that a temporary use of a tool has become specialized.
	Observing that a crystalized appropriation leads to a new tool.

6 Conclusions

Through our partnership with the Hospital we have access to a large number of highly trained health professionals who mainly use a verbal and paper based system for managing and care-taking of chronic mental patients. They have evolved their work practices over more than two decades providing an exceptionally rich environment for us to study how they have created natural workflows without technology. As we progressed through our design processes for improving their work practices, we reoriented our thinking to realize that to be able to design ICT solutions for them, we first needed to design technologies that had no impact on their natural work practice. Then, we could begin to work with them to establish how to appropriate the technologies into improving their practices. However, without their inventiveness in the process, the solutions ended up alienating them from being able to improve their care giving potential. Further, we suspect that without this process, trying to introduce technologies from the existing research to allow them to use accepted best practices will generally fail. Thus, we developed an additional element in the design process that encapsulated the natural ability of the health professional to use objects around them, once they had adopted them. We believe this is a critical component in developing natural user interface solutions for health care workers in this context. We look to future work to see how effective it is and whether it generalizes to other contexts.

Acknowledgements. We thank FAPESP, CAPES-DFAIT, Boeing. CAIS Clemente Ferreira Ethics.

References

1. Gibson, J.J.: The Ecological Approach to Visual Perception, NJ, USA, L. E. Ass (1979)
2. Vuong, C., Ingalls, T., Abbas, J. Transforming clinical rehabilitation into interactive multimedia. In: ACM MM, pp. 1–2, (2011).
3. Geurts, L., et al.: Digital games for physical therapy: fulfilling the need for calibration and adaptation. In: TEI 2011, pp. 117–124 (2011)
4. Norman, D.: Affordances, Conventions, and Design. Interactions 6(3), 38–41 (1999)
5. Bath, P.A.: Health informatics: current issues and challenges. J. Inf. Sci. 34-4, 501–518 (2008)
6. Li, J., Wilson, L., Stapleton, S., Cregan, P.: Design of an advanced telemedicine system for emergency care. In: OZCHI 2006, pp. 413–416 (2006)
7. Reeder, B., Turner, A.: Scenario-based design: A method for connecting information system design with public health ops and emerg mgmt. J. Bio. Inform. 44(6), 978–988 (2011)
8. McGrenere, J., Ho, W.: Affordances: Clarifying and Evolving a Concept. In: GI 2000, pp. 179–186 (2000)
9. Norman, D.A.: Natural User Interfaces are Not Natural. Interactions 17(3), 6–10 (2010)
10. Preece, J., Rogers, Y., Sharp, H.: Interaction Design, 1st edn. John Wiley & Sons, Inc., New York (2002)
11. Schönauer, C., et al.: Chronic pain rehabilitation with a serious game using multimodal input. In: Virtual Rehabilitation (ICVR), pp. 1–8 (2011)
12. Broderick, M., Smaltz, D.: HIMSS E-Health White Paper: e-Health Defined (2008), `http://www.himss.org/content/files/ehealth_whitepaper.pdf`
13. Dix, A.: Designing for appropriation. BCS HCI (2), 27–30 (2007)

Chronicles: Supporting Conversational Narrative in Alternative and Augmentative Communication

Annalu Waller[1], Rachel Menzies[1], Daniel Herron[1], Suzanne Prior[1], Rolf Black[1], and Thilo Kroll[2]

[1] School of Computing
[2] School of Nursing and Midwifery,
University of Dundee, Perth Road,
Dundee, DD1 4HN, Scotland, United Kingdom
{awaller,rachelmenzies,danielherron,sprior,
rolfblack}@computing.dundee.ac.uk, tkroll@dundee.ac.uk

Abstract. Individuals share experiences and build relationships through the medium of narrative. Lifelong personal narratives play a key role in developing social identity. Individuals with little or no functional speech due to severe speech and physical impairments (SSPI) find it difficult to share personal narrative as augmentative and alternative communication (AAC) systems do not support interactive story telling. As a result, people with congenital SSPI who use AAC may not have learned the linguistic skills involved in sharing narratives. The Chronicles software was developed to support the sharing of personal narrative. Conversational analysis of a conversation using Chronicles illustrates how the system can support more natural conversations when using AAC.

Keywords: Augmentative and Alternative Communication, Personal narrative, Social identity, Accessibility, Assistive technology, Disability.

1 Introduction

People with little or no functional speech due to severe speech and physical impairments (SSPI) can benefit from augmentative and alternative communication (AAC) [1]. Although voice output communication aids (VOCAs) provide independent access to spoken output, people who use VOCAs communicate at rates of around one tenth of natural speech and find great difficulty in engaging in conversation [1]. Current VOCAs are well suited to supporting the expression of needs and wants (such as “I am thirsty”), but more complex interactions such as conversational narrative (e.g., speaking about a holiday or childhood experience) are not well supported.

The *Chronicles* project worked with adults with SSPI, their support staff, families and friends to harness existing research technology to support them in sharing their personal narratives. The *Chronicles* software was developed through ethnographic observations, interviews and rapid prototyping methodologies [2]. Researchers were

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 364–371, 2013.

embedded in a local adult care facility during the design, development and pilot testing of the software. Evaluation of *Chronicles* with an adult male with SSPI illustrates how the system can support users in initiating and sustaining interactive narrative.

2 Background and Previous Research

Narrative has been demonstrated to shape and reflect identity and sense of personal continuity [3]. However, for many individuals who use AAC, the ability to participate in personal narrative remains severely restricted. Instead, aided conversation is characterised by one or two word responses with little initiation or elaboration [3, 4]. Most VOCAs allow for the storage and retrieval of longer units of text. However, these "stories" tend to be delivered as monologues with little opportunity for spontaneous adaptation to the natural flow of conversation. Relating personal narrative within conversation is complex and involves the retrieval, sequencing and evaluation of conversational utterances to convey past experience [3, 4].

Research AAC systems have attempted to support more natural conversation flow by storing smaller conversational units so that the order of retrieval can vary. The *Talk* system [5] demonstrated a retrieval system through which the user is supported to navigate through a conversation. *Talk:About*™ [4] included ways to introduce a story, tell it at the pace required (with diversions) and give feedback to comments from listeners; but again this tool was based on a library of fixed texts. Natural language generation (NLG) has been used in more recent AAC projects. Dempster [6] developed a system which uses information either entered by the user or retrieved from sources such as online radio stations to create narratives. The *How was School today...?* (*HwSt*) project [7] uses sensor data to provide information about a disabled child's day at school. Time stamped data related to location (e.g. 11.05am: Hall), people (e.g. 11.12am: Mrs Sound) and objects (e.g. 11.15am: Tambourine) were collected using RFID[1] sensors. Using additional information, e.g. the child's timetable, this data is transformed into narrative utterances using data-to-text technology, e.g. "In the morning, I was in the hall", "Mrs Sound was there", "I played the tambourine". The child is able produce an evaluation that relates to the previous utterance by selecting a positive (happy face) or negative (sad face) button after an utterance, e.g. in relation to the last two utterances: "She is nice." or "I didn't enjoy it.". *HwSt* provided a symbolic interface to support pre-literate individuals as many AAC users are in danger of not acquiring literacy [1].

3 The Chronicles Project

The vision was to develop a system which would guide both literate and pre-literate users in sharing lifelong narratives in an interactive conversation. The technical challenges for the proposed project were: i) to develop strategies to capture lifelong

[1] RFID: Radio Frequency ID.

narratives for storage in the system; ii) to provide an effective retrieval system so that users can retrieve appropriate narratives; iii) to display the retrieved narrative utterances in a manner that makes sense to the user; and iv) to support automatic evaluation as in *HwSt* ("Jane was there" ⇨ "I like her"). An ethnographic study was conducted in a day centre for adults with physical and learning disabilities. This resulted in the establishment of story telling groups to support clients to develop narrative skills [8]. The design phase involved the co-design of an interactive story telling device with four nonspeaking participants [2].

4 The Chronicles System

The *Chronicles* system is designed to allow users to retrieve parts of a narrative non-sequentially to facilitate an interactive conversation instead of delivering a monologue. In addition, users can elaborate by selecting buttons to expand on the characters and places within the story and add emotional evaluations to utterances by choosing the appropriate buttons. Chronicles is implemented within the Tobii Communicator 4[2], an AAC software platform, and was installed on a DynaVox Vmax[3], a rugged touchscreen tablet running Windows XP.

The interface (Fig. 1) follows a logical flow, starting with a story selection bar at the bottom of the screen, moving up into the utterances that make up that story in the middle of the screen, and at the top to the embellishments that can be made, allowing the user to elaborate on the story. Potential users of the Chronicles software might not have functional literacy; images are used in conjunction with a small amount of text to help users identify and understand each story, utterance, and embellishment.

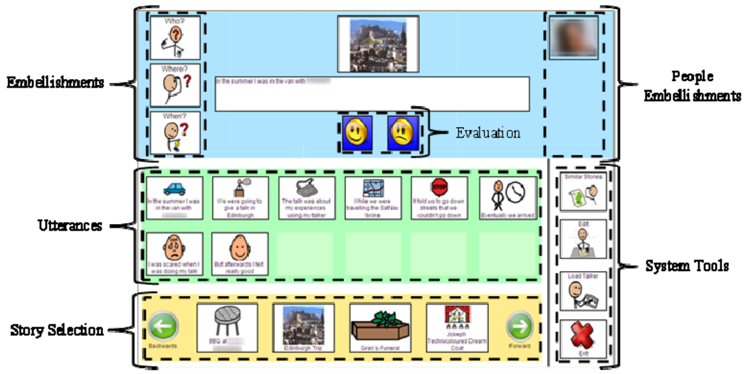


Fig. 1. The final interface

[2] http://www.tobii.com/en/assistive-technology/global/products/software/tobii-communicator/
[3] http://dynavoxtech.com/products/

The interface is divided into six sections:

1. Story Selection: Users can navigate through all of the stories they have stored in the system. When a story is selected, the *Chronicles* software retrieves all its utterances and populates the Utterance section of the screen with this data.
2. Utterances: The utterances are the parts that make up a story. Each utterance can be selected independently of the others, allowing a change of narrated order and repetition during conversation.
3. Embellishment Buttons: Allow the user to embellish on the story, providing them with the opportunity to talk about information that surrounds the narrative, such as where and when the story took place, but is not crucial to the story itself.
4. People Embellishment Buttons: If a person is mentioned in an utterance, a button with their picture links to utterances about that person.
5. Evaluation Buttons: The happy face button generates a positive comment and the sad face button generates a negative comment about the last utterance spoken. The evaluations are contextualised to the preceding utterance; if the utterance was "I had chicken for dinner", the positive evaluation button will generate a positive food experience evaluation, for example, "It tasted delicious!"
6. System Tools: The control buttons provide a variety of functions including selecting a similar story, editing the current story, switching to their existing AAC software (disabled during the evaluation period) and exiting the *Chronicles* system.

5 Evaluating Chronicles

Conversations with and without the Chronicles system were analysed to determine how the structure of narratives differs from those in using a traditional VOCA.

5.1 Participant

Gordon[4] (31 years old) has cerebral palsy and uses a motorized wheelchair with a joystick. Gordon has no intelligible speech but uses some vocalisation and idiosyncratic gestures. He uses a Pathfinder[5] device with Minspeak LLL 128; this system has a symbol keyboard interface – words are stored under symbol sequences. He uses direct selection to access keyboard sequences; this is labour intensive due to Gordon's physical difficulties. Retrieved words are spoken through a speech synthesiser. Gordon is a competent communicator, initiating conversations with communication partners. He has some difficulty shifting between topics and maintaining a prolonged conversation. He relies on producing single words or short sentences using his VOCA.

[4] All names of participants are changed.

[5] `https://store.prentrom.com/`

5.2 Methodology

The evaluation phase was primarily qualitative, with the main objective of assessing the performance and usability of the system in a care setting. Researchers spent an hour a week over a six week period in a local adult care setting, working with the participant. Narratives were elicited by encouraging the participant to talk about his past. This involved the researcher working with the participant and care worker using his existing VOCA (see [8] for details). The researcher then entered the data into *Chronicles*. Gordon was then trained to use the *Chronicles* system and was given opportunities to practice any skills learned with both existing and novel partners. The participant was encouraged to use the system independently, but received support from a care worker when required. During each training session, the researcher set up the device to the *Chronicles* home screen and Gordon selected a story to tell or a topic for discussion. At the end of the evaluation phase, Gordon was asked to speak to two unfamiliar partners. All sessions were video recorded and subsequently transcribed using the following notation:

Talk	– Natural speech	{ button }	– Interface button selected
[Talk	– Overlapping speech	((turns))	– Non-verbal interaction
<u>Talk</u>	– Whole word / phrase computer generated speech	???	– Inaudible speech

5.3 Equipment

During the training sessions with the *Chronicles* system, the participant had access to his own VOCA. This allowed him access to his full vocabulary for narrative elicitation purposes. During the evaluation sessions, the participant had access to either his VOCA or the *Chronicles* system, to facilitate a comparison of the systems.

5.4 Results and Analysis

Transcripts of conversations between Gordon and familiar and unfamiliar partners were analysed to identify whether there was any difference in the way in which conversational narratives were structured and how this impacted on the conversational flow. Using Labov's [9] definition of narrative, narratives were identified in the transcripts as a sequence of at least two clauses, which are temporally ordered, where any change in their sequence results in a change in the sequence of narrative events. Narrative sequences were then analysed in terms of having an abstract or summary, an orientation, evaluations, narrative clauses, a result and a coda (to mark the end of a narrative). Narratives also involve a planned disruption (i.e. what makes the event of interest), orientation to the listener (i.e. not a monologue), serialization of discourse markers (e.g. "and then…"), and context cues (providing background information). It was important to analyse whether a narrative emerged during the conversation as an orally realised discourse unit [9]. Figures 2 and 3 show extracts of conversations, first using the participant's existing VOCA (Pathfinder) and then Chronicles.

Narratives stored within Chronicles were elicited during conversations with the participant using his VOCA (Pathfinder) with help from a care worker.

Conversations using the Pathfinder are characterised by the retrieval of individual words, e.g., ***No yesterday card David Hannah came***. Each word is retrieved and then spoken by pressing a maximum of three keys sequentially. These sequences have been memorised by Gordon over time. Having retrieved the individual words, Gordon selects the "speak display" key to say the sentence as a whole (Fig. 2, line 1). Although Gordon is interacting with Kate, an unfamiliar partner, Gordon wanted Carol, his care worker, to participate. It soon became clear that Carol's role was to act as "interpreter" as she worked with Gordon to understand that he wished to tell Kate that his cousin David (who he regards as a brother) forgot his birthday and didn't even send a card. Much of the 14 lines of transcript are working out what Gordon is trying to communicate. However, Gordon does succeed in giving the vital details of his story. Carol tries to elaborate the story for Gordon in line 12, giving him the opportunity to evaluate his story nonverbally in line 13. The rest of the conversation (not shown) followed a similar pattern to reveal that Hannah is David's wife. Having elicited the story from Gordon, it is not stored for retelling. In future, if Gordon wishes to retell his story, he will have to undergo the same process unless Carol and Kate are with him in which case he might defer to them to tell it for him. If Carol thinks the story is important, she might store all the parts of the story in the notebook facility of the device, but this would be retrieved as a monologue.

1	G:	***No yesterday card David Hannah came. No yesterday card David Hannah came***.
2	C:	((shakes head at GM)) Did he phone? Did he? Ooooooh.
3	G:	((laughs))
4	K:	Somebody forget your birthday?
5	C:	Who's David, G?
6	G:	***Cousin and brother. Cousin and brother.*** ((looks at C and gestures))
7	C:	((nods)) Mmmhmmm.
8	K:	Your brother forgot your birthday?
9	G:	***Cousin and brother.*** ((looks at C and gestures))
10	C:	Cousin's brother? Cousin and brother?
11	G:	***I say brother.***
12	C:	He's like a brother, to you, and he forgot your birthday? You know what you do? This is what I would do G. I would send him a birthday card saying happy birthday - wait a minute, it was my birthday!
13	G:	[((vocalises))]
14	C:	Happy birthday to me! And then just send him that. See if he gets the…

Fig. 2. Transcript 1 – Gordon speaking to Carol (familiar) and Kate (unfamiliar) using Gordon's existing VOCA (Pathfinder with Minspeak LLL 128)

In comparison, the conversation with Richard in figure 3 follows a very different pattern. Unlike transcript 1, the conversation is not primarily concerned with unravelling what the aided conversation partner is trying to say. Instead, the conversation flow is more natural. Gordon spends less energy retrieving words. In transcript 1, he accesses 12 individual words and speaks three full utterances, necessitating 15 separate retrieval sequences (each one requiring at most three key sequences). Using *Chronicles*, he accesses 5 full utterances consisting 60 words using 5 retrieval button selections. The utterance in line 5 (figure 3) prompts a question and answer sequence in which Richard uncovers more detail.

The structure of the narrative reflects some of the characteristics of a typical conversational narrative. It begins with an abstract (line 1), has an orientation (at Christmas, when I was six), narrative clauses (several utterances), a result (line 18) and a coda (line 27). Although Gordon did not use any evaluation buttons, Richard provides access to emotional evaluation in lines 17 and 27, similar to Carol and Kate in transcript 1 (lines 12 and 14). Both narratives involve a planned disruption (e.g. missing the birthday or wearing casts). Gordon responds to the listener in both conversations, e.g., line 3 of transcript 2, and some discourse markers are found in transcript 2 (e.g. "as well" in line 18). Much of transcript 1 attempts to elicit context cues (who is David?) while Gordon is able to provide background information in transcript 2.

1	G:	{Utterance} ***My best Christmas present was a television when I was six.***
2	R:	Well, I have not heard that before, start again because I missed the beginning.
3	G:	{Utterance} ***My best Christmas present was a television when I was six.***
4	R:	Oh, wow, that was a while ago. A television when you were six. Did you like watching some programmes?
5	G:	{Utterance} ***It was to cheer me up because my legs were in plaster a lot when I was younger.***
6	R:	[Oh, right. Did you get lots of operations?
7	G:	((Nods))
8	R:	I couldn't imagine so. Are your hips.
9	G:	((pointing to feet))
10	R:	And your feet as well.
11	G:	((Attempts to press a button))
12	R:	Now I worked with kids before I came here { ??? }and they had a lot of problems with their hips too, yes.
13	G:	((Attempts to press a button))
14	R:	Do you like some specific programmes when you watch the telly then?
15	G:	Aye.
16	G:	{Utterance} ***I could not get to sleep when they were on.***
17	R:	Ah, yes. With the plasters. I know we had some. Really awkward, aren't they? Did you have them spread apart as well, your legs? Well It hasn't changed much that is still the case. They have not found a better method yet.
18	G:	{Utterance} ***I had to wear a green body brace as well.***
19	R:	A green one.
20	G:	Aye.
21	R:	A green one.
22	G:	[Aye.
23	R:	Was that the colour of the plaster?
24	G:	Aye.
25	R:	The whole brace, the whole body.
26	G:	Aye.
27	R:	Oh my goodness, I could not imagine the difficulty in sleeping.

Fig. 3. Transcript 2 – Gordon speaking to Richard (unfamiliar) using CHRONICLES

6 Discussion

The purpose of this paper is to illustrate the potential of a narrative based system to facilitate a different type of conversation compared to that which is produced using a conventional word retrieval system. A single case study has been used to demonstrate the usage of *Chronicles*. Using a conventional device, Gordon's conversation is stilted with a focus on ascertaining the facts of the story. Using *Chronicles*, there is a more natural flow as the nonspeaking conversation partner is more independent of his contributions. Narratives are told using both systems, but different characteristics are more apparent when using *Chronicles*. Although Gordon did not use the evaluation buttons in this transcript, he has done so in other conversations. Analysis of the two transcripts, one with and one without the *Chronicles* system, shows differences in interactions; when using *Chronicles*, the user is better able to initiate topics, sustain topics, modify the narrative content and add emotional evaluation. The system allows the user control when sharing personal narrative and demonstrates that a narrative-based AAC system can facilitate more natural conversation. The analysis of the conversations illustrates how a narrative based AAC system can enhance the flow of an aided conversation.

The results of this case study suggest several avenues for further research. The elicitation of past experience remains a challenge and work is underway to support input of experience by users and carers by developing a framework through which past experiences can be elicited, e.g., using an automated question and answer sequence to elicit the parts of a narrative in terms of an abstract, an orientation, narrative clauses, evaluations, a result and a coda. *Chronicles* is currently being ported to a tablet platform and this will facilitate further case studies.

Acknowledgments. The authors would like to thank all the participants and staff at the day centre for welcoming the researchers to share their lives for the duration of the project. This project was funded by a Digital Economy Research Grant: EP/I031901/1. Project web link – http://aac.computing.dundee.ac.uk/

References

1. Beukelman, D.R., Mirenda, P.: Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults. Paul H. Brookes Publishing Co., Baltimore (2005)
2. Prior, S., Waller, A., Kroll, T., Black, R.: Use of an Agile Bridge in the Development of Assistive Technology. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 1579–1588. ACM, New York (2013)
3. Grove, N.: Using Storytelling to Support Children and Adults with Special Needs: Transforming lives through telling tales. Routledge, Oxford (2012)
4. Waller, A.: Communication Access to Conversational Narrative. Topics in Language Disorders 26(3), 221–239 (2006)
5. Todman, J., Alm, N.: Modelling conversational pragmatics in communication aids. Journal of Pragmatics 35(4), 523–538 (2003)
6. Dempster, M., Alm, N., Reiter, E.: Automatic generation of conversational utterances and narrative for AAC: a prototype system. In: NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies, Log Angeles, USA, June 1-6, pp. 10–18 (2010)
7. Black, R., Waller, A., Turner, R., Reiter, E.: Supporting Personal Narrative for Children with Complex Communication Needs. ACM TOCHI 19(2) (2012)
8. Prior, S., Waller, A., Black, R., Kroll, T.: Investigation of narrative use in the care environment. Communication Matters, 41–42 (2012)
9. Labov, W.: Language in the Inner City. University of Pennsylvania Press, Philadelphia (1972)

Development of Novel eHealth Services for Citizen Use – Current System Engineering vs. Best Practice in HCI

Isabella Scandurra[1], Jesper Holgersson[2], Thomas Lind[1], and Gunilla Myreteg[1]

[1] Uppsala University
[2] University of Skövde, Sweden
{Isabella.Scandurra,Thomas.Lind}@it.uu.se,
Jesper.Holgersson@his.se, Gunilla.Myreteg@fek.uu.se

Abstract. Many new public eHealth Services are now being developed. Often a conventional customer-vendor process is used, where the customer is a public authority, e.g. a county council, and the vendor a commercial actor, e.g. an IT development company. In this case study the engineering process regards a novel eHealth service aiming to provide patients with online access to their electronic health record. A complicating factor in conventional customer-vendor processes for public e-services is that "the future user could be anyone". In the light of best practice in Human-Computer Interaction, this study examines the joint effort of the customer and vendor when developing novel services for citizen use. The results include delimiting factors, recommendations for public authority customers and proposed new actions for the research agenda.

Keywords: user participation, public e-health services, e-government, electronic health records, system development methods, collaborative design, Scrum.

1 Introduction

In present study, health information is brought to the citizen via online public eHealth services on home- or mobile devices. One expectation of public eHealth services in general is to make patients more involved in their own health and empowered in their relationship with caregiver organizations [1]. Consequently, many new eHealth services with the aim to provide citizens with health information are currently being developed in Western Europe by both public healthcare providers and industry. One example is to provide patients with online access to their own electronic health record (EHR). Such a service can be seen as a public e-service which is a part of the e-government imperative. To date, main goals with e-government have been to reduce costs and time needed to provide services to citizens [2]. As a result, public e-services have mainly been developed from an internal government perspective, and external user considerations have been given little attention [3]. Studies show that neither public authorities nor citizens benefit from the introduction of such public e-services [2].

It is evident that increased knowledge of external user needs is essential for successful deployment of public e-services [4] and there is a growing interest for

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 372–379, 2013.

citizen centeredness when developing public e-services in both governmental strategic documents [5] as well as in research [3]. However, it must be noted that user participation in public e-service development is challenging to put into practice. Public e-service development most often has to deal with a heterogeneous target group, i.e. all citizens [6]. Furthermore, citizens cannot be obliged to participate; their participation is voluntary, and performed besides their ordinary duties. Finally, procurement legislation and slim customer-vendor processes hamper use of user-centered requirement analyses. Consequently, and also due to traditional patterns, the process to develop public e-services is often based on conventional system engineering where future users are not involved in the development process. Public eHealth services provide no exceptions.

Present study adheres to Cooperative design [7, 8] as one of the Human-Computer Interaction (HCI) research theories that regards system development with user participation and that considers design a social process [9]. From research literature we know that usability aspects should be brought in early in the development process [9, 10]. Previous research also presents several methods to engage users in the future, like role playing and future workshops [9, 11]. Other methods to bring future needs analysis into system development are iterative prototyping and scenario-based design [8]. The latter is a well-used model in Participatory Design, also known as the Scandinavian tradition, developed since the 1970's. [7]

The degree of user participation may vary. Regardless of activation degree, in cooperative design developers and practitioners/users are seen as active cooperating partners aiming to reduce uncertainty and risk in development of novel systems where no detailed conception of exactly which future needs should be supported and in which way [9, 11]. Also in User-Centered Design (UCD) [12, 13] active participation of users is preferred; there are however other methods and techniques to apply when potential end-users are not accessible, such as personas to shape future users [14], inspection methods to evaluate an interface [15] and use of representatives for real users [10, 13, 16]. In the light of best practice from the HCI domain, the purpose of this study was to investigate the actual development process of the customer and vendor when developing novel eHealth services for citizen use: How did customers from a county council and IT-developers from industry collaborate to jointly develop a public e-service for a third party, i.e. patients or citizens? To what extent were best practice-methods from HCI used in the development of this novel public eHealth service?

2 Methods and Materials

The studied project SUSTAINS[1], is a European Union (EU) financed collaboration that aims to develop and deploy different regional eHealth services on 16 sites in 11 European countries by 2014. At the start of SUSTAINS in 2012, the majority of the partners had already initiated some development and results are now emerging. The County Council of Uppsala (LUL) in Sweden is coordinator of the EU-project and

[1] SUSTAINS `http://www.sustainsproject.eu/`, retrieved 2013-05-15.

subject to this case study. The clinical eHealth services in LUL enable patients in the county to access and read their EHR, containing medical notes, drug prescriptions, medical lab results, diagnoses and referrals [1].

Development of these public e-services was performed in 2011-12 in a customer-vendor setting. LUL (the customer) was the initiator of this development and owner of the resulting eHealth services. The IT Company (the vendor) engaged in development of the eHealth services at LUL had previously been assigned to develop some of the predecessors of this current version. In November 2012 the eHealth services were made available to all, approx. 200 000 patients within LUL, following a minor launch in August limited to LUL employees (also considered patients) for testing purposes. Recently (in March 2013) a national strategy decision was made to deploy LUL eHealth services nationally; to provide all Swedish citizens with online access to their EHRs. This fact increased the interest to further study how novel eHealth services reach a heterogeneous target group, spread over an entire country, of different ages, education, health status and interests to use the services or to participate in the development process.

The Study: Research Team Evaluators and Methods

An action research project (DOME[2]) was created by 16 nationally spread researchers to perform independent studies with multidisciplinary perspectives on the SUSTAINS EU-project. The four authors are members of DOME and each with an expertise in research areas related to eHealth: health informatics & HCI (IS), e-government & cooperative design (JH), IT & deployment processes (TL), organizational change & management control (GM).

When the research project was initiated (August 2012), the SUSTAINS project had already established a customer-vendor organization and current version of eHealth services was being developed. The selected data collection method was therefore to perform semi-structured interviews with SUSTAINS members from both sides. Three roles from the customer; project manager, project owner and medical advisor, were matched to the vendor's development manager and two usability experts. Six interviews explicitly regarding the development process (table 1) were recorded and followed up by complementary questions and answers by face-to-face contact, phone as well as e-mails containing project documentation, requirements and system overviews from the vendor.

Table 1. Respondents from the customer-vendor organization of SUSTAINS

Respondents	Reference in text	Affiliation	Length of interview
Project Manager	R1	Customer	60 min
Project Owner	R2	Customer	120 min
Medical Advisor	R3	Customer	90 min
Development manager	R4	Vendor	90 min
Usability expert 1	R5	Vendor	90 min
Usability expert 2	R6	Vendor	60 min

[2] `http://www.it.uu.se/research/hci/dome/index.php?lang=1`, retrieved 2013-05-15.

The recorded material was transcribed and analyzed by all four researchers using a content analysis model where the development process of the system was described according to the Systems Development Life Cycle (SDLC) [17]. This model describes the development process as constituted of three phases: initial; with various analyses, intermediary; system is realized, and finally; system is put to practice.

3 Results: Current System Engineering vs. Best Practice in HCI

The actual process of SUSTAINS development is compared to best practice methods and key principles in HCI and displayed in table 2. The initial phase (1) of SDLC concerns preliminary analyses, systems analyses, and requirements specifications (future users are identified and the project is defined with regard to its scope and expected outcome), the intermediate phase (2) deals with systems design and systems development (the system is realized), and the final phase (3) concerns integration, testing, installation and deployment (the system is put to practice).

Table 2. Initial (1), intermediate (2) and final phase (3) of SUSTAINS and Best practice in HCI

Current system engineering process	Best practice in HCI (a selection of key principles)
(1) *A type of "Knowledge in the head"*: customer representatives had previously studied pilot projects in Denmark, Estonia and the United States, as well as experiences of a predecessor of SUSTAINS, implemented at a private practice in LUL. These experiences were not systematically documented nor reported in the form of a systems analysis or similar. Future users were generally identified as inhabitants in the county. System specification existed in terms of an EU-consortium negotiation of a list of 12 features that should create the basis for the novel eHealth system. "*Black box-approach":* There were no specified goals other than realizing the features listed by the EU-consortium. The scope was defined in terms of planned launch date, first access in summer of 2012, and a final launch in autumn of 2012. Neither users, nor developers were actively involved in the initial analyses of the novel eHealth services.	a) *Identify main target groups for the intended service*, b) *categorize future users/most frequent users* [9]. c) *Analyze thoroughly potential users* to elicit user requirements in terms of functionality and usability [7, 12]. d) *Use requirements and project goals to steer* development [9-11, 16] e) Stated goals and expected outcome are basic tools to *create formative and summative evaluations* [10, 18]. f) *Participation of real users* is preferred to decrease uncertainty and risks related to system acceptance, and a number of methods are presented in *Introduction* [7-9].

Table 2. (*Continued*)

(2) *Communicating mental models*: The vendor's development team got involved and started by designing a prototype based on the 12 features. It needed to be redesigned with respect to a user perspective.. Usability experts were called to assist; they performed a heuristic evaluation, a conceptual model and created 3 personas; an old demented woman and her relatives, a disabled child and his parents, and a woman with multiple diagnoses. Results were delivered to the customer that accepted to rebuild the prototype. *Changing requirements*: The vendor used Scrum; an agile software development framework [19]. The development manager (i.e. scrum master) took the task to turn vague specifications into practically solvable requirements of the novel system. The Scrum process used iterations of three weeks, each ending with a customer demonstration using *personas* to get feedback for improvements and acceptance of each functionality. Users were not systematically incorporated in development activities but a *focus group test day* with patient organizations was a mandatory step in the EU-project. It is not clear whether/how this day was analyzed to improve the services. The vendor managed to *simulate user participation* when the customer assumed the role of user representative using each of the personas during demos.	g) *Actively involve future users* in design activities. The degree of user participation may vary, h) from an advisory role as a part of design team [7], using e.g. *future workshops* [11] i) to a more *representative role* during testing, prototyping and similar user-centred activities [9, 10, 13, 16]. j) use *mock-ups* [9] *user scenarios and prototypes* to trigger discussion around possible future technologies. k) Incremental development or *iterative prototyping* is advocated to gradually establish understanding between users and developers. (l) When potential end-users are not accessible other UCD-methods like *inspection methods* with guidelines [15], user representatives or (m) *personas* [14] should be applied.
(3) *Practice in practice* Only a few tests were carried out by the customer. The launch in August 2012 giving access to employees in LUL was a non-systematic test as each employee was free to use the system in the way she wanted. There was no systematic way of collecting reactions or questions from the users apart from a provided e-mail address as an option to feed back experiences. This option was sparsely used by the employees and no end-user education was provided. At service launch in November 2012, the same feedback system was used.	n) Actively *involve future users in evaluations* and tests regarding GUIs as well as workflows and functionality [13] o) Active user participation in these stages increases the likelihood for successful deployment. If real users are not accessible, there are other formative and summative evaluations to perform; with or without usability experts [12, 18]. p) *End-user education* is important.

4 Discussion and Conclusion

The SUSTAINS Case: Analysis of current case has shown that HCI best practice was not followed by the customer for any of the three phases. Overall, the project applied a scarce amount of user participation. "The users' needs were known to some extent by the aid of a survey conducted in 1999 and by the experiences made in the private general practice" (R2) Instead of an inventory among patients today, the customer (particularly the project owner) tried to take on the role of a future end-user based on the reasoning that it was future end users' needs which were of importance: "An average patient of today cannot know what she needs and demands from a future eHealth service system, as her views on and ideas of health and health systems will change, along with her own behavior". Moreover, no stated project- or effect goals made development and evaluation difficult. The logic was, that "if a stated goal is invalid according to a future user it will not matter if the system can fulfill the goal or not". According to HCI best practice, that uncertainty can be decreased using different user centered methods and techniques.

The vendor was only involved in the intermediate phase of the project regarding development of the services from specifications provided by the customer. Fortunately, the customer was supported by the vendor in handling some end-user issues and the moving targets. Use of an agile development method brought the common understanding of expected results forward as the customer collaborated closely with the vendor's development team in frequent sprint demos. Further, an initial inspection evaluation led to redesign of the entire prototype as well as creation and use of three personas during development and sprint demos [19]. The usability aspects could be further refined, in many cases usability was neglected by the system's owner: "As far as I remember no one was responsible for usability aspects" (R5). "We have made no usability testing of the system" (R6) and currently "no one knows whether the eHealth services are useful or only an online service" (R5).

Recommendations to Public Authority Customers: Although research has long advocated cooperative design methodologies, current public e-service projects are still technology-driven, instead of focusing on potential user needs. Use of existing and modified methods and techniques from the field of HCI as referenced here are of great value. In order to grasp this knowledge, it may be wise to incorporate HCI experts, not only from the vendor, but also internally, to establish a proper know-how of how to involve users to gain value out of the participation. In industry, usability experts are engaged in other projects, and only when called for their knowledge will be available for e-Government projects. It is therefore important to build in usage considerations early in planning and procurement documents, as well as working towards leaving conventional customer-vendor processes in favor of cooperative and agile methods. When HCI knowledge lacks at the customer end, which is not rare in healthcare, the developer must be given a mandate to handle these issues. In current case, the customer was led into development methods and processes selected by the vendor, in order to jointly develop a public e-service. In our opinion this project was saved by applying evolutionary prototyping, a thorough inspection evaluation and use of personas, where the customer meritoriously played user representatives using the 3 personas during iterative demos. However, working with future user-methodologies in a cooperative manner increases the likelihood for successful deployment according to best practice.

Proposed New Actions for HCI Research Agenda: By this study we inform the HCI domain where practice really is standing, and we propose a call for action, to support public authorities/public care providers by leading them towards a collaborative and user-centered development environment in an action research setting.

Constructive evaluations of novel eHealth services need to be performed, not only usability tests in a user context in current project, but in all projects. HCI researchers could assist e.g. in a triple helix-constellation with public sector and industry to build e-services that meet future users' needs and possibly saving resources in the society.

Moreover, defined goals are needed. In this case e.g. how the system is intended to contribute to the patients' well-being, how or how much the system should ease the pressure on different services provided by healthcare today, or what impact the new eHealth services should have on the workload of certain clinical staff. Support from research in creating e.g. efficiency goals would benefit public e-service projects.

There is a need for educational projects to educate the customer to handle development processes with public user (citizen or patient)-centered perspectives. HCI researchers would also do well as mediators between customers, users and developers.

Neglecting future users and usability aspects is not unique for the studied project, on the contrary. There are recognized delimiting factors that e-government initiatives are associated with and that the HCI community could study and propose e.g. guidelines for. Compared to e-services developed by commercial actors, public e-services must encompass both economic and democratic values at the same time as public administrations are regulated by laws to a large extent [20]. Being a public administration also brings the responsibility to provide e-services that provide clear benefit to the citizens since there is seldom an option to select another similar public e-service; they simply do not exist [21]. At last, users of public e-services are citizens with certain constitutional rights. Public e-services cannot exclude any user groups although they are not a majority of the users, instead all future users must be considered, taking into account different disabilities, geographical limitation, and language issues [22]. These delimiting factors make user participation in public eHealth service development a challenging and complex activity and best-practice methods are needed to support public e-service development. Future work is to further examine which HCI methods, and to what extent HCI methods, can support the process of empowering patients and making them more involved in their own health.

Acknowledgments. The DOME-project is financed by VINNOVA (2012-02233), Swedish Agency for Innovation Systems. SUSTAINS is supported by the European commission (No 297206). Special thanks go to the respondents at LUL and the development company EVRY that invited us to a Triple helix environment for future work.

References

1. Atzori, W., Ünver, Ö.: SUSTAINS User Requirements Recommendations, Report D4.1v1,2013 (May 15, 2013), http://www.sustainsproject.eu/
2. Anthopoulos, L.G., Siozos, P.S., Tsoukalas, I.A.: Applying participatory design and collaboration in digital public services for discivering and redesigning e-Government services. Government Information Quarterly 24, 353–376 (2007)

3. Axelsson, K., Melin, U., Lindgren, I.: Exploring the importance of citizen participation and involvement in e-government projects - practice, incentives and organization. Transforming Government: People, Process and Policy 4(4), 299–321 (2010)
4. Melin, U., Axelsson, K., Lundsten, M.: Talking to, not about, Entrepreneurs – Experiences of Public e-service Development in a Business Start Up Case. In: Cunningham, P., Cunningham, M. (eds.) eChallanges (2008)
5. Commission of the European communities, i2010 eGovernment Action Plan: Accelerating eGovernment in Europe for the Benefit of All, C.o.t.E. communities, Brussels (2006)
6. Henriksen, Z.H.: The diffusion of e-services in Danish municipalities. In: Traunmüller, R. (ed.) EGOV 2004. LNCS, vol. 3183, pp. 164–171. Springer, Heidelberg (2004)
7. Schuler, D., Namioka, A.: Participatory Design Priciples and Practicies. Lawrence Erlbaum Associates Inc., London (1993)
8. Bardram, J.E.: Scenario-based design of cooperative systems: Redesigning an hospital information system in Denmark. Group Decision and Negotiation 9, 237–250 (2000)
9. Greenbaum, J., Kyng, M. (eds.): Design at work: Cooperative design of computer systems, pp. 3–24. Lawrence Erlbaum Associates, New Jersey (1991)
10. Constantine, L., Lockwood, L.: Software for Use: A practical guide to the Essential Models and Methods of Usage-Centered Design. Addison-Wesley (1999)
11. Jungk, R., Müllert, N.: Future Workshops: How to create desirable futures. Institute for Social Interventions, London (1987)
12. Norman, D.: Cognitive engineering. In: Norman, D., Draper, S. (eds.) User-Centered System Design: New Perspectives on Human Computer Interaction, pp. 378–382. Association for Computing Machinery, NY (1986)
13. Gulliksen, J., Göransson, B., Cajander, Å., et al.: Key principles for user-centred systems design. Behaviour and Information Technology 22(6), 397–409 (2003)
14. Mulder, S., Yaar, Z.: The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web. New Riders Press (2006)
15. Nielsen, J.: Heuristic Evaluation, Usability Inspection Methods. John Wiley & Sons, New York (1994)
16. Beyer, H., Holtzblatt, K.: Contextual Design. Defining Customer-Centered Systems. San Fransisco. Morgan Kaufmann Publishers, San Fransisco (1998)
17. Avison, D., Fitzgerald, G.: Information Systems Development. McGraw Hill Higher Education (2006)
18. Brender, J.: Handbook of evaluation methods for Health Informatics. USA Elsevier Academic Press, USA (2006)
19. Cajander, Å., Larusdottir, M., Gulliksen, J.: Existing but not Explicit – The User Perspective in Scrum Projects in Practice. In: INTERACT 2013 LNCS Proceedings, INTERACT 2013, Cape Town, South Africa (2013)
20. Karlsson, F., et al.: Exploring user participation approaches in public e-service development. Government Information Quarterly 29(2), 158–168 (2012)
21. Lindgren, I., Jansson, G.: Electronic services in the public sector: A conceptual framework. Government Information Quarterly 30(2), 163–172 (2013)
22. van Velsen, L., et al.: Requirements engineering for e-Government services: A citizen-centric approach and case study. Government Information Quarterly 26(3), 477–486 (2009)

Increasing Accuracy by Decreasing Presentation Quality in Transcription Tasks

Frank Soboczenski[1], Paul Cairns[1], and Anna L. Cox[2]

[1] University of York, Department of Computer Science, Deramore Lane,
York, YO10 5GH, UK
{fs596,Paul.Cairns}@york.ac.uk
[2] University College London, Interaction Centre, Gowner Street,
London, WC1E 6BT, UK
Anna.Cox@ucl.ac.uk

Abstract. Many tasks with interfaces require entering data accurately, for example, entering patient data in an electronic records system or programming infusion pumps. However, human error is inevitable. Paradoxically, results from cognitive psychology suggest that representing information in a poorer quality format increases the likelihood of memorising the information accurately. This is explained by the dual system account of cognition where slower, more effortful but more accurate thinking is invoked via the poorer quality representation. We present two studies where we transfer these results to the domain of data- entry and show that poorer quality format of to-be-copied information leads to increased accuracy in transcription tasks. Moreover, this is not a consequence of the typical speed-accuracy tradeoffs. The results of our novel approach have implications for the design of data-entry tasks in domains such as healthcare.

Keywords: Human error, cognition, data-entry, presentation quality, number-entry, infusion pump, perception, safety-critical systems.

1 Introduction

Routine data transcription tasks occur in a large number of settings, e.g. entering student marks into a spreadsheet, and paying bills via an online-banking system. But people are also prone to errors and while a wrong letter or number is sometimes little more than an inconvenience, things can get pretty ugly in safety-critical domains. One such domain, depending on a high standard of accuracy and data quality, is healthcare. For example, nurses are often required to enter large amounts of sensitive data in electronic health records (EHRs) and to program drug infusion pumps with information located on a prescription or infusion bag.

A simple slip while transcribing patient data or programming a syringe pump in a hospital can quickly turn a routine task into a serious situation. Previous studies have shown that human errors are the source of more than 50% of incidents in most domains and in aviation can be as high as 90% [20]. In healthcare the process of studying errors is only now emerging [19]. Recent studies have shown that people

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 380–394, 2013.

make transcription errors frequently and many systems do little to detect and prevent errors [18]. These issues strongly motivate our interest in reducing error rates in transcription tasks and we view this work as demonstrating an important new approach in this area.

Transcription tasks require users to perceive the target information, which is then coded into easily remembered parts in their memory. This is followed by the decomposition of the remembered chunks into separate characters to associate the matching keystroke to the relevant character. Once this association is complete, the individual characters are then translated into movements to press the proper key during the data- entry process [14]. Previous research has shown that the ease with which a user can access the to-be-copied text impacts the strategies used to complete the tasks. So, how far away the to-be-copied content [3] is or whether or not the user has to click to reveal it [8], can lead to differences in accuracy of performance. This is explained by the Soft Constraints Hypothesis [8], which suggests that people are more likely to use a potentially less accurate *memory-intensive* strategy over a more accurate *perceptual-motor* one when the cost of accessing information is higher. Gray et al. (ibid) demonstrate that memory-intensive strategies can lead to higher error rates in transcription tasks and therefore argue that reducing the effort required to access information can be beneficial.

Whilst much research has focused on decreasing the costs of accessing information, recent research has demonstrated that there may be disadvantages to this approach if the information processing strategy adopted is too passive [11]. Supporting evidence from [5] suggests that *increasing* the effort required to access information to be transcribed can be beneficial. Specifically, they argue that information represented in a less readable format can increase the likelihood of accurately memorising the information [5]. One explanation for this finding is that, in order to make sense of the text or the numbers, the user experiences cognitive dysfluency when reading and so moves away from more automatic reading processes towards more effortful processes [2]. Different names are used to describe these kinds of processes such as for example *implicit* and *explicit* [7, 13] and the more neutral, and increasingly common terms, *System 1 vs. System 2* [17]. *System 1* refers to automatic, hard to control, effortless and fast processes [10]. In contrast, *System 2* is slow in execution, effortful and is limited in its capabilities [10]. Whether a *System 1* or a *System 2* process is chosen to complete a task is influenced by the task at hand. For example, when a calculation task changes from a relatively simple to a difficult task *System 2* takes over. An example of such a scenario is thinking of tasks such as the question of what is the result of 2x2 or the question of what is the result of 342x319. In the latter example, *System 2* takes control over the thinking process to produce an answer. The literature also presents an expansion of these well established thought processes by a third one namely *System 3*. This third process reflects on complex problem solving and prolonged major decisions, which take a very long time (i.e. months or years) and are therefore not applicable in this experimental context [6].

There are also attempts in the literature to examine ways to stimulate *increased cognitive engagement* in certain situations. In one particular study [5], researchers tried to provoke people to think harder by manipulating the ease with which fonts could be read, and measuring the impact on the participants' ability to memorise specific information. Their results indicate that there is potential for great improvement in performance of people's memory when presenting information in a harder to read font.

In this paper, we present two studies to explore the effects of *System 1* vs. *System 2* on data transcription tasks. This is a novel application of results of [5], which are about memory, to the domain of interaction by which we might contribute to the findings of Gray et al [8]. In our first study we investigate the effects of dysfluency of text in the domain of text-entry. We chose text-entry as an initial study because it is widely known that number-entry studies produce a very low error presentation rate (usually below 5%) [12,18]. As we are specifically looking for errors, a text-transcription task seemed ideal to simply get enough data to work with. The research instrument in the first experiment was the text-entry game *TypoMadness,* which we designed specifically for this purpose (see Fig. 2). Participants had to transcribe text phrases as quickly and as accurately as possible in two different conditions *Clear* & *Obscured.* In the second study, we transferred the results from text- directly to number-entry, comparable to the tasks in a healthcare environment. The task for participants in the second study was to re-enter given numbers in a calculator style interface i.e. serial interface in three different conditions *Clear*, *Medium* and *Obscured*.

2 Text-Entry Experiment

At first pass, it appears intuitive that text that is less easy to read (illegible or dysfluent text) would result in a higher number of errors in a text-transcription task as a user may be more likely to misperceive the text. However, evidence from [5] & [11] suggests that dysfluent text could encourage the adoption of an alternative information processing strategy, resulting in fewer errors. The hypothesis for this study was that people would make fewer errors and therefore archive a higher score in the *Obscured* condition than in the *Clear* condition.

2.1 Participants

20 participants (9 men) with a mean age of 27.2 years (st.dev = 8.5) and mostly students at the University of Anonymous were randomly recruited by email as well as personal invitation. Nine participants stated that the task was difficult. However, only four of these nine people were part of the *Obscured* group. Moreover, these four participants further stated other reasons such as being unable to use the backspace-key than the visibility of the given text as the reason why they felt the task was difficult. None of the participants had dyslexia or any other disability related to vision and reading. Additionally, all of the participants stated that they had normal or corrected to normal vision.

2.2 Design

A between-subjects design was used with ten participants in each group. The independent variable is the grade of visibility of the text. There were two conditions: *Clear,* which used clear visible content, (font-colour; #000000) and *Obscured,* which used less clear i.e. less visible content (font-colour: #BABABA) (see Fig. 1).

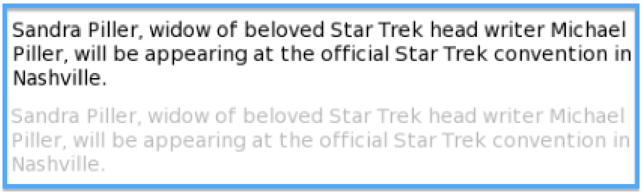


Fig. 1. Example of the text displayed in the *Clear* (upper) and *Obscured* (lower) conditions

The dependent variable was the score. It is the accumulated sum of correct entered characters in each task reduced by errors. An error in this study is anything that is not identical to the target text. This could be an additional space, upper or lower case difference, wrong or missing letters generally any sign different from the displayed text. Due to limitations of space, we have not reported an analysis by error type (slip, mistake etc.) in this paper. The set of 26 text phrases was counterbalanced across the study to avoid ordering effects.

2.3 Materials

The task was framed as a game in which the aim was to enter as many sentences as accurately and as quickly as possible within 15 minutes. Each sentence had between 95 and 161 characters and was displayed in either *Clear* or *Obscured* format. Additionally, the sentences were not connected in prose or any other form, so participants could not have predicted them. The sentences were randomly chosen from news and boulevard press websites such as theguardian.co.uk or cnn.com.

The game used for this study was implemented as a website in PHP with JavaScript. The website was connected to a MySQL database allowing the entered text phrases to be stored in the database for each participant. The displayed score was calculated out of the matching characters of each submitted text phrase. Each score for individual text phrases was stored in the database and displayed via cookies on the game website. The game website consisted of a display area in the centre of the page where the text paragraphs were displayed, a textbox for typing the text and a red button to submit the entered text and continue with the next task. The backspace button and the cursor keys were disabled in the text field so that participant could not change their entered text. All entered sentences no matter what they might be were saved in the database after a participant pressed the red button. The entered sentences were saved as a string in the database and compared to the original text. If there was any difference the score for that task was modified accordingly. It should be noted that where an error arose due to inserting an extra character, this did not cause all subsequent characters to be classed as erroneous.

The game was played on a MacBook Pro with a UK English keyboard layout and a 13" size display at 1280x800 resolution. The browser used for the experiment was Google Chrome running under Apple Mac OSX 10.7.5 (Lion). The game website ran in the browser, the browser window was maximised and controls were disabled from being shown so that only the game itself was visible. All other applications were closed. In addition, there was also an informed consent to sign at the beginning of the study and a brief paper-based demographic questionnaire to fill out by participants after they completed the main task. The demographic questionnaire was used to specify the participants but not further used in the analysis reported here. A smartphone was used to stop the time by the experimenter.

2.4 Procedure

After reading the instructions, participants completed two training sentences, which were separate from the main sentence set of the experiment. The target text was presented in a textbox in the centre of the individual webpage. After the participants finished entering the training session a new start page was displayed only with a start button. Participants could see their score (in form of a character count as seen in Fig. 2) displayed on the right side of the text field. On completion of the training, the participant proceeded to the main study by clicking on the start button.

The target text was presented either *Clear* or *Obscured* depending on the condition. Participants were instructed to enter the target text as quickly and as accurately as possible. In order to get the highest possible score, all tasks had to be completed within 15 minutes though participants were not shown how long they had been playing. The experimental design in form of a game made it necessary to put participants under some form of time-pressure. This was also done as means of increasing the error rate. Each time the participant pressed the red button, their score was updated though there was no direct indication of any errors. If they completed the task within the 15 minutes, the final pages displayed the overall score. The experimenter controlled the smartphone and stopped the participant at the end of the given time limit. After finishing the entry task, the experimenter offered each participant a small break before asking the participants to fill out a demographic questionnaire. The participants were then debriefed as to the goals of the study.

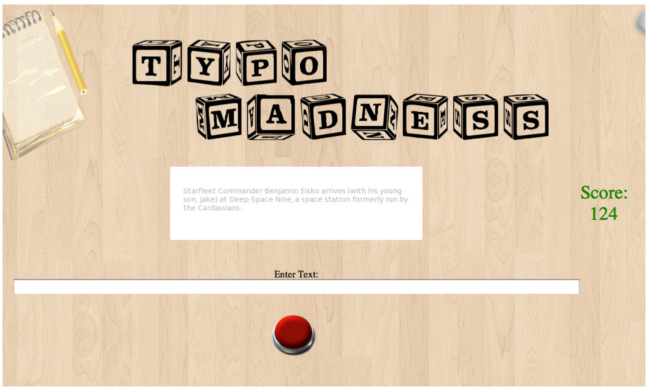


Fig. 2. The interface of the game *TypoMadness*

2.5 Results

The hypothesis for this study was that there would be significantly fewer errors in the *Obscured* group (i.e. a higher score) than in the *Clear* group. The mean of the errors made in the *Clear* group was 830.1 (sd = 516.3) whereas in the *Obscured* group was 360.4 (sd = 127.0). The spread of errors is much higher in the *Clear* group than in the *Obscured* group (see Fig. 3).

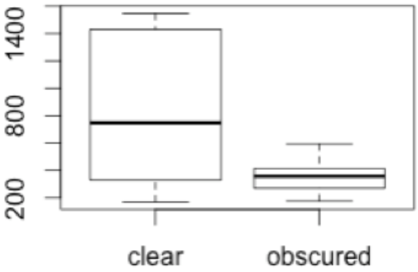


Fig. 3. Boxplot of total number of errors made by all participants in each condition

A Mann-Whitney-W test was conducted, as error data does not follow a normal distribution, which revealed a significant difference between the two groups (W = 23, p = 0.04). Furthermore, though we had a directional hypothesis, a more acceptable two-tailed test was chosen to avoid the controversies around one-tailed testing [1]. This move also resulted in a higher threshold for significance and therefore would give us more confidence in our findings.

In terms of speed of completing the task (see Fig. 4), the mean rate for entering sentences for the *Clear* group was 1.69 sentences/min (sd = 0.45) and for the *Obscured* condition 2.02 sentences/min (sd = 0.49). There was no significant difference in the rate of entering sentences (W = 26, p = 0.07).

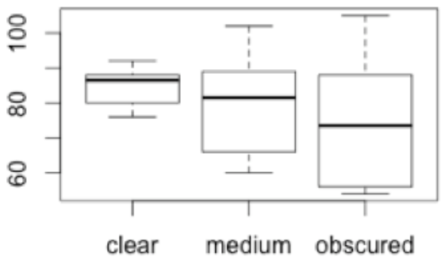


Fig. 4. Boxplot of the rate of typing (sentences/min) of participants in each condition

Moreover, the participants in the *Clear* group entered a mean of 23.1 (sd = 2.99) sentences and the in the *Obscured* group 25.0 (sd = 1.63). A boxplot illustrates the sentences entered by group (see Fig. 5). There were also no significant differences between the number of sentences participants completed (W = 32, p = 0.15).

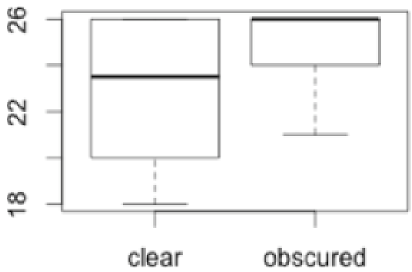


Fig. 5. Boxplot of the number of sentences entered by each participant in each condition

Four participants from the *Clear* group and seven participants from the *Obscured* group managed to enter all the 26 given text phrases and finished ahead of time. This leaves six participants from the *Clear* and three from the *Obscured* group who could not finish in time and therefore used the full 15 minutes.

2.6 Discussion

As predicted, the results showed that participants in the *Obscured* group achieved higher accuracy scores (i.e. fewer errors) than those in the *Clear* group. One explanation for this reduction in errors could be a speed accuracy trade-off. However there was no significant difference between the two conditions in neither the rate of entering sentences or on the total number of sentences entered by each group. Indeed, the participants in the *Obscured* group entered more sentences, working at a faster rate suggesting that the reduced readability of the font was not slowing the participants down. Nonetheless, they made fewer errors than the *Clear* group.

Our manipulation is identical to that of [5] who suggest that *increasing* the effort required to access information results in an increase in the likelihood of accurately memorising the information [5]. However, we only focussed on representing the to-be-entered information in a faded font-colour but not also in a different font-style, which indicates a more robust outcome. The results therefore suggest that a more memory intensive strategy may have been adopted by the participants in the *Obscured* group which, in this case, resulted in enhanced performance.

Focusing on limitations, it can be argued that the size of our sample (20 participants) is simply too small to have confidence in our data. This might be true and requires further studies to get a more robust set of data, which is one reason why we decided to present a second study on the effect of *System 1* and *System 2* in this paper. On the other hand a larger sample size might just present a larger effect of our detected result as we have chosen to examine our data with a higher threshold of significance (two-tailed test). An improved experimental design would suggest to use sentences, which are all of the same length to avoid possible control confounds. Additionally, grey text can be more legible than black text under specific circumstances, depending on the level of contrast with its background and the overall level of luminance. Some types of dyslexia respond well to contrast reduction for improving readability. However, there were no dyslexic participants (at least not that they were aware of) in our study. Additionally, prior experience in data-entry work of participants ore even any prior knowledge of the material might have influenced their performance in our task; therefore a more careful selection of participants for future studies needs to be considered. Moreover, there is also the argument that a too high grade of obscurity of text may result in participants getting frustrated or suffer from fatigue and therefore abandon the task. We regard this as major issue and designed our next study to explore performance across a range of levels of obscured information.

3 Number-Entry Experiment

As the results from our initial text-entry study were very promising we were keen on exploring the discovered effect of *System 1* vs. *System 2* in number-entry. Therefore, we directly transferred the task from text- to number-entry in the next experiment. The hypothesis for our second experiment was that there is a significant difference in the number of errors between the conditions *Clear*, *Medium* and *Obscured*.

3.1 Pilot

The reason why the number-entry study consists of three conditions resulted out of an initial pilot study among 53 first-year undergraduate students in computer science. This pilot to measure the effect of *System 1* vs. *System 2* in the domain of number-entry was designed accordingly to the previous text-entry experiment with two conditions *Clear* and *Obscured* to measure a significant difference in the amount of errors between these two groups. We discovered that participants did make more errors in the *Clear* condition but not significantly more errors. We also noted that our initial design of our *Obscured* condition was too strong i.e. the obscured numbers were too faded, which made it too hard to read for participants. It was then decided to design this study with three conditions to explore a wider range of different levels of obscurity. These three conditions were all less obscure than in the pilot study. The participants of this pilot study were not allowed to take part in the described follow-up study.

3.2 Participants

Participants were randomly recruited by personal invitation or by email. There were 30 participants (16 men) with a mean age of 26.5 years (st.dev = 8.2) in total.

Participants were mostly familiar with a calculator based number-entry interface as used in the study but none of them had previous experience of working in the healthcare sector. However, three participants stated that they were familiar with infusion pumps. Two people stated that they found the task difficult (one in the *Clear* and one in the *Obscured* group) and two different participants stated that they did not consider themselves to have normal vision (also one in the *Clear* and one in the *Obscured* group). Five participants had never used a number pad before, 7 rarely used it, 14 sometimes used it and only 4 participants (one each in group *Clear* and *Obscured* and two in group *Medium*) often use a number pad to enter numbers.

3.3 Design

A between-subjects (3 groups) design was used with 10 participants in each group. The independent variable is the grade of visibility of the presented numbers and the textbox where participants enter the numbers. There were three conditions: *Clear,* which used clear visible content (font-colour; #000000), *Medium* (font-colour; #BDBDBD) and *Obscured* (font-colour: #DFDFDF), which used less clear i.e. less visible (light-grey colour) content (see Fig. 7). The dependent variables were the total number of errors made, particularly the rate of errors and also the number of corrected errors in each group. A single error is counted when any number is entered other than the displayed number. Further, a corrected error is an error where participants noticed that they made an error and entered the number again. These errors were not added to the total number of errors. As previously, error type is not considered in our analysis.

3.4 Materials

The task set for participants was designed as a game in which the aim was to enter as many numbers as accurately and quickly possible in a calculator-based interface within 5 minutes. Compared to the first experiment a five-minute limit was chosen for this study as the pilot study showed that participants could transcribe content that consists of numbers considerably faster than text. The visual design of the serial i.e. calculator-based number-entry interface was carefully created after real world infusion pumps such as the Baxter© AS40A or the Graseby™ 3400 syringe pump (see Fig. 6). For the purpose of this study only the number pad of the infusion pumps with an *Enter* and a *Clear* key was replicated all other buttons of the original interface design were removed from the interface as they could have a possible unwanted effect on participants.

The interface was implemented in PHP and JavaScript with a logging feature, which recorded all user interactions with the virtual device and wrote all interactions into a MySQL database. The interface provided thirteen functional buttons usable by the keyboard-only interaction (numpad).

The display itself consisted of 16 components. Three number displays (number to be entered, score and number input field), *Enter* and *Cancel* buttons and 0-9 number buttons including the decimal key. By pressing the *Enter* key the current displayed number was saved and at the same time the display was cleared. The display itself provided participants to enter a 5-digit number with two digits after a separating decimal point as previously gathered numbers from infusion pump logs confirmed the validity of this number range [19].

Fig. 6. Infusion pump models Baxter© (left) and Graseby™ (right)

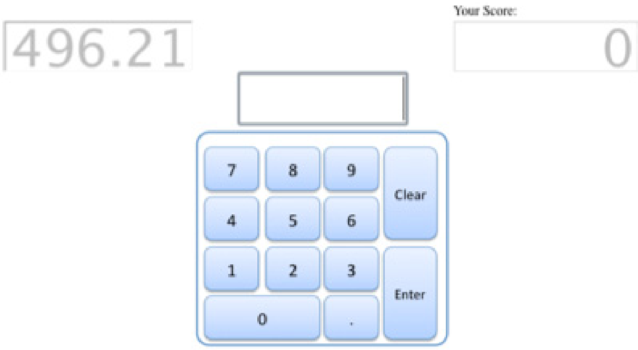


Fig. 7. Experimental setup; serial interface in the *Medium* condition

There were three versions of the interface used in the study. One where the number to be entered was displayed with a normal font-colour (black on white background) (group *Clear),* one where the number was displayed with a slightly lesser visible font-colour (light-grey on white background i.e. #BDBDBD) and the last one where the number to be entered was displayed in a hard to read faded font-colour (strong-faded grey on white background, i.e. #DFDFDF) in group *Obscured.*

In addition to the interface described above, which was accessed via logging into a website, the instructions as well as an informed consent document were printed out on paper for the participants to read and sign. Additionally, there was also a brief demographic questionnaire, which was implemented as a Google form. Again, this was not used in the analysis reported here.

The computer used was an Apple MacBook Pro 13” connected to an external keyboard and monitor with a 17” size display at 1920x1080 resolution. The browser used for the experiment was Mozilla Firefox 10 running under MacOSX 10.7.5 (Lion).

The interface ran in the browser, the browser window was maximised and controls were disabled from being shown so that only the interface area itself was visible. All other applications on the machine were closed and the auto-completion mode of the browser was deactivated. Furthermore, the keyboard was prepared so that participants could only use the numpad to enter numbers all other keys were disabled. Stickers were put in place on the '+' key to make it easier for the participants to associate the particular '*Clear*' key function.

Each number a participant had to enter, whether it was for the training to get familiar with the controls or the main study, was randomly generated in the game and displayed in the left upper corner of the interface (see Fig. 7). A smartphone was also used for this study to stop the total duration each participant used to enter as many numbers as possible.

3.5 Procedure

After a participant had read the instructions, he or she was asked to become familiar with the use of the keyboard controls and the display of the interface. For this purpose they were asked to run 10 training trials, which were separate from the main experiment. Participants were asked to sit straight in front of the monitor to avoid a possible effect on the perception of the numbers as the *Medium* and *Obscured* information could be perceived differently from a different angle. After a participant became familiar of how the serial number-entry interface worked he or she was then asked to proceed to the main study. Randomly selected participants (first come first serve basis) were asked to enter the given numbers either in the *Clear, Medium* or *Obscured* group within 5 minutes as quickly and as accurately as possible. The experimenter stopped the participant at the end of the 5-minute timeframe. During the experiment participants themselves could not see how much time had passed. Each time a participant pressed the *Enter*-key the number he or she just entered which was shown in the centre display was saved to the database and a new number was instantly and randomly generated at the upper left display. Additionally, at the same time the numbers in both displays were compared and if they matched the score were increased by 100 points. Any other entry than the correct number resulted in a decreasing of the score by 50 points. After finishing the number-entry task, the experimenter offered each participant a small break. Each participant was then asked to fill out a small online demographic questionnaire and the participants were debriefed as to the goals of the study.

3.6 Results

The hypothesis for this experiment was that there is a significant difference in the number of errors between the three conditions *Clear*, *Medium* and *Obscured*. Participants entered in total 2391 numbers for all three groups (see Fig. 8), which resulted in 44 total errors (1.8%) and 149 corrected errors. 849 numbers were entered in group *Clear* (28 errors), 796 in group *Medium* (12 errors) and 746 in group *Obscured* (4 errors), respectively. The mean of the errors made in the *Clear* group was 2.8 (sd = 2.20) whereas in the Medium group was 1.2 (sd = 1.39) and in the Obscured group 0.4 (sd = 0.52).

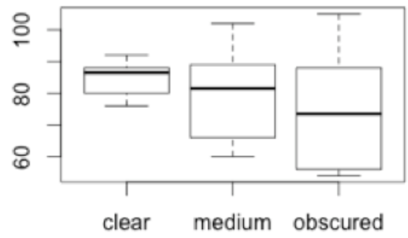


Fig. 8. Boxplot of amount of numbers entered by all participants in each condition

A Kruskal-Wallis test was conducted, as error data does not follow a normal distribution (see Fig. 9). The results showed a highly significant difference between the errors made in the three groups (H = 9.78; df = 2; p = 0.007). However, there was no significant difference in the total numbers entered between the groups (H = 2.08; df = 2; p = 0.35) (see Fig. 8) and the number of corrected errors (H = 0.23; df = 2; p = 0.89). The mean of the numbers entered in the *Clear* group was 84.9 (st.dev = 5.45) whereas in the *Medium* group was 79.6 (st.dev = 13.95) and in the *Obscured* group 74.6 (st.dev = 17.61). Additionally, the mean of the corrected errors in the *Clear* group was 5.2 (st.dev = 5.18), whereas in the *Medium* group was 5.3 (st.dev = 6.80) and in the *Obscured* group 4.4 (st.dev = 3.59).

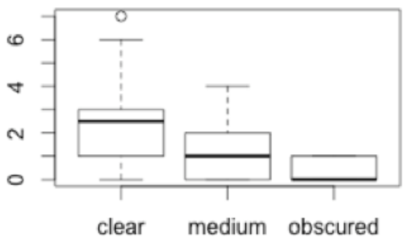


Fig. 9. Boxplot of the number of errors made by all participants in each condition

Additionally, taking into account the number that each individual entered in relation to the errors, the rate of errors was also significantly different (H = 8.89; df = 2; p = 0.01). Moreover, the follow-up multiple comparison testing (p = 0.05) of the total errors made by each participant in each group revealed, that only the difference between group *Clear* and *Obscured* is significant (observed difference of 11.75 > critical difference of 9.43). A second follow-up test (p = 0.05) of the rate entered by each participant in each group revealed also that only the group *Clear* is significantly different from group *Obscured* (observed difference of 11.30 > critical difference of 9.43).

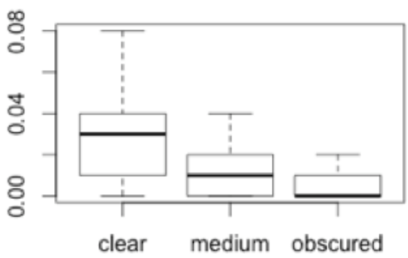


Fig. 10. Boxplot of the error-rate of each participant in each group

Eleven participants (36%) made at least one error with only one person making no error in the *Clear* group, 4 persons making no error in the *Medium* group and 6 persons making no error in the *Obscured* group.

3.7 Discussion

Like the text-entry study the results showed that participants in the *Medium* and *Obscured* group achieved higher accuracy i.e. less errors than those in the *Clear* group. However, only the difference between *Clear* and *Obscured* is significant, which again suggests that the experimental manipulation similar to those suggested by the pilot has worked and there is a trend indicating that a more obscured representation of information can lead to less errors being made in data-transcription tasks.

Still, as observed in the pilot a presentation of content that is too obscure might have the opposite effect of making numbers simply too hard to read and therefore resulting in people abandoning the task.

Nevertheless, the discovered result is not due to possible speed-accuracy tradeoffs. Yes, participants entered fewer numbers in the *Obscured* group than in the other groups however not significantly different.

Interestingly, whilst the number of committed errors was fewer there is no difference in the number of the corrected errors. This may suggest that people use different strategies to correct errors or people are self-detecting their slip errors. While people who commit the errors think their action was correct, those who correct errors are aware of their mistakes suggesting that they are self-evaluating their actions on the fly. However, if people use a memory intensive strategy i.e. a strong encoding in memory then perhaps the corrected errors are made by comparison with memory rather than comparison with the original. It is interesting to note that there are fewer corrected errors, which also suggests different mechanisms behind corrected and uncorrected errors.

Like other comparable studies [12, 18] the rate of number-entry errors is still low (1.8% in this case) but this does not mean it is less important to investigate why these errors happen. It is known that misprogrammed infusion pumps are the second most frequent cause of medication errors making it imperative to reduce as many errors as possible [16]. An error rate of 1.8% in medical device context can still result in grave consequences. And therefore, whilst the total number of errors is modest, there is still much to explore and to learn.

4 Overall Discussion and Conclusions

We have presented two studies, which showed that less visible content could affect the information processing strategy adopted in a text-entry as well as in a number-entry task and, counter-intuitively, improved accuracy in the specific task.

This therefore suggests that less visible information can result less errors being committed in both text- and number-entry tasks. Moreover, the results show a strong support for the effect of *System 1* vs. *System 2* in the context of interaction. If tasks require a higher cost of access then evoking *System 1* vs. *System 2* might overcome the various problems presented by [8].

Further work employing eye-tracking measures is required to explore whether the information processing strategies adopted by those in the two groups result in differences in the number of fixations on, and the speed of reading of the target text. Specifically, it will be important to understand whether the enhanced performance of participants in the *Obscured* group is the result of slower more deliberate processes such as additional perceptual monitoring of the target text resulting in enhanced memory for the text or more careful execution of the motor movements required to enter the content. When focusing on design, our discovered results represent a potentially important contribution for the design of future medical devices.

However, there is the possibility that this effect will wear off over time. Future work will focus on a long-term study, where we can explore how long the discovered effect lasts in data-entry tasks.

Additionally, fatigue is another issue that needs to be examined in this context. There is evidence in the literature that, for example, nurses who are on extended work duty possess a significantly decreased level of vigilance [15]. The grade of frustration is another interesting aspect to explore. For example, if the visibility of the *Obscured* text is too low i.e. if the text is too hard to read then there is the potential that people will get too frustrated in entering numbers (or text) and simply lower their efforts towards a successful completion of the task or even abandon the task completely. In future studies, we are keen to consider how this effect can be applied further and explored to improve safety-critical systems such as infusion pumps.

Moreover, as healthcare personnel do not usually transcribe information from an electronic source but also from paper-based sources there is further room to explore the application of our findings in this context.

There is also the argument in the literature that text is processed differently in sense of string and individual characters [4, 9]. Additional work is required to explore the relations between these paradigms. For example, an alternative design for a text-entry study could focus on using content in welsh-language and let participants re-enter the text again in form of a game to investigate how people process text either in individual characters or as connected strings.

We believe our results revealed important findings based on a new approach to understanding the text-entry task. This may have significant implications for the design of tasks in safety-critical systems. We are therefore keen to continue to explore this seemingly fruitful research direction.

References

1. Abelson, R.P.: Statistics as principled argument. Lawrence Erlbaum Associates, Hills-dale NJ (1995)
2. Alter, A.L., Oppenheimer, D.M.: Unity the tribes of fluency to form a metacognitive nation. Personality and Social Psychology Review 13, 219–235 (2009)
3. Back, J., Brumby, D., Cox, A.: Locked-out Investigating the Effectiveness of System Lockouts to Reduce Errors in Routine Tasks. Psychology 1, 3775–3780 (2010)
4. Bradshaw, J.L.: Three interrelated problems in reading: A review. Memory and Cognition 3, 123–134 (1975)
5. Diemand-Yauman, C., Oppenheimer, D.M., Vaughan, E.B.: Fortune favors the Bold (and the Italicized): Effects of disfluency on educational outcomes. Cognition (2011)
6. Dijksterhuis, A.: System 3 thinking. Presented at the Cognitive Psychology Section Annual Conference Proceedings (2012)
7. Evans, J.S.B.T.: In two minds dual-process accounts of reasoning. Trends in Cognitive Sciences 7, 454–459 (2003)
8. Gray, W.D., Sims, C.R., Schoelles, M.J., Fu, W.: The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. Psychological Review 113, 461–482 (2006)
9. James, K.H., James, T.W.: Letter processing in the visual system: Different activation patterns for single letters and strings. Cognitive, Affective, and Behavioral Neuroscience 5, 452–466 (2005)
10. Kahneman, D.: Explorations of the Mind Intuition. UC Berkeley Graduate Council Lectures, Humanities (2008)
11. Morgan, P.L., Patrick, J.: Designing Interfaces That Encourage a More Effortful Cognitive Strategy. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 54, pp. 408–412 (2010)
12. Oladimeji, P., Thimbleby, H., Curzon, P., Iacovides, J., Cox, A.: Exploring Unlikely Errors using Video Games: An Example in Number Entry Research. In: Workshop on Safety-Critical Systems and Video Games: Contradictions and Commonalities, Held at Fun and Games 2012, Toulouse, France (2012)
13. Reber, A.S.: Implicit Learning and Tacit Knowledge. Journal of Experimental Psychology: General 118, 219–235 (1989)
14. Salthouse, T.: Perceptual, cognitive, and motoric aspects of transcription typing. Psychol. Bull. 99, 303–319 (1986)
15. Scott, L., Rogers, A., Hwang, W.-T., Zhang, Y.: Effects of Critical Care Nurses' Work Hours on Vigilance and Patients' Safety. American Journal of Critical Care 15, 30–37 (2006)
16. Smetzer, J., Cohen, M.: Preventing medication errors related to drug delivery devices. In: Cohen, M.R. (ed.) Medication Errors, 2nd edn. (2006)
17. Stanovich, K.E., West, R.F.: Individual differences in reasoning: Implications for the rationality debate. Behav. Brain Sci. 23, 645–726 (2002)
18. Thimbleby, H., Cairns, P.: Reducing number entry errors solving a widespread, serious problem. Journal of the Royal Society, Interface / the Royal Society 7, 1429–1439 (2010)
19. Wiseman, S., Cairns, P., Cox, A.: A Taxonomy of Number Entry Error. In: HCI 2011: The 25th BCS Conference on Human-Computer Interaction (2011)
20. Zhang, J., Shortliffe, E.H., Patel, V.L., Freed, M., Remington, R.: The Nature of Human Errors An Emerging Interdisciplinary Perspective. In: Proceedings of the 22th Annual Conference of the Cognitive Science Society, CogSci 2000 (2000)

You Can't Touch This: Potential Perils of Patient Interaction with Clinical Medical Devices

Paul Noble and Ann Blandford

UCL Interaction Centre, University College London,
Gower Street, London, UK
Paul.Noble@cs.ucl.ac.uk, A.Blandford@ucl.ac.uk

Abstract. Clinical medical devices are designed with the explicit assumption that trained medical team members will operate them in appropriate hospital environments. As technological complexity increases, along with the possibility to create specific ward configurations, the potential for unusual interaction combinations poses challenges for safety and training. Resilience engineering proposes that a system should cope with disturbances and unexpected conditions. Consequently, an important consideration for design is to examine medical device interactions that can be considered 'non-routine'. In recognition of the localised nature of clinical practice, and in order to investigate the broad range and type of non-routine occurrences, a novel interview approach was adopted involving medical researchers and practitioners. Examples of non-routine interaction were obtained across a diverse range of localities. Covert patient interactions and dangerous configuration combinations were identified which adversely affected treatment. Drawing on these concerns the potential role of patient involvement in bolstering system resilience is discussed.

Keywords: Medical Devices, Safety, Resilience Engineering, Customisation.

1 Introduction

Any patient or visitor to a medical facility cannot fail to notice modern technological equipment situated in wards, departments, and operating theatres. These widely used interactive devices undergo natural technology cycles in which manufacturers drive advancement by demonstrating the limitations of yesterday's models, and market the latest feature sets. As a result, it is not uncommon for these devices to increasingly suffer from 'feature bloat', in which new and seldom-used functionality makes operation increasingly confused and complex. An advantage of this approach is that the same medical device can be used across many hospital departments, where specific functionalities are all accessible through the device interface. However, following an institutional drive for simplification, medical device manufacturers have provided the ability to adapt and configure aspects of the user interface; for instance to show or hide menu items, or to re-order procedural input steps and set default values. This permits device tailoring by hospital department and ward type.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 395–402, 2013.

Human-computer interaction (HCI) has long concerned itself with the inadequacy of systems that fail to reflect working practice, and are inflexible in use. In particular, many ethnographic studies observing interactions ‘in the wild’ have emphasised frequently the unplanned and situated nature of technology use [10]. Medical devices operate within behaviourally complex and emergent environments, and the latest systems can simply relocate error by introducing as many new issues as existing problems that they are designed to solve [5]. In this regard, the opportunity for device configuration and tailoring poses a number of design decisions balancing safety, training, and efficiency.

Resilience engineering [4] offers an alternate perspective to that of traditional risk management, examining particular strategies and procedures adopted by those within a system to allow it to succeed. Of specific interest is how the system is prepared for, and reacts to, disturbances and unexpected events. This shift in thinking has contributed to a growing interest in the examination and identification of resilient strategies surrounding interaction within complex socio-technical systems.

Bearing in mind the highly localised nature of medical device use, the objective of this paper is to understand the clinical environment by examining unexpected events and disturbances, and acquire generalisable insights for design. To address this aim, the paper presents a study of these ‘non-routine’ interactions. We recognise that the phrase non-routine is not ideal, however we deemed it the best general descriptor which could convey the temporal, normal, and legal aspects of the following properties: atypical, infrequent, unapproved, unauthorised, and untrained. Initially, the goal was to discover the breadth of technological interactions within the local clinical environments that were described as non-routine by participants. Non-routine dimensions were then extracted from these examples in order to facilitate understanding of the tensions within this space. Secondly, a resilience engineering perspective was applied in order to uncover areas of system vulnerabilities and weaknesses.

In this paper, existing literature from the fields of resilience engineering, appropriation, customisation, and ethnomethodology are initially reviewed. Subsequently, the novel methodology used in the study is described where interviews with medical researchers and practitioners are conducted. Situated non-routine dimensions are then presented and selected examples discussed. The following section then discusses an identified system area with low resilience (brittleness), and discusses the potential for unsafe complications with customised devices. The paper concludes by arguing for patient centered care in device design and practice.

2 Previous Work

Health care quality assurance literature [6] describes successful healthcare organisations as having a requirement to be both flexible and efficient in order to succeed. Lillrank suggests that this is achieved through a mix of standard, routine, and non-routine process elements, where repetitive processes can be standardised, routine processes consisting of both repetition and variability can be executed in dissimilar ways, and non-routine processes can be considered so variable as to be chaotic. This definition of non-routine complements resilience engineering's argument that a system should handle disturbances and unexpected events. Sujan [11] describes a resilience observation study of a hospital dispensary where individual staff cope with non-routine occurrences on a daily basis, prioritising tasks in order to anticipate and moderate further disturbances. It is the understanding of the dynamics and differences from mundane practice to those surrounding the extraordinary event that is important for the resilient design of medical devices.

Historically, the introduction of new technologies into a variety of work domains has often failed due to misconceptions about local working practice. No system is a perfect fit, and the way in which technologies are successfully adopted, adapted and incorporated into working practice is through the process of appropriation. Interactive technology can also allow some degree of customisation to better meet the needs of particular individuals and groups of users. Therefore, customisation refers to a process of technology ownership where the needs of the individual are balanced with those of the group. Appropriation 'concerns the adoption patterns of technology and the transformation of practice at a deeper level' [2], and differs from customisation in that there is a co-evolution [9] between the fitting of the technology within current organisational structures, and with the adaptation of working practices to support collaboration around the new technology. Randell [8] examined appropriation and customisation within the intensive care unit (ICU), observing situations where technologies are molded into working practices and are made to work in the way that the nurses want them to work, and not as envisaged by the manufacturer. Importantly, there is a strong motivation to share information and develop a working practice around the technology. This is described also by Wenger [12] as how an individual develops membership of a Community of Practice (CoP).

In this paper, a distinction will be made between non-routine incidences occurring due to local appropriations and customisations, and all others. This is primarily because departments and wards generally have very good reasons for behaving in a particular way, based upon a solid history of experience. To an individual observing outside the particular CoP however, the reasons for these behaviours can be unclear. In addition, local practices that are not generalisable are not of interest to this study.

Although much work and attention has focused on case studies where medical device interactions have led to undesirable outcomes, less research has been conducted on the analysis of situations where devices are used in a non-routine way but do not directly cause harm. System stability and quality of care may be impacted even though there is no perceived association with error. The underlying objective of this work is to analyse unusual socio-technical interactions in order to better understand how to support patient health care through technology. Prior research leaves much about resilient situated device use unanswered. In particular, there is little understanding of the experiences of those who are outside the CoP but who still interact with the situated device, and what the implications are for system resilience and patient safety.

3 Methodology

Resilient strategies along with appropriated and customised interactions can be considered non-routine, dependent upon the comprehension of those inside or outside the workgroup community. In order to investigate the breadth of these non-routine medical device interactions a novel interview approach involving medical researchers and practitioners was adopted. A meta-analysis of the experiences of situated researchers and clinicians was conducted, probing their findings and research, and reinterpreting them in a different way. To be clear, this does not simply imply that a literature review was undertaken; the experiences and observations of those who conducted research in particular clinical situations were directly sought.

7 academics and research associates (including 2 clinicians) on the CHI+MED medical project[1] were chosen as participants in order to allow access to a broad range of professional experiences. All academics had conducted a number of situated medical studies as prior work. To facilitate the collection of rich and varied examples, participants were left to decide for themselves what the term ‘non-routine’ implied. All were invited to prepare some examples of non-routine interaction that they had encountered prior to interview. The interview consisted of two parts where prepared examples of non-routine interaction were initially recounted, before examples and experiences contributed by other participants were shared. In this way, an increasingly rich collection of examples was progressively used to stimulate and sharpen discussion. Researchers were interviewed in ascending order of experience, and prior to clinicians in order to allow examples and interpretations to be challenged or endorsed by final practitioner interviews. Semi-structured interviews with participants lasted approximately one hour and were recorded and transcribed. Initially data was categorised and filtered to remove customisation and appropriation examples. A thematic analysis was then conducted in order to draw out and identify dimensions influencing non-routine interactions. The initial analytic goal was to broadly understand the dynamics of those within the clinical environment in their efforts to manage non-routine events. Subsequently, a resilience engineering perspective was applied in order to elicit design insights from generalisable examples of use.

4 Results

The participants together contributed 29 examples of non-routine device interaction. 12 examples described work practice appropriations or customisations, 7 described ‘forced’ situations where there was compelling motivation for interaction within a timeframe, 4 described situations where medical research interactions were conducted by medical staff and patients, 4 described interaction errors primarily due to gaps in knowledge and experience, and the remaining 2 are what we termed ‘covert’ patient interactions.

[1] CHI+MED (Computer-Human Interaction for Medical Devices, EP/G059063/1) is an EPSRC-funded project to improve the safety of interactive (programmable) medical devices.

4.1 Understanding the Clinical Environment

Four main aspects influencing whether an interaction is considered routine or not were identified. In order to assist understanding these have been arranged on a diagram representing a space of potential clinical device interactions (Figure 1). The *legality* of the interaction examines formal procedures, where subsequent approval must be sought if none exist. Work *practice* ascertains if best and local work practice is being applied, or whether workarounds are being followed. The *novelty* of the situation considers the requirement for an entirely original solution, or if a close variation can be adopted. User *training* distinguishes between complete and essential system training. The fluid and temporal nature of each dimension should be noted.

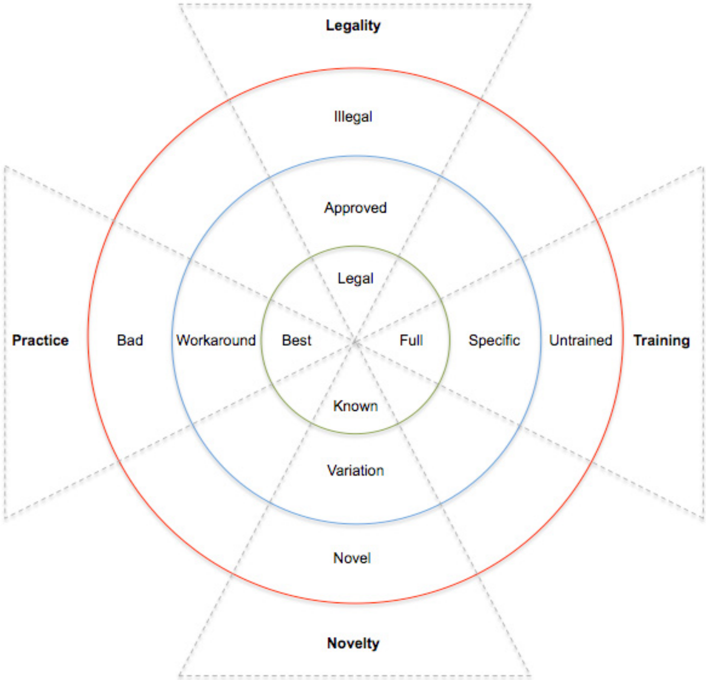


Fig. 1. Space of potential clinical device interactions

From Figure 1 it is clear that non-routine interactions are only possible in the middle and outer rings, with systemic risk increasing as a function of radius. Any individual dimension has the potential to pull the system into a region of risk. This analysis highlighted a vital role played by those community members who gave *approval* for responses to non-routine interactions. Prompt approval is critical as to how such a system resiliently and legally operates in an unpredictable environment and functions by both providing a damping mechanism through the coordination of an appropriate response, and that of information distribution by disseminating awareness of a particular event.

One participant's example described the use of a newly introduced portable heart monitor to provide visibility of a patient's condition during ward transfer:

On this occasion, the heart monitor failed to activate after charging and displayed a "BATT COND" message. After some experimentation, it was discovered that if the battery was removed and reinserted, the message cleared and the unit functioned once more. Subsequently it was realised that the heart monitors' rechargeable batteries are intended to be discarded after 50 charge cycles in order to guarantee safe operation. The ward does not have the resources to register every time the batteries are recharged, and it has now become practice to remove and reinsert the battery to reset the device when the event reoccurs.

An analysis of this example using the identified non-routine dimensions highlighted that initially this was a completely *novel* situation and so no *training* could have reasonably been provided to tackle this event. However, a solution was found and the team muddled through. Approval was then given for the technically *illegal* procedure, and the workaround became (possibly bad but known) *practice*.

4.2 Covert Patient Interactions with Medical Devices

The majority of the examples of non-routine interactions contributed were appropriations and customisations. This is unsurprising considering that these processes are normal means of embedding a technology and tailoring for practice. Of the other examples, those that described interactions in which patients would covertly interact with their clinical device were of particular interest, due to their potential impact on working practices and system resilience.

In this study, there were two main motivators for patient interaction with clinical medical devices; the desire for unrestricted doses of a controlled substance, and for relief from device alarm noise. These examples all involved infusion pumps; devices that intravenously pump medicinal fluids into a patient's circulatory system. Most modern infusion devices provide functionality in which the keypad panel can be locked out in order to restrict access[2]. A physical button located on the casing, or selected through the software menu screen typically activates this type of locking mechanism. Understandably, patients are motivated to attend to this activity, as described by Participant A:

'But some patients are also very wise to the fact that there is a keypad lock button...because they watch, and I always teach nurses that if you are going to put the keypad lock on don't make it obvious that you are looking for the button to press that's not in a standard place...and that the patient actually works out what you're doing. You can actually do it on this particular pump by just sort of putting your hand on the top and looking like you're holding the pump as you're doing everything else'

[2] Functionality generally aims to prevent dosage rate increases.

Another strategy used by patients is to look up the device manuals online, and discover the default locking codes or physical locking locations [3]. In these manuals, interactions to be prevented are variously described as 'unauthorised', 'tampering' and even 'malicious tampering'. This may seem a little harsh in the case of the patient who simply seeks relief from a constantly alarming device, however frequently resetting the device alarm can also temporarily shut off the pump with unintended consequences. Participant B:

'Nobody wants to be shut in a room with something that's persistently alarming. So the problem was that when they were rechecking the blood of this patient they discovered that clearly he couldn't have been having all of his therapy. The problem is that if its [the pump] not infusing at a certain rate then you are not getting the target dose within a certain time frame, so the concentration never reaches the critical level'

In this case, the patient had been covertly resetting his alarm so frequently that in the average period that the pump was operating correctly, there was not enough time to build up a particular concentration of the infused drug in his bloodstream.

Looking up device manuals online can appear a responsible alternative for a patient to exercise in preference to device menu exploration. However, the potential consequences of looking up information online can be serious, because as we have discussed it is possible for different wards to have the same medical device but with appropriated and customised configurations. Participant C recalled an example emphasising the dangerous potential for the untrained interaction of a customised device by medical team members or patients:

'I remember with the volumetric [infusion] pump, the trainer he pointed out that the bolus function on the thing doesn't mean bolus, or the primer, I can't remember but there was a specific function that either said prime or bolus or whatever, and he just said it doesn't do that. Press this button for something else essentially. So I think that's more to do with how the engineers set up the pumps'

Irrespective of what the particular pump functionality technically does here, the pump has been configured in a non-standard manner where a standard and labelled functional button delivers a different operation.

5 Conclusions and Future Work

As interactive medical devices move towards an increasingly feature rich future, it is apparent that some means of interface simplification is necessary. This paper examined potential complications caused by the ability to configure and customise, where the removal of standardisation increases the opportunity for non-routine interaction. The dynamics of non-routine interaction in a clinical medical environment suggest that device customisation will increase the burden of localised training by experienced community members, and impact safety. However, these disadvantages come at the gain of efficiencies for community members, and if possible, a safe balance must be found. One particularly interesting result of this study is the identification of covert patient interactions. These covert interactions already incur safety consequences, but are made potentially riskier when customised devices are operated using standard manuals found online. Restriction of information as means to control [1] has now largely been circumvented by these manuals. Using the resilience engineering perspective as a lens to uncover the *absence* of resilience can be illuminating. Particular areas of concern are situations where humans are prevented from contributing to system resilience. According to Participant D: *'The patients aren't there to think about what's happening with their treatment that's the nurses job'*. However, involving patients in their own treatment would appear to be a particularly resilient action. Considering the management of false medical device alarms, this has the possibility to free up significant amounts of nursing time, as well as enormously improving the wellbeing of the patient. The patient should be integrated into the workgroup community and be made aware of interaction implications. Device designers also need to recognise and anticipate cases of patient interaction. Resilient infusion pump designs such as that by proposed by Nemeth [7] offer visualisation of a complete cycle of a treatment, and would have prevented some of the issues discovered in this paper. Achieving a safe mix of device customisation and safety would appear to be a delicate balance, and future work will explore the deeper issues surrounding the situated trade-offs incurred.

References

1. Adams, A., Blandford, A., Lunt, P.: Social empowerment and exclusion: A case study on digital libraries. ACM Trans Comput. Hum. Interact. 12(2), 174–200 (2005)
2. Dourish, P.: The Appropriation of Interactive Technologies: Some Lessons from Placeless Documents. Computer Supported Cooperative Work 12(4), 465–490 (2003)
3. Grissinger, M.: A Patient's Wandering Eye and the Internet Can Lead to Pump Tampering. Pharmacy and Therapeutics 34(6), 283–328 (2009)
4. Hollnagel, E., Woods, D., Leveson, N.: Resilience Engineering: Concepts and Precepts. Ashgate, Burlington (2006)
5. Koppel, R., Metlay, J., Cohen, A.: Role of computerized physician order entry systems in facilitating medication errors. The Journal of Urology 293(10), 1197–1203 (2005)
6. Lillrank, P., Liukko, M.: Standard, routine and non-routine processes in health care. Intl. Journal of Health Care QA Inc. Leadership In Health Services 17(1), 39–46 (2004)
7. Nemeth, C.P., Hollnagel, E., Dekker, S.: The Infusion Device as a Source of Healthcare Resilience. In: Resilience Engineering Perspectives: Preparation and Restoration (2009)
8. Randell, R.: User customisation of medical devices: the reality and the possibilities. Cognition, Technology & Work 5(3), 163–170 (2003)
9. Rogers, Y.: Exploring Obstacles: Integrating CSCW in Evolving Organizations. In: Conf. on Computer Supported Cooperative Work CSCW, pp. 67–78 (1994)
10. Suchman, L.: Plans and Situated Actions. Cambridge University Press (1987)
11. Sujan, M., Pozzi, S., Ingram, C.: Resilience as Individual Adaptation: Preliminary Analysis of a Hospital Dispensary. In: CEUR Procedings 4th Workshop HCP, pp. 48–51 (2011)
12. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press (1998)

A Dog Tail for Utility Robots: Exploring Affective Properties of Tail Movement

Ashish Singh and James E. Young

Department of Computer Science
University of Manitoba
Winnipeg, MB, Canada R3T 2N2
{ashish,young}@cs.umanitoba.ca

Abstract. We present a dog-tail interface for utility robots, as a means of communicating high-level robotic state through affect. This interface leverages people's general knowledge of dogs and their tails (e.g., wagging means happy) to communicate robotic state in an easy to understand way. In this paper, we present the details of our tail construction, and the results of a study which explored a base case of people's reactions to the tail: how various parameters of tail movements and configuration influence perception of the robot's zoomorphized affective state. Our study indicated that people were able to interpret a range of affective states from various tail configurations and gestures, and in summary, we present a set of guidelines for mapping tail parameters to intended perceived affective robotic state.

Keywords: human-robot interaction, animal-inspired interfaces, affective computing.

1 Introduction

As robots continue to enter people's spaces and environments it will be increasingly important to have effective interfaces for interaction and communication between the people and robots. One such aspect of this communication is people's peripheral awareness of robots' actions and motions, a communication channel which has proven important for human-human interaction in work roles similar to utility robots; for example, as office assistants, robots should be as non-intrusive as possible yet still provide ambient awareness of their tasks [10, 32]. In this paper we introduce the use of a dog tail for robots as a peripheral awareness communication mechanism (Fig. 1).

Dogs are one of the most preferred pets in many countries around the world, and as such we posit that even non-dog owners have a passing knowledge of dogs from the general "social stock of knowledge" [4, 14]. Thus we believe that dogs' tail gestures and vocabulary are generally well understood, at least on a basic level, and that utility robots can leverage this to unobtrusively broadcast various aspects of their state to people while they work. For example, a cleaning robot could keep its tail high on a full battery and slowly lower it to indicate less energy, it could lower its tail (between its wheels) to indicate a stuck wheel, or a delivery robot could wag its tail quickly to

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 403–419, 2013.

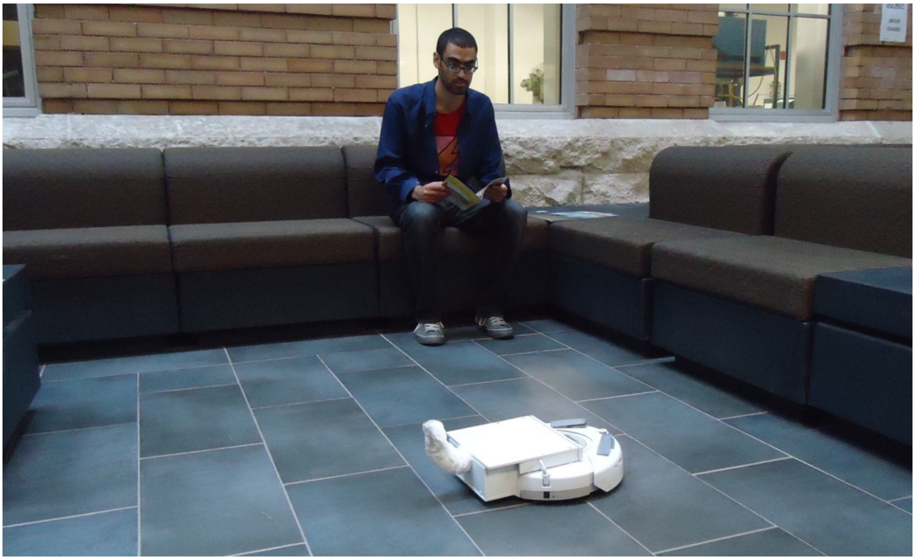


Fig. 1. A person notices the ambient tail state of a passing robot

indicate it found the person it was looking for. This tail interaction would further leverage people's existing tendencies to anthropomorphize and zoomorphize utility robots, for example, people give them names and buy them clothes [7, 30].

Affect is easily recognized by people (at least at the abstract level) and utility robots that find themselves collocated with people can use this communication channel. While a tail would enable a robot to broadcast zoomorphic affective information, it is not clear how this information will be interpreted by people in relation to its state; for one, a real dog uses its tail in concert with the rest of its body language, while our utility robot uses the tail in isolation. In addition to building a utility-robot dog-tail prototype, we conducted an experiment that explored how people respond to the tail in general, and how various tail parameters (e.g., wagging speed or height) impact the robot's perceived affective state. From this we summarize our findings into a toolkit set of guidelines for developing dog-tail interfaces to broadcast desired state information. The contributions of this paper are: a) a novel utility-robot dog-tail interface design and implementation, and b) a set of design guidelines, grounded in a study, for developing dog-tail behaviors for specific affective response.

2 Related Work

Part of the affective computing tradition in human-computer interaction is to incorporate human or animal-like affect and emotion directly into interfaces [20, 21], for example, a picture frame which uses an ambient color display to communicate emotion between people when they are apart [5]. There is a well-established application of ideas from affective computing to human-robot interaction, where impressions of robotic affect can be used to help users gain high-level state information without requiring them to read complex sensory information [6, 11, 12, 31]. Some have suggested the use of facial expressions and embodied gestures, where examples include mechanized faces with eyebrows, mouths, etc. [1–3, 34], animated faces on screens [16, 18], using mixed reality to superimpose graphics faces on robots [31], human-like whole gestures with arms, etc. [2], or even using gaze [29]. More abstract methods such as colors and sounds have similarly been used for communicating emotion to represent state. For example, both the Breakbot and AIBO robots use color in this way, and the Sony AIBO further uses puppy-like sounds [8, 23]. Our work follows this overall approach by using affect and emotion, via a robotic tail, to encourage attribution of affect for the purpose of conveying high-level robot state.

Animals have commonly been used for robotic interface inspiration. Leonardo, for example, was designed as a fantastical mammalian creature that people can relate to and communicate with [2], the Sony AIBO was designed explicitly to act as a puppy [8], and one robot can follow its owner on a leash similar to a dog [33]. Several robots have also used tails in concert with other features to help communicate with people, to convey emotions as part of their animal persona or design [8, 26]. In our work, we investigate how a dog-inspired tail interface can be applied to utility robots, and how people perceive the affective impact of its various motion possibilities.

Zoological research tells us that dogs can convey a broad range of states through their tails, for example, suggesting a happy state by wagging, high arousal or self-confidence by raising their tails, or fear by lowering their tail [1, 5]. Robotic pets such as the AIBO only use simple tail wagging [8], and it is still unclear how a wide range of behaviors can scale to robots. One study discovered that a tail can invoke memories and that interpretations vary with wag speed in one dimension [27], although this work did not aim to formally parameterize the tail motions and measure affect as we do, and also did not find consistent results across people. Further, robotic tails could produce motions not found with real dogs, such as moving in a full circular motion. In our work, we extend tail movement capabilities and formally investigate how people perceive a full range of motions in terms of affect.

3 Implementation

We based our dog-tail interface implementation (Fig. 2) on a technique used in hobbyist animatronics [28, 35] where the tail was constructed from a modified common construction toy kit (Klixx): the interlocking pieces were sanded to achieve smooth movement and to increase range of motion, and paperclips were inserted through drilled holes to strengthen the joints (Fig. 3). Tail deformation was achieved using a two-dimensional cable and heavy-duty servo pulley mechanism attached to a wooden board. The cables were attached to the tail by being threaded through the paper clips.

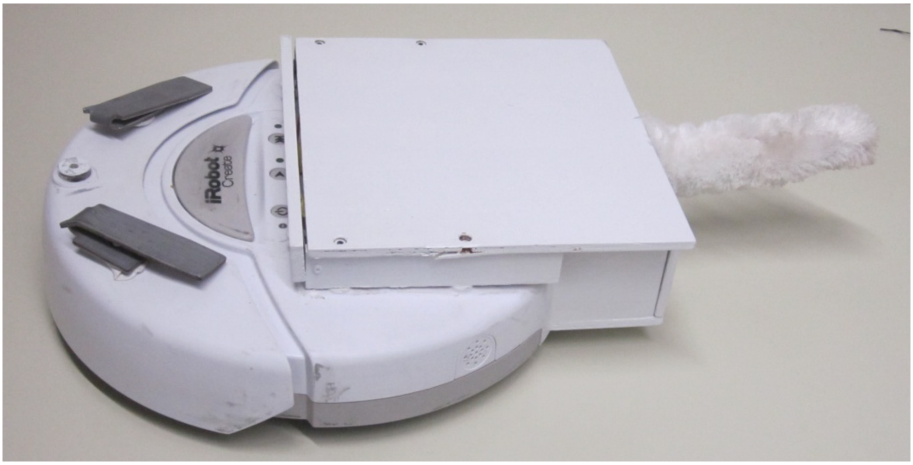


Fig. 2. An iRobot Create with our dog-tail attached

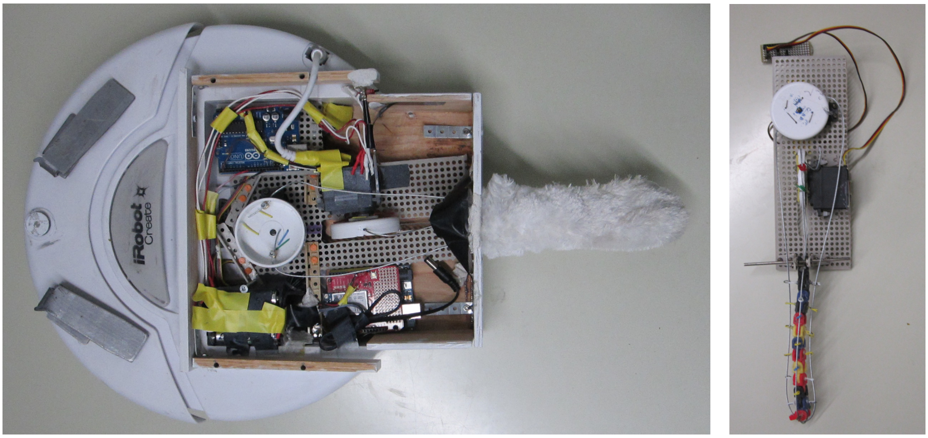


Fig. 3. Top-view of the tail assembly and before putting it inside the wooden box (right)

A key point with this mechanism is that we had to ensure that the two mounted servos were directly in line with the central tail axis, both in terms of height above the platform and lateral offset; slight alignment deviations added a tilt to the tail movements, making direct left-right or up-down movements difficult, and larger deviations resulted in the tail twisting (crumpling).

The entire tail mechanism was mounted on an iRobot Create disc-shaped robot, covered with a wooden box to hide the internal workings, and the tail itself was covered with white furry fleece (taken from a stuffed animal) to improve the appearance of being a dog tail.

The electronic implementation was achieved by using two Arduino Uno prototyping platforms, one to control low-level tail behavior and one to drive the robot around a space, and a WiFly wireless internet module for remote control.

The basic tail design was determined through informal pilot studies (5 participants). We experimented with tail coverings: no covering was seen as too mechanical or unpleasant, a white spandex (hose) cover was seen as somewhat reptilian and left a negative impression, and our fur covering was seen as cute and fun. Despite worries that fur would be out of place with a plastic robot, this was not a strong concern in the pilot. We tested tail length in terms of now natural it appears on our robot; we built short (10 cm), medium (15 cm), and long (32 cm) versions. The medium length was favored strongly in the pilot and used in our final version.

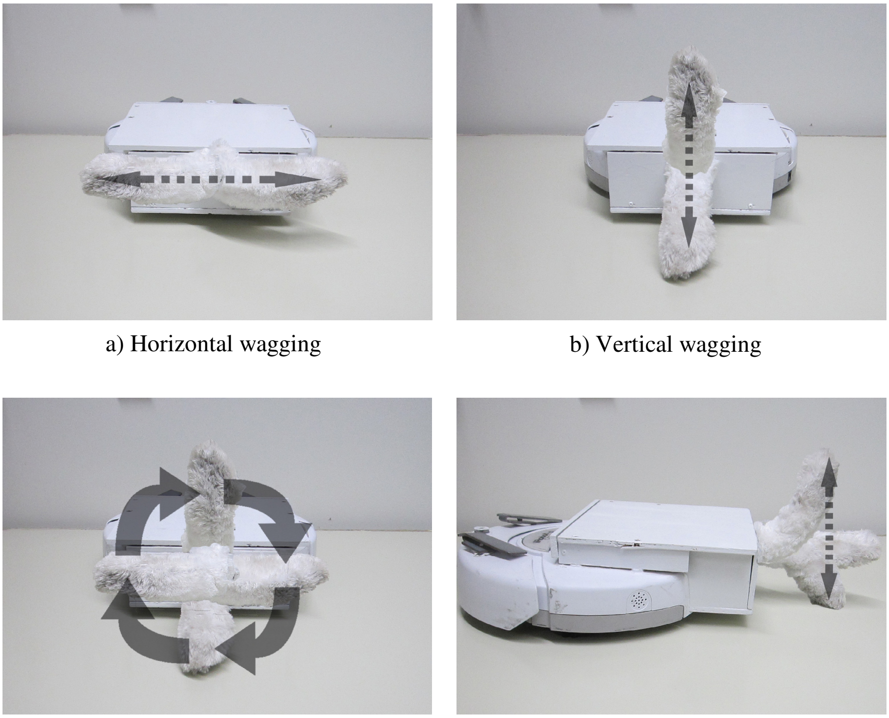
4 Exploratory Methodology

Zoological research tells us much about how dogs use their tail to communicate [9, 13], and we believe that people have some basic understanding of dog-tail communication from the general social stock of knowledge [10], for example, that wagging is a positive or playful state. However, we did not yet know if this knowledge transfers to interacting with a robot via a dog-tail. We also did not know to what extent people understand what a tail may be trying to communicate, or how different kinds of tail motions will be interpreted by people. As an initial step, through pilots we observed that people do not (at all) understand intricate dog-language specifics, e.g., different meanings when wagging on the left vs. the right. However, in the pilot people understood well that dog tails are used for communication, and could understand basic language such as wagging and raising or lowering the tail. Our exploratory goal is to leverage this general understanding (i.e., we do not want to require people to have training) and to systematically explore how people interpret a full range of tail motion possibilities: we implemented a broad range of tail motions and configurations, at different speeds, wag-sizes, and so forth, and conducted a study to evaluate how people perceive the resulting robotic affective states.

4.1 Measuring Perceived Affect

To classify people's perceptions of the robot we applied Russell's Circumplex Model of Affect, a standard psychological model of affective and emotional states [17, 24]. This model breaks affect into two scales or dimensions: valence (pleasure), from displeasure to pleasure, and arousal, from low to high energy.

To measure people's perceptions of the robot's affect we employed the Self-Assessment Manikin (SAM) [19]. SAM is a standard psychological instrument for rating affective states on the above affect model, where valence and arousal are represented by a series of easy to understand comic-like pictorial representations: from a very unhappy to a very happy character on the valence dimension, and from a sleepy low-energy to a high-energy awake character on the arousal dimension. People can rate an affective state simply by selecting the most appropriate pictorial on each dimension; in our case we used seven-point scales. Although generally used for a person to rate their own feelings, this method can also be used to rate the perceived state of others [22].



a) Horizontal wagging

b) Vertical wagging

c) Circular wagging

d) Side-view of robot, in this case vertical wagging

Fig. 4. The three continuous motion patterns used in our study

As part of our analysis, we correlate our observed valence and arousal with previous work on classifying emotions on these scales [19]. This provides a means to generate descriptive keywords that are capable of explaining how people may interpret the particular motion.

4.2 Exploring Tail Vocabulary

In developing a broad range of tail motions and states, we used nature-inspired possibilities (how real dogs act) as well as additional possibilities enabled by our tail configuration but not used in nature. We categorize our selected tail settings into: continuous *wagging*, the tail is always moving, *gestures*, the tail does an action at a certain time, and *postures*, the tail maintains a static state.

We developed three forms of wagging: horizontal, where the tail moves left and right on a plane roughly parallel to the floor similar to as in nature [9, 13] (Fig. 4a), vertical, where the tail moves up and down perpendicular to the floor (not found in nature, Fig. 4b), and circular, where the tail moves in a complete circle (not found in

Table 1. The full range of tail configurations we developed and tested

category	sub-type	attributes
continuous wagging	horizontal	speed: low, medium, high wag-size: small, medium, large height: low, parallel to floor, high
	vertical	speed: low, medium, high wag-size: small, medium, large
	circular	speed: low, medium, high
action gestures	raising	speed: low, medium, high height: low, high
	lowering	speed: low, medium, high height: low, high
static postures		height: low, medium, high

nature, Fig. 4c). All three types were created at low, medium, and high speeds. We varied horizontal and vertical to have either a small, medium, or large wag-size (how *tight* or *wide* the wag was, no change in speed). Finally, we varied horizontal wagging to have a low, parallel, or high offset in relation to ground.

We developed two tail gestures, a raising and a lowering action, to mimic how dogs act in nature: the tail was kept at a non-moving neutral state slightly below center (as with a real dog) except when it moved to complete a gesture. The tail would hold the gesture for 0.5 seconds before returning to the neutral state. We created low, medium, and high speed versions of the gestures, referring to the time taken to change from neutral to target state (raised or lowered), and a low and high offset version of each, representing how far the gesture moved from neutral. Finally, we had three static postures: a low tail, a straight tail parallel to the floor, and a high tail.

Table 1 presents an overview of our 31 motions deriving from the above configurations. Note that attributes are manipulated independently of others and thus some entries of the table are identical. For example, for the three wag sizes of horizontal wag, the other two attributes (speed, height) were kept fixed; for horizontal wag, the medium speed and the medium height settings were effectively identical. This reduction yielded 26 unique behaviors that were shown to participants.

4.3 Anticipated Interpretations

In general, we assume that high tail height will have high valence, and that valence values will decrease when the height is decreased, as this is naturally how dogs communicate with their tails [9, 13]. Additionally, based on previous motion and emotion work we expect that, in general, higher speeds will have higher arousal [25]. We expect this to happen for wagging, gestures, and postures.

5 Study

We recruited 20 participants from our local university population to participate in our study: 12 males / 8 females, aged 18-47 (M=24.25, SD=6.79). Our study was reviewed and approved by our university research ethics board, and all participants received $10 for their participation in the 60 minute study.

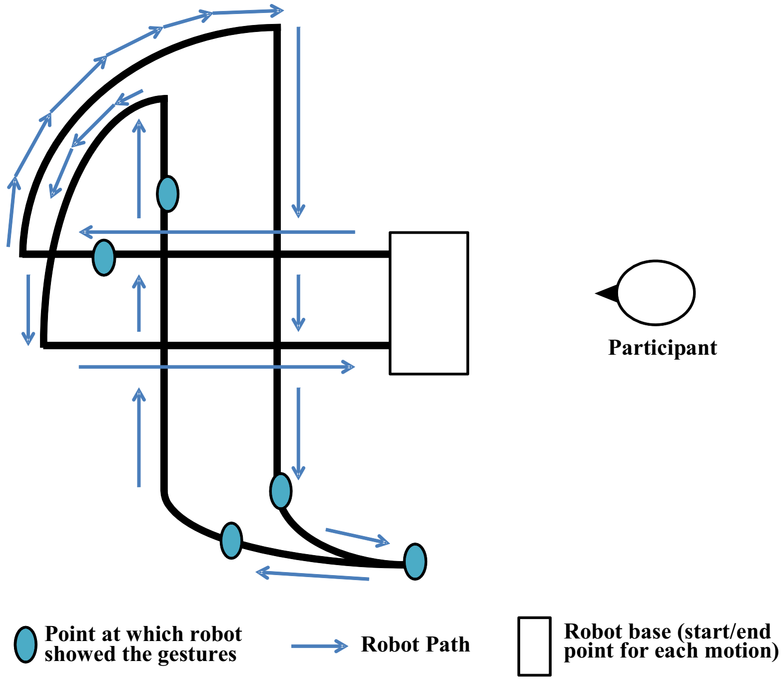


Fig. 5. Robot motion plan, blue ellipses define the points where robot showed action gestures

Participants were brought to our lab environment, and after a brief introduction, signed an informed consent form and received their compensation. We introduced the robot and the tail, the concept of the robot using the tail to communicate mood, and introduced the SAM scales based on its recommended text [15]. Participants proceeded to view the tail behaviors, order of appearance counterbalanced across participants. Participants were given 15 seconds post-demonstration of each tail behavior to rate the configuration on the SAM scales. Finally, we conducted a semi-structured interview, to investigate general views on the tail interaction, and debriefed the participants before ending the study. All studies and interviews were videotaped.

The layout of the study environment is shown in Fig. 5, where the participant was seated at a desk positioned to easily view the robot's motion, as it followed the path indicated. We designed this path to provide views of the tail from the front, sides and behind. The robot used the same path for all tail configurations, where the tail action was the only thing that changed. Blue ellipses on the robot path represent the spots where the robot showed the action gestures such as "raising the tail" which can only happen at certain points. We used side and back views of the robot for showing action gestures so as to provide a clear view of the tail to the participants; we did not have a view from the robot's front as the robot might have occluded the tail. Other than action gestures, all tail motions were programmed to initiate when the robot started to move and were stopped when robot came to a halt. The path took 35 seconds to complete, after which the experimenter returned the robot to its original position to minimize drift over the cases.

5.1 Results

We performed six primary analyses based on our configurations highlighted in Table 1 and our anticipated interpretations; our dependent variables were the participant ratings of affect on the valence and arousal dimensions.

Speed vs. Wag Type. We conducted a 2 way ANOVA on wag type (horizontal, vertical, circular) versus speed (low, medium, high). As the assumption of sphericity was violated for the main effect of speed on valence (Mauchly's test, X^2_2=14.93, p<.05), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (e =.631).

All effects are reported as significant at p < .05. There was a significant main effect of the wag speed on both valence $F_{1.26,24.30}$=9.79 and arousal $F_{2,36}$=71.38. Planned contrasts (we predicted that more speed would express more energy and more positive valence) revealed that, on both the valence and arousal dimensions, high speeds were rated significantly higher than medium speeds $F_{1,18}$=18.53 valence, $F_{1,18}$=42.92 arousal, and low speeds $F_{1,18}$=11.79, valence, $F_{1,18}$=99.42, arousal.

There was also a significant main effect of wag type on both valence $F_{2,36}$=15.52 and arousal $F_{2,36}$=39.63. Post-hoc tests (with Bonferroni correction) reveal that vertical wagging (M=-0.56, SD=1.49) was rated as lower valence than both horizontal (M=1.44, SD=0.98) and circular (M=0.98, SD=1.83), although there was no difference found between horizontal and circular. For arousal, all differences were significant: horizontal wagging (M=0.63, SD=1.3), vertical (M=-0.56, SD=1.45), and circular (M=1.61, SD=0.97). These relationships are shown in Fig. 6.

There was a significant interaction effect between the wag type and speed on valence $F_{4,72}$=3.74 and arousal $F_{4,72}$=3.02, indicating that speed's effects on perceptions of valence and arousal depends on the wag type. For valence, post-hoc tests (with Bonferroni correction) revealed that all three speeds yielded different results for horizontal wag, but no significant effects were found for vertical or circular wags, as suggested by Fig. 6. For arousal, post-hoc tests (with Bonferroni correction) reveal that speed is a significant predictor of measured arousal for horizontal and vertical wagging, but for circular wagging, low speed is significantly different from medium and high, which themselves are not different.

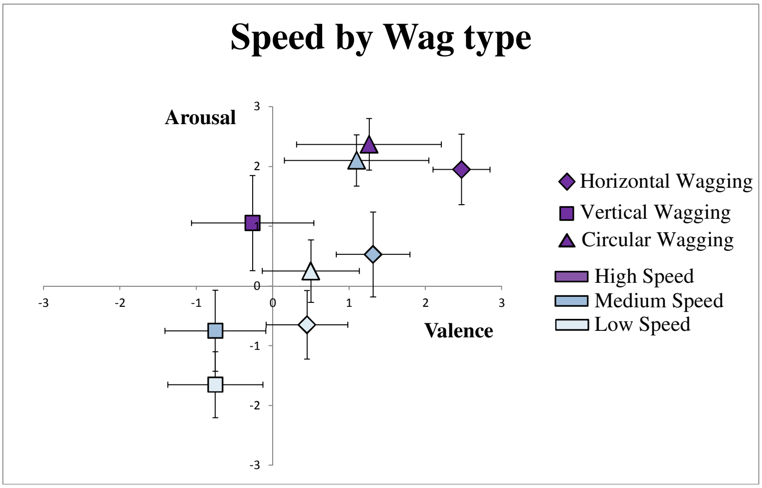


Fig. 6. Speed by Wag Type, error bars show 95% confidence interval

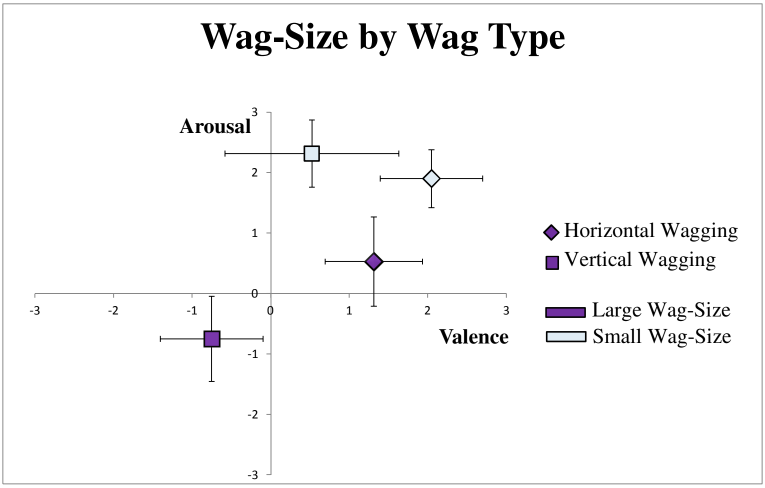


Fig. 7. Wag-Size by Wag Type, error bars show 95% confidence interval

Wag-Size vs. Wag Type. We conducted a 2-way ANOVA on wag-size (small and large) versus wag type (horizontal, vertical); all effects reported significant at $p<.05$. There was a significant main effect of wag-size on both valence $F_{1,17}=7.77$ and arousal $F_{1,17}=48.39$, showing that smaller wag-size increases perception of both valence and arousal (Fig. 7). A significant interaction effect was found between the wag type and how different wag-size were perceived for arousal $F_{1,17}=12.19$ indicating that small wag-size is significantly different from large wag-size in the arousal dimension.

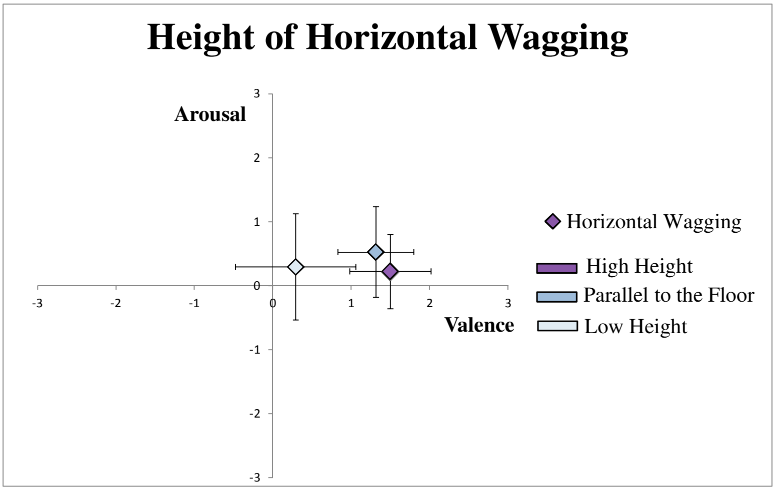


Fig. 8. Height of Horizontal Wagging, errors bars show 95% confidence interval

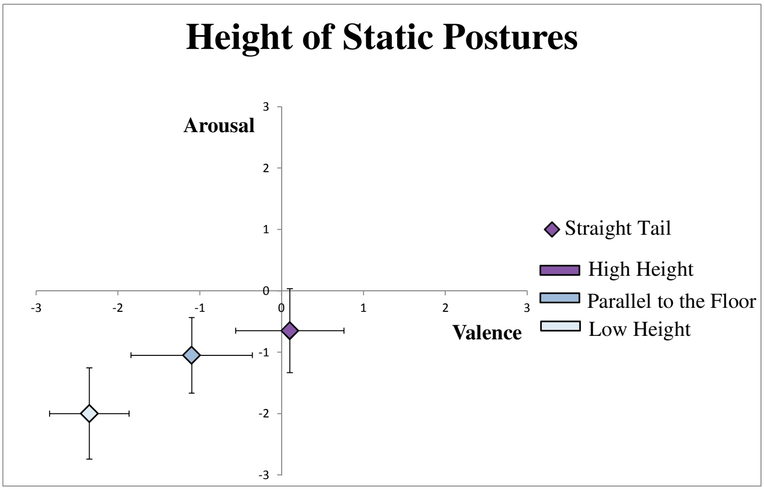


Fig. 9. Height of Static Postures, error bars show 95% confidence interval

Height of Horizontal Wagging. We conducted a 1-way ANOVA on height (low, parallel to floor, high) with horizontal wag type (3 levels of wagging); all effects reported significant at $p<.05$. There was a significant main effect of height on valence $F_{2,32}=6.601$, with planned contrasts highlighting that both medium $F_{1,16}=4.69$ and high height $F_{1,16}=12.48$ were higher valence than low height. This indicates that high height had more valence than low height (Fig. 8). There was no effect on arousal.

Height of Static Postures. We conducted a 1-way ANOVA on height (low, parallel to floor, high) with static postures (3 levels of straight tail); all effects are reported significant at $p<.05$. There was a significant main effect of the height on perceived valence $F_{2,38}=21.4$ and arousal $F_{2,38}=6.36$. Planned contrasts for valence showed that low height (M=-2.35, SD=1.04) was lower rated than both medium (M=-1.1, SD=1.59) and high (M=-0.1, SD=1.41), and for arousal low height (M=-2.00, SD=1.59) was lower rated than high (M=-0.65, SD=1.46) (other contrasts non-significant). This explains that high height had more arousal and more valence as compared to low height (Fig. 9).

Non-significant Tests. No significant effects were found using ANOVAs on speed (low, medium, high) by action gestures, or height (low, parallel to the floor, high) by action gestures.

Semi-structured Interview. From the post-study semi-structured interview, we found that: 17 participants (85%) zoomorphized the robot, for example, saying "it looks like an animal," "it felt like a dog." Additionally 2 female participants (25% of females) asked the name and gender of the robot and 6 of them (75% of females) discussed its "cuteness."

19 participants (95%) responded positively when asked if they found the dog-tail interface easy to understand and read, saying such things as "I am able to perceive its feelings," "it was easy to understand feelings of the robot," (we note that participants were introduced to the gestures as "feelings" through the explanation of the SAM affective measurement instrument). Some, however, suggested that we add other dog elements, such as puppy sounds to improve the communication clarity. Many mentioned that they were also interested in seeing the dog-tail interface on other utility robots, and some (2 participants) were interested in seeing a cat-tail version.

5.2 Discussion

We observed that people readily accepted a dog-tail interface on a utility robot, easily understood the concept of the robot communicating through the tail, and that the tail interface has a broad and detailed communication vocabulary that people can consistently understand. We observed that basic dog-tail language such as higher tail height is understood, and also that higher tail wagging speeds result in perceptions of higher arousal, and in general, also result in more positive perceived valence. This was also echoed in the static tail postures, although with no movement, results were all generally less aroused and less positive than their moving counterparts.

We further found consistent differences in wagging types. While horizontal wagging was generally perceived as positive valence, vertical wagging was seen as being more negative – even with faster wags – and circular was somewhere in between with less clear results on valence. Thus, different wag types can be used depending on what a robot is trying to communicate. We feel that this inclusion of non-natural motions did not hinder our results as designers are free to stick to natural ones, and our statistically-significant results (including non-natural motions) indicate that there is a base-line common interpretation between people that can be used in design.

Upon consideration of our lack of consistent findings for our action gestures, we realized that the robotic motion itself and the neutral tail state held when a gesture was not being performed was a likely confound, where people perhaps rated those constant elements instead of the periodic gesture. This is supported by the fact that the perceived valence and arousal of all action gesture movements were tightly grouped.

One perhaps unexpected result was that smaller wag-sizes result in perceptions of higher valence and arousal; we expected these motions to have lower results given their lower movement profiles. Upon consideration, however, we realized that smaller wag-sizes at the same tail speed will result in more wags per second, perhaps increasing the perception of speed.

Overall, our results show that people were able to understand affective robotic states as conveyed using a tail, and as such this technique could be used as a peripheral-awareness channel for conveying high-level robotic state to people. For example, energetic vs. fatigued tail motions could be used to show battery level (e.g., fast/slow tail wag), or a robot could appear depressed (low-arousal/valence, e.g., slow-moving low tail posture) to show navigational confusion such as being lost. By communicating these abstract states, a utility robot can indicate its present state using people's existing knowledge of dog-tail movement and help them understand when and how they should interact with the robot.

6 Design Guidelines

From our results we formed the following tail guidelines to help designers in creating their own dog-tail interfaces that convey affect. We present our guidelines from three perspectives: motion parameters (speed, wag-size and height), wag-type (horizontal, vertical and circular wagging) and postures (straight tail-high and straight tail-low). In addition, we present the results from an exercise where we correlated our average arousal and valence ratings to existing knowledge of affective states on these continuums. We envision that this section will be useful to those creating dog-tail interfaces, aiding them in selecting tail movements to represent a desired affective state that can be understood by people without having to undergo training.

6.1 Dog-Tail Motion Parameters

Speed. A higher speed projects a higher valence and arousal (e.g., elated) and a lower speed projects a lower valence and a lower arousal (e.g., uninterested).

Wag-Size. A smaller wag-size projects a more positive arousal (e.g., energetic) and a larger wag-size projects a less arousal (e.g., lazier).

Height. A higher tail projects a more positive valence (e.g., happier), and lower tail a more negative valence (e.g., sadder).

6.2 Dog-Tail Wag-Types

Horizontal Wagging. This is the natural form of wagging, as found in dogs. This type of wagging can convey a range of valence and arousal values, starting from medium to high.

Circular Wagging. A tail wagging in circular motion may be able to project a more positive arousal as compared to horizontal and vertical wagging at the same speeds.

Vertical Wagging. A tail wagging in vertical motion generally projects a more negative valence and a slightly more negative arousal as compared to horizontal and circular wagging, although medium high arousal states can be achieved with high speeds or small wag sizes.

Table 2. Adjectives matching participant ratings of tail motion

category	sub-type	parameter	attributes and descriptive keywords
continuous wagging	horizontal	speed	low - modest medium - wondering high - joyful and elated
		wag-size	small - strong, mighty and powerful large - interested
		height	low - contempt parallel to floor - awed high - wonder
	vertical	speed	low - solemn medium - shy and disdainful high - aggressive
		wag-size	small - aggressive large - selfish and quietly indignant
	circular	speed	low - reverent medium - aggressive and astonished high - overwhelmed
action gestures	raising	speed	low, medium and high - shy, selfish, disdainful or weary
		height	low and high - shy, selfish, disdainful, weary timid and fatigued
	lowering	speed	low, medium and high - shy, selfish, disdainful or weary
		height	low and high - shy, selfish, disdainful, weary timid and fatigued
static postures		height	height: low - lonely parallel to floor - fatigued high - concentrating

6.3 Static Dog-Tail Postures

Static dog-tail postures provide more subdued impressions of affect and valence than the moving counterparts. A low, static tail projects a very low valence and arousal, while a higher tail makes this impression more moderate.

6.4 Correlating Tail Motions to Affective Adjectives

We correlated our results to existing work that maps data points on the arousal-valence space to affective adjectives [18], as a means of generating loose-yet-informative keywords to roughly describe how various tail configurations may be perceived. We took the average rating for each motion and correlated it with the closest point on the previous work. A summary is given in Table 2.

7 Future Work

Our work is limited to giving a robot a dog-tail only for communicating different affective states. In contrast, a real dog uses its face, eyes, voice body language, etc., to accompany its tail motions to create more complex expressions for deeper communication. We hope to continue our line of work to investigate what other aspects of dog communication can be used by robots in similar ways, or even which other animals can serve as inspiration for developing this type of interface.

While the aim of this study was to develop an understanding of how a robot may communicate using a dog tail, and how people may perceive the communication, moving forward it will be important to further develop our guidelines to provide researchers with more concrete tools for tail-interface design. For example, although our results and guidelines help designers decide how to communicate a desired affective state, we do not yet address how to move from low-level robotic state (e.g., battery level, malfunction, etc.) to affective ones. While this is a broad question for HRI in general, we believe that we can follow the dog metaphor as one promising direction for developing this kind of mapping.

Currently, we have only placed our dog tail on a small robot that sits close to the ground (similar to a small dog). We will explore how our tail will translate to other morphologies such as a humanoid robot or flying robot, and other domains such as toy robotics. Part of this question will be to explore the limits of use. For example, while we focused currently on utility robots, we will explore other less obvious applications such as inanimate objects (e.g., a printer) to help convey the devices' state, and will consider where the tail interface may not be applicable, for example, for remote control robotics or industrial machines where the tail may not be monitored.

8 Conclusion

In this paper we presented an original dog-tail interface and conducted a formal evaluation to investigate how people perceived the affective states of a robot equipped with a dog tail, across a full range of tail behaviors. We found that the tail was able to convey a broad range of affective states and that people reliably interpreted the tail motions in a consistent fashion. From this, we summarized our results into design guidelines for creating dog-tail interfaces.

Overall, we anticipate that our contribution of exploring and mapping how robots can use dog tails to communicate affect will be of use to HRI designers, providing them with a new paradigm for robotic communication. Further, we hope that robots using this kind of periphery communication will help people in understanding their state and help in deciding when and how they should interact with the robot.

Acknowledgement. We are thankful to Dr. Jacky Baltes and Andrew Winton for their valuable time and input. We would also like to thank B and K Winton Industry for lending their tools to design our interface. Finally, we would like to express our gratitude to Derek Cormier, Masayuki Nakane and Ibrahim Shahin for their support in the study.

References

1. Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies 59(1-2), 119–155 (2003)
2. Breazeal, C., Hoffman, G., Lockerd, A.: Teaching and Working with Robots as a Collaboration. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems - AAMAS 2004, pp. 1030–1037 (2004)
3. Van Breemen, A., Yan, X., Meerbeek, B.: iCat. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems - AAMAS 2005, p. 143 (2005)
4. Brown, S.E.: Self Psychology and the Human–Animal Bond: An Overview. The Human-Animal Bond and Self Psychology: Toward a New Understanding 12(1), 67–86 (2004)
5. Chang, A., Resner, B., Koerner, B., Wang, X.: Ishii. H.: LumiTouch. In: Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems - CHI 2001, p. 313 (2001)
6. Duffy, B.R.: Anthropomorphism and the Social Robot. Robotics and Autonomous Systems 42(3-4), 177–190 (2003)
7. Forlizzi, J., DiSalvo, C.: Service Robots in the Domestic Environment. In: Proceedings of the ACM International Conference on Human-Robot Interaction - HRI 2006, p. 258 (2006)
8. Friedman, B., Kahn, P., Hagman, J.: Hardware companions?: What Online AIBO Discussion Forums Reveal About the Human-Robotic Relationship. In: Proceedings of the Conference on Human Factors in Computing Systems - CHI 2003, p. 273 (2003)
9. Galac, S., Knol, B.W.: Fear-Motivated Aggression in Dogs: Patient Characteristics, Diagnosis and Therapy. Animal Welfare 6(1), 9–15 (1997)
10. Hamill, L., Harper, R.: Talking Intelligence: a Historical and Conceptual Exploration of Speech-Based Human-Machine Interaction in Smart Homes. In: Proceedings of the International Symposium on Intelligent Environments, pp. 121–128 (2006)
11. Hyunsoo, S., Young-Min, K., Jong-Chan, P., Chong, H.K., Dong-Soo, K.: Design of a Robot Head for Emotional Expression: EEEX. In: International Symposium on Robot and Human Interactive Communication. RO-MAN, pp. 207–212 (2008)
12. Kang, S., Jaehong, K., Joochan, S.: A Robot Platform for Children and Its Action Control. In: International Conference on Control, Automation and Systems, pp. 2095–2098 (2008)
13. Kleiman, D.G.: Social Behavior of the Maned Wolf (Chrysocyon brachyurus) and Bush Dog (Speothos venaticus): A Study in Contrast. Journal of Mammalogy 53(4), 791–806 (1972)
14. Kriglstein, S., Wallner, G.: HOMIE. In: Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems - CHI 2005, p. 2094 (2005)

15. Lang, P., Bradley, B., Cuthbert, B.: International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-8 (2008)
16. Lee, M.K., Makatchev, M.: How do People Talk with a Robot. In: Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA 2009. p. 3769 (2009)
17. Mehrabian, A.: Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. Current Psychology 14(4), 261–292 (1996)
18. Michalowski, M.P., Šabanović, S., DiSalvo, C., Busquets, D., Hiatt, L.M., Melchior, N.A., Simmons, R.: Socially Distributed Perception: GRACE plays social tag at AAAI 2005. Autonomous Robots 22(4), 385–397 (2006)
19. Morris, J.D.: Observations: SAM: The Self-Assessment Manikin an Efficient Cross-Cultural Measurement of Emotional Response. Journal of Advertising Research 0021-8499, 63–68 (1995)
20. Norman, D.A.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books (2004)
21. Picard, R.W.: Affective computing. The MIT Press, Cambridge (1997)
22. Pollick, F.E., Paterson, H.M., Bruderlin, A., Sanford, A.J.: Perceiving Affect from Arm Movement. Cognition 82(2), B51–B61 (2001)
23. Reeder, S., Kelly, L., Kechavarzi, B., Šabanović, S.: Breakbot. In: Proceedings of the ACM Conference on Designing Interactive Systems - DIS 2010, p. 61 (2010)
24. Russell, J.A.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
25. Saerbeck, M., Bartneck, C.: Perception of Affect Elicited by Robot Motion. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction - HRI 2010, pp. 53–60 (2010)
26. Schreiner, K.: Retail Robots on the Move. Intelligent Systems 16(2), 4–7 (2001)
27. Shibata, T., Yoshida, M., Yamato, J.: Artificial Emotional Creature for Human-Machine Interaction. In: IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, vol. 3, pp. 2269–2274 (1997)
28. Singh, A., Young, J.E.: Animal-Inspired Human-Robot Interaction: A Robotic Tail for Communicating State. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction - HRI 2012, pp. 237–238 (2012)
29. Staudte, M., Crocker, M.W.: The Utility of Gaze in Human-Robot Interaction. Metrics for Human-Robot Interaction (2008)
30. Wolftronix.com, `http://www.wolftronix.com/tail/spine`
31. Young, J.E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H.I., Grinter, R.E.: Evaluating Human-Robot Interaction. International Journal of Social Robotics 3(1), 53–67 (2010)
32. Young, J.E., Xin, E., Sharlin, E.: Robot Expressionism Through Cartooning. In: Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction - HRI 2007, p. 309 (2007)
33. Young, J.E., Hawkins, E., Sharlin, E., Igarashi, T.: Toward Acceptable Domestic Robots: Applying Insights from Social Psychology. International Journal of Social Robotics 1(1), 95–108 (2008)
34. Young, J.E., Kamiyama, Y., Reichenbach, J., Igarashi, T., Sharlin, E.: How to Walk a Robot: A Dog-Leash Human-Robot Interface. In: International Symposium on Robot and Human Interactive Communication, RO-MAN, pp. 376–382 (2011)
35. Zecca, M., Endo, N., Momoki, S., Itoh, K., Takanishi, A.: Design of the Humanoid Robot KOBIAN - Preliminary Analysis of Facial and Whole Body Emotion Expression capabilities. In: International Conference on Humanoid Robots, pp. 487–492 (2008)

Pebbles: User-Configurable Device Network for Robot Navigation

Kentaro Ishii[1,2], Haipeng Mi[1,2], Lei Ma[1,2], Natsuda Laokulrat[1,2], Masahiko Inami[3,2], and Takeo Igarashi[1,2]

[1] The University of Tokyo, Tokyo, Japan
[2] JST, ERATO, IGARASHI Design Interface Project, Tokyo, Japan
[3] Keio University, Yokohama, Japan
kenta@sakamura-lab.org, mi@iii.u-tokyo.ac.jp, malei@satolab.itc.u-tokyo.ac.jp, natsuda@logos.t.u-tokyo.ac.jp, inami@inami.info, takeo@acm.org

Abstract. This study proposes devices suitable for use by non-experts to design robot navigation routes. The user places landmarks, called pebbles, on the floor to tell navigation routes to a robot. Using infrared communication, the pebbles automatically generate navigation routes. The system is designed such that non-expert users can understand the system status to configure the user's target environment without expert assistance. During deployment, the system provides LED and voice feedback. The user can confirm that the devices are appropriately placed for the construction of a desired navigation network. In addition, because there is a device at each destination, our method can name locations by associating a device ID with a particular name. A user study showed that non-expert users were able to understand device usage and construct robot navigation routes.

Keywords: Robot Navigation, Tangible User Interface, Navigation Landmark, Non-Expert User.

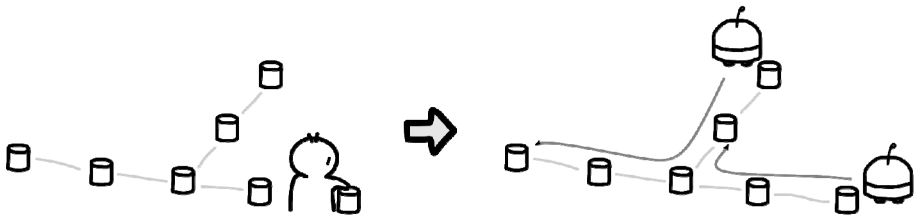


Fig. 1. Pebbles Usage. The user places pebbles at desired locations. Pebbles automatically construct a network topology. After deployment, the robot can travel to any pebble in the network.

1 Introduction

Robots can be used for various applications in both home and office environments. For example, a robot can travel to predetermined locations at a specific time to collect environmental data. Robots can also deliver a physical object, such as a cup of coffee

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 420–436, 2013.

to a colleague, upon a user's request. If a robot moves an electric fan closer to a person, the fan can work more energy efficiently. However, it is difficult for a robot to know how to reach a specific target location in an unknown environment. Meanwhile, regarding user interaction, it is preferable to specify target locations using natural expressions such as names of the locations. In this study, we propose a method that allows a non-expert end-user to provide navigation information to a robot and to naturally specify locations.

While there are many ways to characterize existing methods to configure an environment for robot navigation, we divide the existing methods into two approaches: automatic mapping by robots and environmental configuration by humans. Automatic mapping by robots, represented by simultaneous localization and mapping (SLAM) [1], is an effort saving method from the user's point of view. The robot acquires map information by automatically exploring the environment, and the robot can plan navigation for a specific location after constructing a map. However, the internal representation of the constructed map is typically a set of coordinates, which may be too complicated for non-expert users to associate with a specific named location, such as "kitchen." We call this name association "labeling." In addition, in such a system, the robot may enter into areas where the user does not want the robot to enter.

On the other hand, there are several implementations that fall into the category of environmental configuration by humans. Physical guides, such as rails for trolley robots and lines for line following robots are the simplest ways to realize such environmental navigation support. A robot can reach target locations by tracing these environmental guides. Most other techniques employ absolute positioning because odometer-based positioning has the fundamental limitation of accumulating positioning errors. Some directly measure a robot's position using environmental sensors [2-3], while some other employ robot-readable devices or tags that work as landmarks that allow the robot to determine its position in the environment [4-5]. For a robot to navigate in these environments, the map of the entire environment is required and is typically provided by the person who manually calibrates and installs sensors, devices, or tags. Combining this manual registration of map information with manual labeling of locations enables a user to naturally ask the robot to reach a specific location. However, existing methods require experts to deploy devices, and therefore the home or office user needs expert assistance to configure locations and labels.

We propose a method that allows non-expert users to configure the environment for robot navigation employing infrared communication devices called pebbles. The robot receives topology information from these devices and moves to locations by tracing infrared signals. We extend previously demonstrated implementation [6], and discuss usage scenario and deployment assistance. We also conduct a user study to see if non-expert users can understand device usage. Fig. 1 illustrates the basic concept of the proposed method. The user only has to place pebbles at desired locations, and the pebbles will automatically construct a network topology. After deployment, the robot can move to any pebble location in this network. Since infrared signals travel in a straight line, if one pebble has unobstructed line-of-sight to another pebble, the robot can be expected to travel between these two pebbles. With our method, destinations are determined by pebbles, and thus it is easy to label destinations by associating a pebble device ID with a name. In our current implementation, the user can easily do this by putting physical labels on buttons on a remote control, for example, “kitchen” or “entrance” (Fig. 5). The user can select a labeled button for the target location. In addition, our method explicitly prevents the robot entering into the area where the user does not want the robot to do, because the robot moves by tracing the constructed network trails. Although our method requires the user to place each pebble so that it is within line-of-sight of neighboring pebbles, each pebble can provide LED feedback, and the robot can provide voice feedback to confirm connection status. None of these configuration processes require expert knowledge.

Our contribution is presenting a user-configurable method for robot navigation, rather than the system’s algorithms, communication technology, or network construction. LED and voice feedback helps the user to construct an appropriate network, and we describe the interaction design between the user and the devices.

2 Related Work

Several positioning techniques have been proposed for indoor environment, where GPS technology provides only limited performance. Radio-based fingerprinting [7] is one of the trends in indoor positioning; in most cases, the system does not require major changes to infrastructure because of the recent widespread presence of wireless LAN [8]. Some other approaches calculate positions by installing sensors in the environment [9-13], including a commercial system [13]. Another popular approach is to utilize cameras fixed in the environment [14]. Other approaches use passive tags to calculate position. Saito et al. implemented an indoor marker-based localization system using coded seamless patterns [15]. Similar indoor localization systems have been subsequently proposed using invisible markers installed on the ceiling [16]. These absolute positioning techniques can be used to determine robot position, but navigating in the environment also requires environmental knowledge, in particular, map information. Typically, registering such map information requires expert knowledge and is not easy to reorganize after registration.

On the other hand, relative positioning by devices emitting beacon signals is another method of robot navigation. Similar to physical guides such as rails and lines, the robot can trace signals to move among the devices. These devices also work to understand environmental structure by detecting if the beacon signals reach others as expected. There have been some work where a robotic system uses such beacon devices to determine the environment [17-20], and we share the basic idea of this previous work. However, the previous work does not discuss if an average user can deploy such devices, which is our focus here. In contrast, a robot places the devices [17], or robots themselves are the devices [18-20].

The automatic mapping techniques represented by SLAM are widely used for robots and autonomous vehicles [1]. Various sensors are used for observing an environment, such as sonar [1,21], laser [22], and vision [23-25]. The robot constructs a map of the environment by exploration, and the robot can use the constructed map to navigate toward a specific location in the map. However, a typical internal representation of the map is a set of coordinates, and it is not trivial for a user to specify a destination or label a location within it. In general, this method requires another representation or visualization of the constructed map so that the user may specify a point in the map. This paper deals with the issue of providing a user-understandable representation, as described in the next paragraph.

In the proposed method, tangible user interface [26] is a key concept for providing a user-understandable representation of the environmental map. Placing a tangible device at each destination bridges the specific location in the user's physical world and the specific node in the system's internal topology map. In our prototype system, the internal representation of the environmental map is a network topology, but the user does not need to know about it. The user only has to specify and label a device at desired locations in the physical world, and the system determines the steps to reach the devices from the network topology.

3 Pebbles

3.1 Device Hardware

A pebble is a self-contained device that is capable of communicating with other pebbles. Pebbles construct a network topology to generate navigation trails for the robot. Each pebble consists of a microcontroller unit, infrared communication unit, and battery set. All components are packed in a plastic cylinder case 10 cm in diameter and 10 cm in height (Fig. 2). Each pebble has a unique ID that distinguishes it within the network. A button labeled with this ID sits on the top for user input.

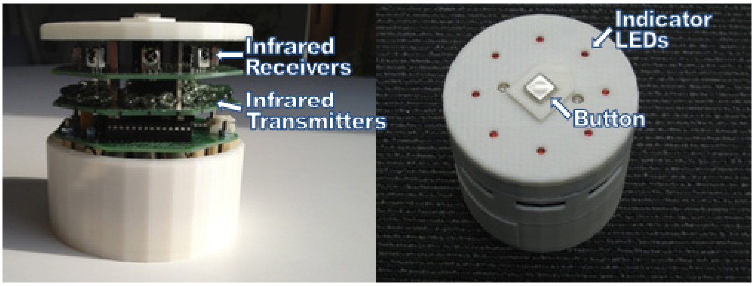


Fig. 2. Pebble Hardware. A pebble has 24 infrared transmitters, 8 receivers, 8 indicator LEDs, and a button. The infrared transmitters and receivers are arranged to cover all directions. The indicator LEDs are used to provide primitive feedback to the user. The button is used to receive user input.

The infrared communication unit has eight infrared transceiver modules; each one contains a red LED indicator, infrared receiver, and three infrared signal transmitters. The modules are placed 45 degrees apart so that all eight modules achieve 360-degree coverage. Infrared signals are modulated using a 38 kHz carrier, a frequency typically used for remote-controlled appliances, to prevent errors caused by noise. The indicator LEDs on the top provide primitive feedback.

3.2 Deploying Pebbles

The robot plans its navigation based on topological information from the pebbles and moves toward a goal according to pebble signals. Therefore, each pebble must communicate with at least one other pebble, and no pebble should be completely isolated. To arrange for each pebble to communicate with its neighbors is the user's responsibility. The system provides LED and voice feedback to help the user to ensure pebble connectivity.

The user can see LEDs on the top of each pebble, each of which indicates its connection status to another pebble in a specific direction (Fig. 3). When transmitting and receiving infrared signals, the indicator LEDs of corresponding channels synchronously blink to indicate adjacent relationships. The user can see if a pebble is visible to its neighbor pebbles. In this way, the user knows if two pebbles are placed too far apart or the path directly between them is blocked by some obstacles. The user can then correct the problem.

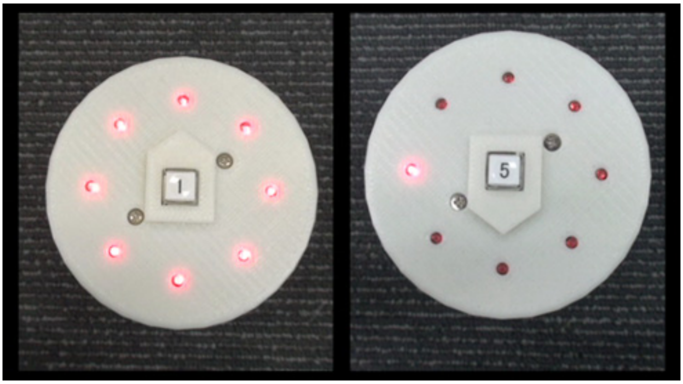


Fig. 3. LED Feedback. When transmitting and receiving infrared signals, the indicator LEDs of corresponding channels synchronously blink to indicate adjacent relationships to the user. The LEDs of the transmitting pebble (left) indicate the transmission of the signal to all directions. The LEDs of the receiving pebble (right) indicate the reception of the signal in the direction of the transmitting pebble.

In addition to the LED feedback, the system also provides voice feedback. The robot enumerates the IDs of pebbles in the network, for example, "I can locate number one, number three, and number five." The user then knows which pebbles are not visible to the robot and can identify which connection is not working as expected. If the robot does not receive any signals from pebbles, it says, "I cannot see any pebble," to indicate that the user should rearrange it.

Utilizing LED and voice feedback, the user places pebbles at desired goals and their transit points. If the user places pebbles as in Fig. 4, for example, the robot can go to all rooms except the top middle room. Since only routes connecting two arbitrary neighbor pebbles are considered possible trails, the user can explicitly keep the robot from entering the top middle room.

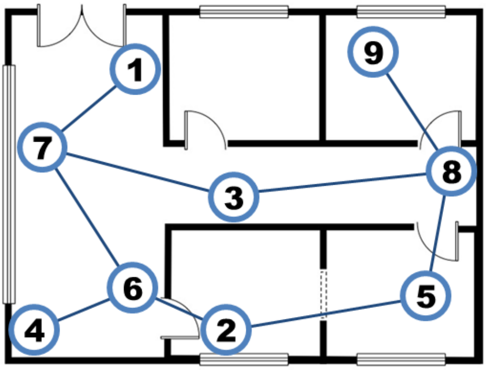


Fig. 4. Sample Deployment. The robot can go to all rooms except the top middle room. In this case, the user explicitly keeps the robot from entering the top middle room.

When the user changes the layout of the rooms or introduces a new object that obstructs a route of the robot, the user has to rearrange the pebbles near the change. Pebbles periodically transmit signals and automatically detect topology changes. Therefore, the user can reconfigure the network by adjusting pebbles, confirming that a pebble communicates with its neighbors as it did initially. This is useful, for example, when the user places an obstacle, such as a piece of luggage, in the room temporarily and wants to specify a path that avoids it.

Locations are labeled by attaching physical labels on the buttons of the remote control (Fig. 5). After labeling, the user can select the target location of the robot by the name of the label. We consider three kinds of labeling: static, semi-static, and dynamic targets. Static targets are the places whose name will not be changed after initial

Fig. 5. Labeling by Physical Labels. The user attaches physical labels on a remote control to name locations. The labels of the targets can be static like “kitchen” and “entrance,” semi-static like “sofa,” or dynamic like “father”.

installation, such as “kitchen,” “entrance,” and “meeting room.” Semi-static targets are usually fixed locations, but might be relocated occasionally, such as “dining table,” “sofa,” and “the desk of Ms. XXX.” Because the user might make a change to the layout of a room in home or office, the location associated with the table, sofa, or desk may be changed after the installation. The user can easily reconfigure the navigation network to fit the new layout by relocating the associated pebble. The user can also carry a pebble around to define dynamic targets, such as “father” and “Mr. YYY.” For instance, users could carry a pebble with them to continuously indicate their current locations. When one user wants to send a robot to another user, the sending user can simply specify the goal by pressing the button associated with the receiving user. Note that the sending user does not need to know where the receiving user is in this scenario. Pebbles can deal with all three scenarios in the same manner. In particular, the ability for non-experts to deal with semi-static and dynamic targets is the feature that differentiates our method from existing methods.

3.3 Specifying a Goal

After the navigation routes have been appropriately designed, the user can instruct the robot to move to any pebble within the network. A basic way to specify a goal is by pressing a button on the remote control (Fig. 6 (left)). The application services of the robot will request the user to specify a goal when required to proceed with the task. The user can also specify a goal by pressing the button on the pebble (Fig. 6 (right)). This is useful when the user want to call the robot or describe a series of target locations by traveling to the locations. If the button of the pebble is pressed, LEDs on the pebble blink more slowly than when transmitting and receiving the signals, which is described Section 3.2. The robot will reset the button press state after the task is completed, or the user can cancel the button press state by pressing the button again.

Fig. 6. Two Ways to Specify a Goal. The user can specify a goal of the robot by pressing a button on the remote control attached to the robot (left). The user can also specify a goal by pressing the button of the pebble (right).

3.4 Robot Navigation

The robot navigates by tracing infrared signals from pebbles. Therefore, the robot has the same hardware as a pebble for receiving infrared signals (Fig. 7). Our prototype employs the robotic platform iRobot Create, the developer's version of iRobot Roomba vacuum cleaner. The robot has a sound speaker for voice feedback and a roof rack for delivering objects (Fig. 7).

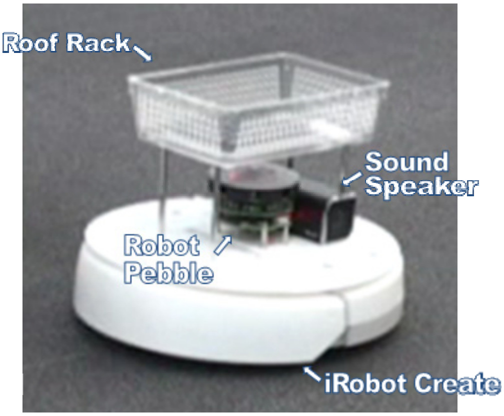
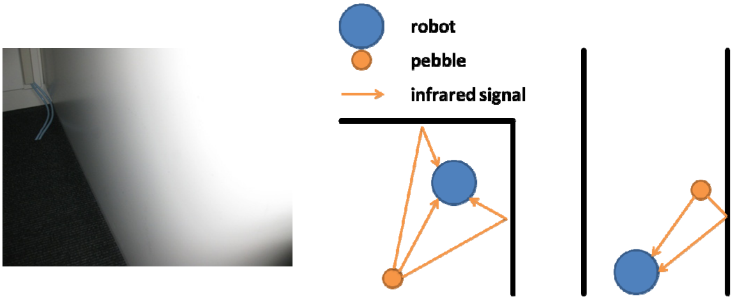


Fig. 7. Robot Platform. The robot has the same hardware as a pebble for receiving infrared signals. The robot also has a sound speaker for voice feedback and a roof rack for delivering objects.

The closer a neighbor pebble is placed, the more receivers out of eight receive signals. Using this characteristic, the robot roughly estimates its proximity to the target pebble by monitoring the number of receivers receiving signals from the target pebble. The robot has a bumper switch to detect collision and will change its direction if it detects obstacles, including pebbles. Because the bumper switch is weaker than a pebble's static friction, the robot does not push the pebble to a new location when the robot bumps into it. In this way, the robot in the current implementation can navigate past objects on the floor. Note that obstacle avoidance is beyond the topic of this paper, and if the robot is equipped with some mechanisms to detect an obstacle before collision such as a laser range finder, the robot could avoid obstacles without collision. The robot may happen to go out of line-of-sight of the network of pebbles. In such a case, the robot moves around in the local area to search for infrared signals as well as searches for another path to the final goal. If it finds the current target pebble, the robot resumes the navigation. If it finds a new path to the final goal, the robot starts a new navigation. If it cannot find the way, it stops navigation and says, "Can you help me? I'm lost."

Based on the constructed network topology, we apply Dijkstra's algorithm [27] with a uniform edge cost for finding the shortest path. All pebbles directly visible to the robot are tested as a starting point of a path. After finding the shortest path, the robot moves to each pebble in the path one by one. For example, if the shortest path is 7-3-2-5, the robot first moves toward pebble No.7. After the robot reaches pebble No.7, the robot then moves toward pebble No.3, and the robot repeats this process to finally arrive at pebble No.5.

One fundamental limitation with the current implementation is confusion caused by highly reflective surfaces such as metal-coated walls (Fig. 8(a)). We observed that the robot receives infrared signals from all directions if the robot is in a corner made of these walls and there is a pebble nearby (Fig. 8(b)). In such a case, the robot does not understand the direction of the target pebble and cannot move forward using a tracing strategy. We implement a short random walk in the local area when receiving the target pebble's signals from all directions until some of channels do not receive the signals. Although this problem can happen in an environment with highly reflective walls, it is the case that the robot happens to move into a corner, and we observed that the robot reached the final goal in most navigation trials. We confirmed that navigation can be done when a pebble is near a highly reflective wall (Fig. 8(c)). In this case, the angle estimation of the robot has slight errors, but the robot can move in a similar direction and can try the next step of angle estimation. We also confirmed that normal wallpaper-covered walls do not cause this problem.



(a) (b) (c)

Fig. 8. Possible Limitation. If there are highly reflective walls like (a) in the environment, the robot may receive the same signal from all directions and be confused (b). On the other hand, the robot can deal with the case that a pebble is near a highly reflective wall (c).

4 Sample Deployment

We took a pebble system to a demonstration house and conducted a trial deployment and navigation in a relatively realistic environment. We took a first-time user from our group and asked her to deploy the pebble system. The goal of the deployment was to get a robot navigate between a kitchen and a studio. Fig. 9 shows part of the

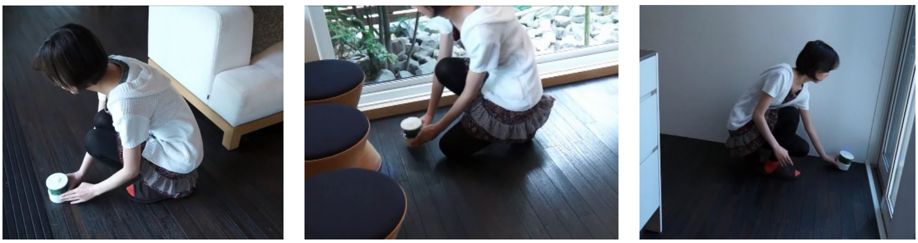
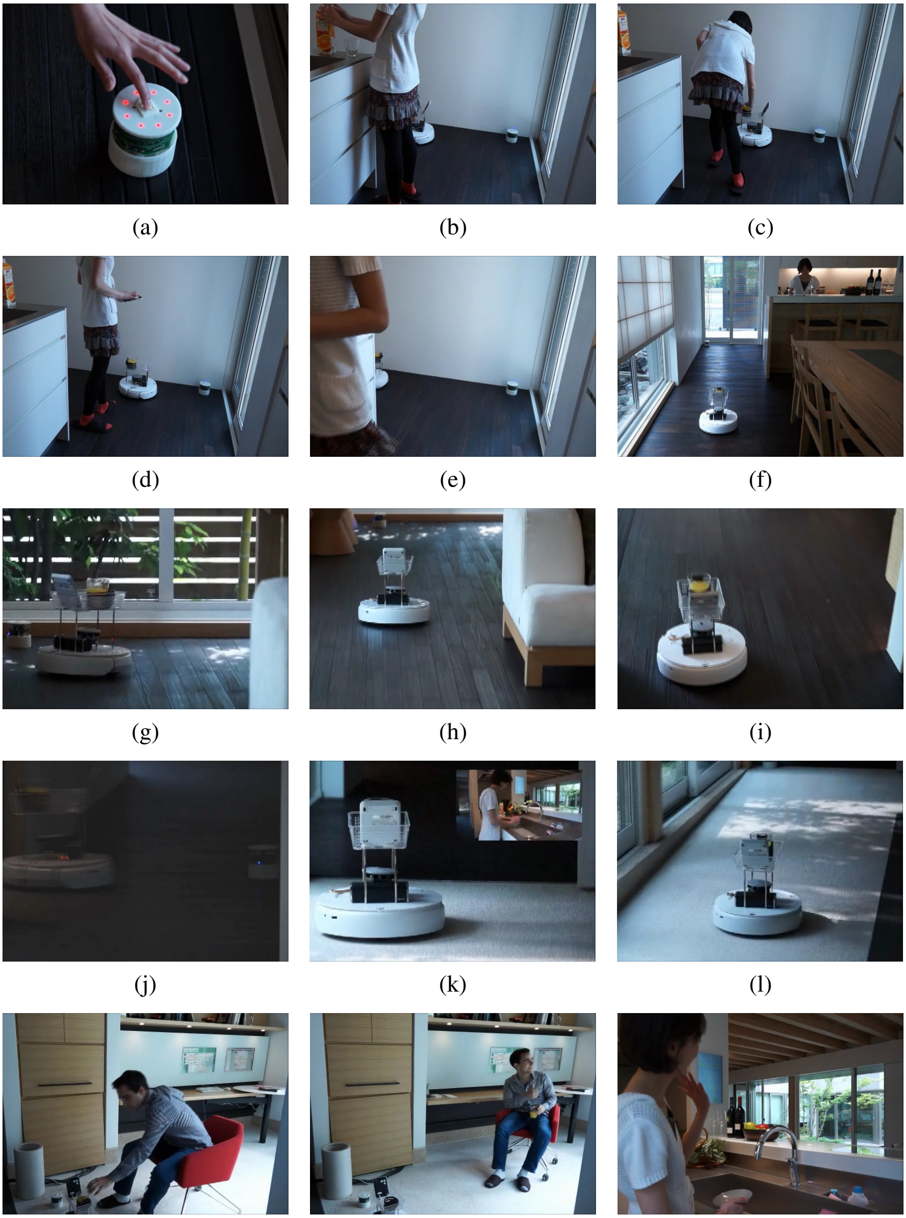


Fig. 9. Deployment Process. We took a pebble system and a first-time user to a demonstration house and asked her to design navigation routes in this environment.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o)

Fig. 10. Navigation Procedure. The user in a kitchen called a robot by tapping a pebble button (a). The user poured juice into a glass, while the robot was moving toward the pebble (b-c). The user directed the robot to travel to a studio using a remote control (d). The robot autonomously traveled out of the kitchen (e-f), through a living room (g-i), a passage (j-l), and the studio (m). The other user received the glass of juice (n-o). In this way, users were able to pass physical things between remote locations by employing the robot.

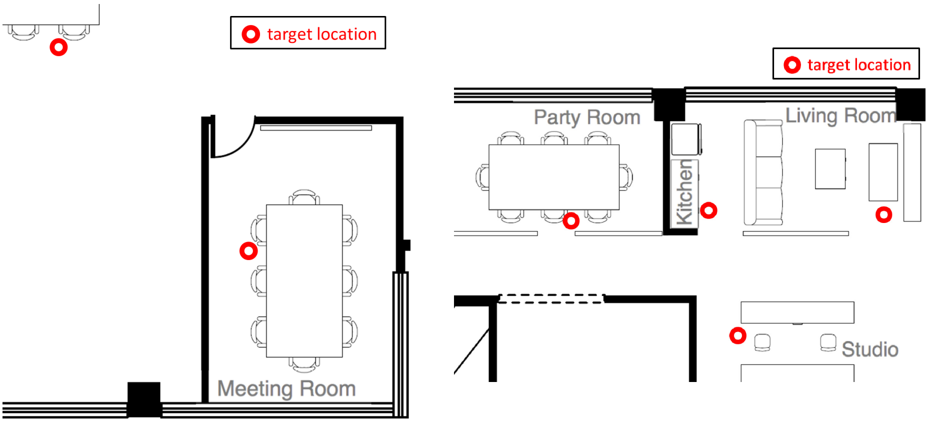
deployment process, which is just placing each pebble so that neighboring pebbles are within line-of-sight. A total of seven pebbles were placed in a kitchen, a living room, a passage, and a studio.

After deployment, we asked her to try using the robot. We gave the scenario: “A wife in the kitchen wants to deliver a glass of juice to her husband in the studio.” Fig. 10 shows stills from the video record of the navigation procedure. She called the robot and directed it to deliver a glass of juice to the studio using functions derived by the pebble system. The robot successfully navigated and delivered a glass of juice, indicating that the pebble system can provide environmental knowledge for robot navigation even in an unfamiliar environment.

5 User Study

5.1 Goals and Methods

We designed a user study to test if the pebble system is understandable and usable for end-users to configure an environment. The user study involves two tasks. The first task (Task 1) is to investigate the clarity of our system. The participants are given only a user guide describing the system. Task 1 uses a room and a passage (Fig. 11(a)). The two target locations do not have a direct view between each other, and thus the participant needs to place pebbles at transit points. After Task 1, the experimenter asks about the system to determine the understanding of the participant and demonstrates full function of the system. The experimenter also accepts questions from the participant. The second task (Task 2) is intended to test the pebble system in a more complicated environment with four different room spaces and a passage (Fig. 11(b)).



(a) Task 1 (b) Task 2

Fig. 11. Experimental Configuration. We use a room and a passage for Task 1, and four rooms and a passage for Task 2.

In both studies, the participant is asked to think aloud so that the experimenter can observe the participant's thoughts. The experimenter places the robot in the environment and hands out nine pebbles that are switched off.

5.2 Detailed Procedure

In Task 1, the participant is instructed to read the user manual. After the participant finishes reading the manual, the participant is asked to configure the environment where the robot can navigate toward two target locations using pebbles. The user manual mentions how to use LED and voice feedback, but it does not instruct or encourage the participant to use the feedback. After deployment, the participant is asked to direct the robot to go to two targets using two different methods: by the remote control and by a pebble button. The participant is allowed to adjust the positions of pebbles in case the robot cannot arrive at the target. If the participant has difficulty completing the task, the experimenter regards it as the participant could not understand the system, and provides necessary help.

In Task 2, the participant is asked to design a network for the robot with four target locations. The participant is allowed to revise their deployment. After deployment, the experimenter checks to see if the network is constructed so that the robot can go to all four of the locations. After that, the participant is asked to move the robot between some of the target locations, given some scenario, such as "A wife in the kitchen wants to deliver snacks and coffee to her husband in the living room."

After Task 2, we conduct a semi-structured interview with the participant to ask opinions about their experiences using the system and how the participant thinks the system should be improved.

5.3 Participants

Eight non-expert participants (five female and three male with ages ranging from 26 to 62; 26 year old female (denoted as 26f), 28 year old male (denoted as 28m), 29m, 29m, 30f, 31f, 50f, and 62f) were invited to participate. The participants received 3000 yen, roughly 25 dollars, for taking part in the user study.

5.4 Results

Task 1. Six out of eight participants correctly deployed pebbles at appropriate places to generate navigation trails and successfully navigated the robot to the specified targets. One participant (28m) did not understand that the pebbles connected as a network. He attempted to move the robot step by step using only one pebble. He placed one pebble at a transit point to move the robot and then took the pebble to the next transit point. Another participant (26f) did not notice the visibility restriction, and the pebble network was disconnected between the inside and outside of the meeting room.

We observed a participant (29m) who tried to use pebbles in an unexpected manner. He wanted to call the robot, so he placed a pebble on the table and pressed the button as if he was pressing a service bell. However, because elevated communication is not supported by the current implementation, the pebbles on the floor did not receive the user request and the robot failed to come. Another participant (31f) attempted to place pebbles in inconspicuous places, e.g., under chairs or at corners. She argued that navigation landmarks should not disturb users' daily activities and they should not be placed at the center of the corridors.

Task2. We confirmed using the voice feedback that all participants completed an appropriate landmark deployment without the experimenter's help. During the deployment, all participants used the voice feedback feature to confirm navigation trails. The participants reported that it was not much more difficult to plan more complicated navigation trails than the ones given in Task 1. After the deployment, eight participants tested a total of 21 navigation trails. In 17 of the tested navigation trails, the robot successfully navigated to the specified place without the participant's help. In the other four navigation trails, the participants first experienced unsuccessful navigations caused by loss of the target pebble location. After adjusting the pebbles, the robot completed the navigation in all four trials.

We observed the participants' tendency to reposition the navigation landmarks several times to find more efficient routes, for example, when trying to reduce the number of pebble landmarks. We observed a participant (28m) who tried to extend the navigation network to an additional room that was not required in Task 2. These observations provide evidence to support our claims that a pebble system has the advantage of being reconfigurable, which provides users more freedom when designing navigation routes.

Interview. We conducted an interview with each participant. All participants were asked their opinions about the feedback. Overall, we obtained positive comments regarding the feedback that assists deployment, and seven out of eight participants gave a more positive evaluation of the voice feedback. One participant (30f) answered that she never watched the LED indicators; instead, she always checked the possible trails by listening to the voice feedback. We conclude that LED feedback could be improved in future revisions.

Two participants gave negative feedback for the entire system. One (31f) argued, "it seems too costly to maintain." In the current implementation, pebbles operate on battery power and require battery replacement after a certain period of time. Since most pebbles, with the exception of dynamic targets, are placed at fixed locations for a long time, we plan to revise the hardware so that power may be supplied from an outlet. The other (26f) did not want to have to "think where to place them to construct a connected network." We, however, think that deploying devices appropriately should remain the user's responsibility. A user can have all the advantages of using pebbles, such as explicitly marking areas to avoid and labeling locations, by the small work of ensuring a direct line-of-sight between each pebble, which is difficult for the system to do automatically. Some people may prefer an effort-saving method, but others may prefer a more customizable method such as the proposed method. Our proposed method provides yet another way to realize environment configuration for robots.

Most of the participants (26f, 28m, 30f, 31f, 50f, and 62f) mentioned, "it would be better if a pebble was smaller." In the current implementation, infrared communication reaches to at least 5m, and this may be more than needed for the home and office environment. We thus think that with careful performance considerations, the device could be made smaller. We think the participants would be satisfied if we eventually made them small enough to be attached to home or office furniture.

One participant (29m) commented that he would like to "call the robot using a mobile phone." This is possible with our current implementation. When a pebble is near the user, the user just tells the system the ID of the pebble using a mobile phone. If the user is taking a pebble as a dynamic target as described in Section 3.2, any place could be the robot's goal. As an alternative approach, a mobile phone that comes with infrared transmitters could work as a pebble to call the robot.

6 Discussion and Future Work

The proposed method provides a simple way for manually configuring an environment for robots, in particular, describing the structure of the environment and naming specific locations. This is, in a manner of speaking, the technique of "Simultaneous Labeling and Mapping." Our method is one option for users having preferences as described in this paper, and we do not aim to completely replace existing methods with the proposed method. If a user does not care about the drawbacks of existing methods for mapping the environment, those techniques can be used to achieve more sophisticated navigation. For example, in the case where the user can receive daily expert assistance, more sensors can be installed to achieve finer robot localization. In the case where the user does not mind the robot exploring the environment, automatic mapping helps the robot to perform more efficient navigation. Conversely, labeling is fundamentally a manual process, and not well fitted for autonomous mapping. Therefore, the proposed technique can be used to add the labels to locations in the already mapped environment.

Although we proposed a simple way of mapping and labeling, non-expert but proactive users may prefer more flexible solutions. Seeking out different possibilities and providing an adequate degree of flexibility for each user's preference is a continuing topic of interest for us. One direction is to use text data for labeling. Inputting text may be a little more complicated than attaching physical labels on buttons, but it still does not require expert knowledge. If the system stores the text of the location names, they can be used to specify locations in various forms such as typing, drop-down menu, and voice input. We are now developing a smartphone interface for label registration based on text data (Fig. 12).

Another direction is to conduct a longitudinal study that could possibly reveal more practical limitations than our user study, as well as uncover further research challenges. Ideas presented in this paper may be elaborated by an additional user study of the system.

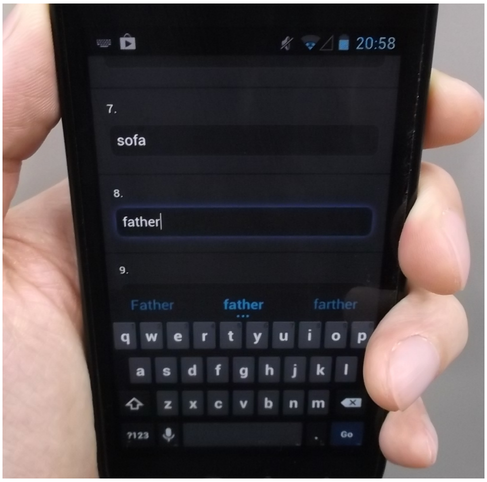


Fig. 12. Labeling by Text Data. We are developing a smartphone interface to label locations. Labels based on text data can be seamlessly used to specify locations by typing, drop-down menu, or voice input.

7 Conclusion

We proposed a tangible user interface for designing navigation routes for robots. Our method is designed for non-expert users, and solves problems associated with existing methods, such as the technical complexity of labeling, invasion in unwanted areas, and complications for reconfiguration. Our user study showed that the proposed devices are easy to use and provide users with flexibility where planning customized navigation routes in simulated home and office environments.

We believe that robots, as autonomous mobility entities, improve convenience in home and office environments by providing more functionality such as monitoring and delivery, but it is difficult to realize completely autonomous robots. In this paper, we demonstrated that a little help from the user improves the functionality of the robot as well as the naturalness in specifying locations. We also assert that, besides improving performance and functionality, improving usability and other human factors is an important topic when using robots with humans.

Acknowledgments. We thank Akihiko Ishida and Greg Saul for their valuable suggestion in the initial stage of this work and their help in implementing an earlier version of the proposed system. We also thank Youichi Kamiyama for his help in implementing the proposed system. We also thank Thien-Anh Tran and Gary Mialaret for their help in testing the proposed system and taking a video at the demonstration house.

References

1. Leonard, J.J., Durrant-Whyte, H.F.: Simultaneous Map Building and Localization for an Autonomous Mobile Robot. In: IEEE/RSJ International Workshop on Intelligent Robots and Systems, vol. 3, pp. 1442–1447 (1991)
2. Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C.T., Ishiguro, H., Hagita, N.: A Semi-autonomous Communication Robot - A Field Trial at a Train Station -. In: ACM/IEEE Annual Conference on Human-Robot Interaction, pp. 303–310 (2008)
3. Ishii, K., Takeoka, Y., Inami, M., Igarashi, T.: Drag-and-Drop Interface for Registration-Free Object Delivery. In: IEEE International Symposium on Robot and Human Interactive Communication, pp. 228–233 (2010)
4. Park, S., Hashimoto, S.: Indoor localization for autonomous mobile robot based on passive RFID. In: IEEE International Conference on Robotics and Biomimetics, pp. 1856–1861 (2009)
5. Park, S., Hashimoto, S.: Autonomous Mobile Robot Navigation Using Passive RFID in Indoor Environment. IEEE Transactions on Industrial Electronics 56(7), 2366–2373 (2009)
6. Mi, H., Ishii, K., Ma, L., Laokulrat, N., Inami, M., Igarashi, T.: Pebbles: An Interactive Configuration Tool for Indoor Robot Navigation. In: Annual ACM Symposium on User Interface Software and Technology, Demonstrations, pp. 11–12 (2012)
7. Bahl, P., Padmanabhan, V.N.: RADAR: An In-Building RF-based User Location and Tracking System. In: The Conference on Computer Communications, Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 775–784 (2000)
8. Haeberlen, A., Flannery, E., Ladd, A.M., Rudys, A., Wallach, D.S., Kavraki, L.E.: Practical Robust Localization over Large-Scale 802.11 Wireless Networks. In: Annual International Conference on Mobile Computing and Networking, pp. 70–84 (2004)
9. Nishida, Y., Aizawa, H., Hori, T., Hoffman, N.H., Kanade, T., Kakikura, M.: 3D Ultrasonic Tagging System for Observing Human Activity. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 785–791 (2003)
10. Chang, H., Choi, J., Kim, M.: Experimental research of probabilistic localization of service robots using range image data and indoor GPS system. In: IEEE International Conference on Emerging Technologies and Factory Automation, pp. 1021–1027 (2006)
11. Niwa, H., Kodaka, K., Sakamoto, Y., Otake, M., Kawaguchi, S., Fujii, K., Kanemori, Y., Sugano, S.: GPS-based Indoor Positioning system with Multi-Channel Pseudolite. In: IEEE International Conference on Robotics and Automation, pp. 905–910 (2008)
12. Gonzalez, J., Blanco, J.L., Galindo, C., Ortiz-de-Galisteo, A., Fernaindez-Madrigal, J.A., Moreno, F.A., Martinez, J.L.: Combination of UWB and GPS for indoor-outdoor vehicle localization. In: IEEE International Symposium on Intelligent Signal Processing, pp. 1–6 (2007)
13. Ubisense, `http://www.ubisense.net`
14. Ishiguro, H.: Distributed Vision System: A Perceptual Information Infrastructure for Robot Navigation. In: International Joint Conference on Artificial Intelligence, pp. 36–41 (1997)
15. Saito, S., Hiyama, A., Tanikawa, T., Hirose, M.: Indoor Marker-based Localization Using Coded Seamless Pattern for Interior Decoration. In: IEEE Virtual Reality Conference, pp. 67–74 (2007)
16. Nakazato, Y., Kanbara, M., Yokoya, N.: Localization System for Large Indoor Environments Using Invisible Markers. In: ACM Symposium on Virtual Reality Software and Technology, pp. 295–296 (2008)

17. Miyama, S., Imai, M., Anzai, Y.: Rescue Robot under Disaster Situation: Position Acquisition with Omni-directional Sensor. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 4, pp. 3132–3137 (2003)
18. Pugh, J., Martinoli, A.: Relative Localization and Communication Module for Small-Scale Multi-Robot Systems. In: IEEE International Conference on Robotics and Automation, pp. 188–193 (2006)
19. Pugh, J., Raemy, X., Favre, C., Falconi, R., Martinoli, A.: A Fast Onboard Relative Positioning Module for Multirobot Systems. IEEE Transactions on Mechatronics 14(2), 151–162 (2009)
20. Roberts, J.F., Stirling, T.S., Zufferey, J.C., Floreano, D.: 2.5D Infrared Range and Bearing System for Collective Robotics. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3659–3664 (2009)
21. Yap, T.N., Shelton, C.R.: SLAM in Large Indoor Environments with Low-Cost, Noisy, and Sparse Sonars. In: IEEE International Conference on Robotics and Automation, pp. 1395–1401 (2009)
22. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem. In: National Conference on Artificial Intelligence, pp. 593–598 (2002)
23. Davison, A.J., Murray, D.W.: Mobile robot localisation using active vision. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 809–825. Springer, Heidelberg (1998)
24. Hwang, S.Y., Song, J.B.: Monocular Vision-Based SLAM in Indoor Environment Using Corner, Lamp, and Door Features From Upward-Looking Camera. IEEE Transactions on Industrial Electronics 58(10), 4804–4812 (2011)
25. Pirker, K., Ruther, M., Bischof, H., Schweighofer, G., Mayer, H.: An Omnidirectional Time-of-Flight Camera and its Application to Indoor SLAM. In: International Conference on Control, Automation, Robotics and Vision, pp. 988–993 (2010)
26. Ishii, H., Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In: ACM SIGCHI Conference on Human Factors in Computer Systems, pp. 234–241 (1997)
27. Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs. Numerische Mathematik 1(1), 269–271 (1959)

Robots for Real: Developing a Participatory Design Framework for Implementing Educational Robots in Real-World Learning Environments

Lykke Brogaard Bertel[1], Dorte Malig Rasmussen[1], and Ellen Christiansen[2]

[1] Danish Technological Institute, Robot Technology, Odense, Denmark
{lbb,dmrn}@dti.dk
[2] Aalborg University, Department of Communication and Psychology, Aalborg, Denmark
ech@hum.aau.dk

Abstract. As educational service robots become increasingly accessible, the demand for methodologies that generate knowledge on r-learning applicable to real world learning environments equally increases. This paper proposes a participatory design framework for involving users in the development of robot-supported didactic designs and discusses its applicability to existing educational contexts on the basis of a case study on the implementation and use of the therapeutic robot seal Paro at a school for children with an autism diagnosis.

Keywords: Participatory design, Human-Robot Interaction, Education.

1 Introduction

For the educational robots of tomorrow to succeed as trainers, teachers and playmates, their appearance, interaction style and social behavior must match the users' needs and meet the demands of a changing educational context. This call for well-established guidelines to how such robots should be designed and in what ways interacting with them can support learning [1]. Thus new areas of research within Human-Robot Interaction (HRI) have emerged, focusing particularly on HRI in education, introducing terms such as r-learning [2] and Child-Robot Interaction [3]. Also the applicability of robots in specific educational curricula such as STEM teaching and language learning [4-6] and special needs education [7-10] has gained increased interest in recent years.

Robots for education are often categorized by their design and functionality [11-12] e.g. as educational robotics (or hands-on robotics) as opposed to educational service robots (social, anthropomorphized robots) [2]. Since hands-on robotic kits generally have been argued to support a constructionist approach to learning [13] research within educational robotics has a tradition for including users in the design and development of such robots through participatory innovation processes [14-15].

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 437–444, 2013.

Within educational service robots, on the other hand, a much clearer boundary is drawn between users and developers [2]. Users are usually not included in the design and development process and limited research has been conducted on the actual implementation of educational service robots in real-world learning environments [16]. Due to high technological complexity and cost combined with low system stability and usability, the majority of research on educational service robots is still based on highly controlled experiments in lab settings which cause certain limitations to the transferability and applicability of the results to real-world learning environments.

First of all, the majority of HRI studies use the Wizard of Oz method (WoZ), in which the robot is fully or partially tele-operated (often without the participants knowing) [17]. Besides the ethical implications WoZ may have (imagine a child revealing his deepest secrets to a robot pet, only to find it is being remotely operated by a researcher, teacher or perhaps parent), the constant demand for an operator reduces the usability of the robot and entails the risk of unrealistic expectations and thus disappointment and rejection of "real" (and perhaps less impressive) autonomous robots. Secondly, though the need for “long-term interaction” studies is often made explicit in calls for papers, the notion of “long-term” is still not really well-defined and the majority of these studies are characterized by what one might call ”serial short-term interaction” in which for instance users participate in experiments once a week for a period of time [18]. Although these studies do explore users’ ongoing interest in a robot when interaction is repeated, they do not take into account contextual factors such as the communication and motivational structures, social dynamics or cultural values in the learning environment that may also affect the interaction and user motivation. In addition, long-term r-learning studies are often conducted in dedicated "playtime" labs [19] overlooking the fact that children, and particularly children with special needs, need familiar, predictable environments to feel safe and can behave very differently in unfamiliar settings. Thirdly, many r-learning experiments are evaluated by quantitative metrics of measurements. This means, that statistical analysis is often applied to too small data sets [7] and that e.g. the number of dialogues between a child and a robot (e.g. [19]) is given priority rather than the quality of such dialogues. Finally, the advantage that experimental methods are said to have over field research when it comes to generalization induces a predominant preference for generalizable results; completely overlooking the importance of individualization as stressed in technology-enhanced learning and special needs education in particular.

To summarize, although r-learning experiments can indeed be of great relevance to the research community, knowledge produced within a theoretically construed setting will necessarily remain theoretical. Thus, a growing number of researchers argue that r-learning research, to have an impact *on* society, must be conducted in and in collaboration *with* society [16]. This paper is in line with this view and argues that a lot can be gained from the hands-on educational robotics’ constructionist and participatory approach to didactics when exploring and developing educational service robots. Thus, the paper proposes a participatory design framework for developing robot-supported didactic designs in close collaboration with users and practitioners, and demonstrates its applicability to real-world learning environments in a case study on the implementation and use of the therapeutic robot seal Paro in autism education.

2 Time-Space-Structure: A Participatory Design Framework

As users are increasingly viewed as active consumers of technology and co-producers of technological practice, new approaches to designing, developing and testing technology are also emerging, requesting systematic methods for including users' experiences, views and needs as resources and including users and practitioners as potential co-developers of future products and services. Participatory Design is particularly useful for this purpose [20] since it is indeed in line with the constructionist perspective, viewing people, social systems and technological development as interrelated, emphasizing the importance of stakeholder involvement and the value of concrete, context-dependent knowledge. This particular participatory design framework focuses on three components: *Time*, *Space* and *Structure* (TSS) as central to design processes with and for users and practitioners. These components are individually present in all user studies. However this paper proposes that particularly the balance between these components determines (1) the level of user participation and (2) the nature of the research questions that a particular user study can address.

2.1 Time

Time covers the length of the study in question. For studies addressing users' first impressions and initial attitudes towards educational service robots time is not necessarily a crucial component, although the often accompanied interest in attitude changes over time naturally increases the importance of extending the user study in time. The time invested is in itself a strategic choice, since a longer case study naturally requires more resources from both researchers and practitioners. At the same time, though, it allows for more research and development iterations as well as ongoing adjustment of the project according to newly gained knowledge and experiences and thus more reliable and replicable results. Our case study extended over a period of 3 months, which allowed users and practitioners to become familiar with the technology and to a certain extent include it as part of their everyday lives and practice.

2.2 Space

Space covers the place in which the study is conducted (e.g. field research as opposed to lab experiments) but do also contain several symbolic meanings. Firstly, a certain stability of the technology must be ensured before sending it out into 'space'. Early educational service robot prototypes are not well-suited for these studies unless developers are present at all times, which in turn require extensive resources. Secondly, researchers must ensure a "creative space", making room for ideas and views of users and practitioners. Participatory design is to a certain extent democratic, providing the opportunity for explorative and creative processes rather than dictating the use of the technology [20]. Finally, a "safe space" must be provided for participants to disclose their personal views and experiences, without the fear of being exposed, since dissenting views and opinions are also essential and enable the discovery of new aspects of the implementation and interaction with the technology.

2.3 Structure

Structure covers the level of control in the study. In experiments control is crucial to ensure validity and generalizable results. In participatory design processes, however, too much control can demotivate the participants or even induce reluctance towards the researchers or the technology. On the other hand a total lack of control could possibly leave the development unfocused, participants insecure and unmotivated, and the technology unused. In order to provide the best conditions for creative processes, we argue the necessity of a structured but flexible framework for participants to navigate. In our case study this was done by devising a blueprint for the process in collaboration with the participants, by planning workshops to collect and share experiences as well as by introducing goal-setting as a way to facilitate direction and motivation.

3 TSS in Practice: A Case Study

To investigate the applicability of TSS in practice we conducted a three month case study on the use of the therapeutic robot seal Paro at a school for children with autism. The purpose of the case study was two-fold; to explore the potential of robots to promote motivation, communication, play and learning in autism education, to investigate the importance of structure in the TSS-framework as well as to identify additional contextual factors that both influence and are influenced by the technological intervention. A total of three Paros were implemented in 3 existing teaching teams of 4-5 teachers and 6-8 students each. At the school the teaching teams are divided by age: T1 (6 children aged 6-9), T2 (7 children aged 10-13) and T3 (8 children aged 14-17) although the students are all at a similar stage of cognitive development (approx. 0-1 years of age) in which most have no spoken language and use alternative tools for communication. To evaluate the importance of structure in the TSS-framework the groups were differentiated by the stage in which goal-setting activities were initiated and thus the level of structure (ranging from structured, to semi-structured and non-structured). The research setup comprised the following three phases:

1. A 3 week *exploration* phase initiated the study. In this phase the Paros were introduced and made available to the teachers without any limitations or restrictions of use. The robots were explored with all students in different setups (i.e. individually, in pairs and in groups) with different levels of teacher involvement (independent, guided or prompted interactions). During this phase we were present most of the time, mainly to ensure no technical issues were obstructing the introduction as well as to gain knowledge on the context and to create a baseline on the basis of the student's immediate responses to Paro. Data collection methods consisted of participating and non-participating observations, in situ interviews and video material. The teachers in T3 defined the goals for the use of Paro in their team early at this stage, i.e. facilitating social interaction and supporting the transition from familiar to unfamiliar settings (e.g. going somewhere new, meeting new people etc.)
2. A 3 week *co-ideation* phase followed consisting of a series of cross-team workshops for the teachers to share knowledge and ideas on the use of Paro with different

children for different purposes. Based on short videos and small narratives derived from the data obtained in phase 1 they collaborated on developing didactic designs, i.e. different types of *use scenarios* (e.g. type of practice and its context), *use applications* (e.g. type of interaction and the role of Paro) and *user personas* (e.g. student's individual needs, preferences and communication style), which could then be compared across teams. The teachers in T1 defined the goals for the use of Paro in their team at this stage, i.e. increasing bodily awareness through sensory stimulation and maintaining or shifting focus in certain crucial situations.

3. In the final *co-creation* phase the concepts of use were applied to systematize the implementation process in connection to the goals defined. At this point the teachers in T1 and T3 themselves restricted the use of Paro to 3-4 students, whom they considered were gaining the most from interacting with Paro. T2 made no restrictions on the use at any stage. During this phase we did follow-up visits at approximately one-weekly intervals, and the study was concluded with a focus group interview with all teachers involved in the final phase.

4 Results and Discussion

Results showed that the stage at which goal-setting activities were initiated did have an impact on *participation* (more teachers participated in T1 and T3 when compared to T2), *motivation* (the use of Paro was more frequent in T1 and T3 even though they had made restrictions on use and user personas, whereas T2 had not) and the potential for *innovation* (more didactic designs were developed in T1 and T3). When comparing the structured (T3) and semi-structured (T1) design processes, results indicate that although a structured design process can help ensure particularly ongoing *participation* and *motivation*, the semi-structured process seem to facilitate more reflection and thus *innovation* of the pedagogic practice, possibly since T1 had not decided on specific didactic designs prior to phase 2 and thus were more open to suggestions from the other teams. In addition to the results on participation, motivation and innovation we identified the four categories of contextual factors (*knowledge, relations, values* and *flow*) that affected the way in which Paro was implemented and used in practice. Thus these contextual factors should be considered when designing educational service robots for real-world educational settings.

4.1 Knowledge

In the case study we found, that the TSS framework enabled us to gain insight into how the technological intervention affects and is affected by the embedded knowledge in the environment and its traditions for sharing knowledge. For instance, one teacher's reflections on a workshop session show the opportunity that TSS brings for inspiration across teams: *"It was a really good idea to try and use Paro as a companion during the mandatory psychological tests as one other teacher suggested"*. Another teacher reflected upon the time management system in the environment and how it did or did not support the introduction of the robot: *"So far my sessions with Paro have been* *introductory with sequences of five minutes. Maybe ten minutes. Actually it is difficult to estimate the time. It's actually a little strange that I lose track of time, since we keep very strict time schedules"*. A third teacher explained how he had gained knowledge on the application of new technology in his context of use: *"We should not replace existing support systems that work. The point is to find areas where the child shows potential for development but where we somehow lack the tools"*. These results show that the framework provided a platform for the participants to both share and create new knowledge by reflecting on their own practice. Thus, the framework not only provides researchers with knowledge *about* practice, it provides the opportunity to develop, share and use knowledge in collaboration *with* practice.

4.2 Social Relations

Experimental r-learning often focuses on the social relations between child and robot. However, in the case study we found that other social relations within the context can be of equal importance, i.e. child/child, child/teacher, teacher/robot or teacher/teacher relations. In some cases the social relation between a teacher and a child alone ensured a successful intervention with Paro: *"We brought Paro along tobogganing and the student gave Paro a ride on the sledge. It has been several years since we were able to get him to go tobogganing with us. Of course, it was also a little stressful for him, but he enjoyed it. It's also kind of a special case because I've known him for many years and I am able to see that he is experiencing something positive, even though outside viewers may think he's a little stressed out"*. Here the teacher's relationship with the child enables her to explore the potential for Paro to motivate otherwise rejected opportunities for play. In yet other situations the social relationship itself was the actual goal of the intervention. For instance, in one case Paro was used as a social mediator between two children of which one was afraid of the other: *"Paro has been an object of joint attention. During the sessions with Paro they have been able to approach each other while petting and grooming Paro. At one point they also briefly touched each other"*. In another case, the goal was for a student to be able to approach new people: *"We brought Paro when visiting the school where a student will be going next year. She introduced Paro to the other students and was more confident than usual"*. These examples illustrate ways in which existing social relations can mediate the use of technology and affect the outcome of the intervention.

4.3 Values

From the case study we found that underlying values, and particularly competing values, in the educational environment mediated both the teachers' attitudes towards the robot as well as new technology in general. For instance, issues related to both attending students' individual needs and meeting outside academic standards was a source of frustration for some teachers: *"It doesn't really make sense that the parents have to apply for dispensation from the common national academic tests when these students' level of cognitive development is estimated equivalent to 0-1 years of age"*. For other teachers, this conflict of values caused an increased awareness of one's own professionalism and reflections on whether the teacher should promote *independency* or *intentionality* in daily activities, e.g. by continuously prompting culturally defined "meaningful" play as opposed to letting the children play on their own in a perhaps to outside viewers less meaningful way. This shows that not only the attitudes of direct users and the values in the immediate context of the intervention are of importance when implementing robots in educational environments. Also indirect and potential users, parents and policymakers represent a group of stakeholders whose views and values affect and are affected by the implementation of educational service robots.

4.4 Flow

Here the term *flow* has several meanings. Firstly, the TSS framework provided the opportunity to gain insight into the daily flow of people in the context which is very much related to the above mentioned identification of indirect users and stakeholders. For instance, some children were living with their parents whereas others were permanently placed in care, visiting their parents once a week. Secondly, it emphasized the advantages of true long-term interaction in comparison with serial short-term interactions. The accessibility to the technology in the case study created opportunities for both consistency and flexibility in the r-learning designs and thus flow in the interaction. For instance, one teacher was able to apply Paro at certain crucial time to create and sustain flow: *"We were at the church for Christmas. One student was feeling uneasy, covering his ears and repeating "close the door, close the door". Luckily I had brought Paro, so I showed it to him and said "look who it is" and he took down his hands to touch Paro and smiled".* Here, the teacher's experience made it possible to exploit an optimal moment for intervention. Such moments occur spontaneously and cannot be orchestrated. Thus they are much less likely to occur in experimental settings where time and space are completely elided in favor of structure.

5 Conclusions, Limitations and Directions for Future Research

This paper presents work related to the development of a participatory design framework for involving users and practitioners in robot-supported didactic design processes, arguing a balanced level of time, space and structure e.g. by using goal-setting strategically at certain stages in the process. The applicability of the framework to real-world educational environments was demonstrated through a case study in robot-supported didactic designs for children with autism which revealed knowledge, social relations, values and flow as important factors when developing didactic designs with and for users and practitioners in autism education. The case study focuses on special needs education and the results are thus limited to this particular educational context. For the framework to be applicable to different educational contexts, e.g. specific domain didactics it must be challenged and further developed in different educational contexts, using different robotic platforms for different educational purposes.

References

1. E.g. Institute for Personal Robots in Education, http://www.roboteducation.org/
2. Han, J.: Robot-Aided Learning and r-Learning Services. In: Chugo, D. (ed.) Human-Robot Interaction. InTech (2010), doi:10.5772/8143, ISBN: 978-953-307-051-3
3. Belpaeme, T., et al.: Multimodal Child-Robot Interaction: Building Social Bonds. Journal of Human-Robot Interaction 1(2), 33–53 (2012), http://www.aliz-e.org
4. Mataric, M.J., Koenig, N., Feil-Seifer, D.: Materials for enabling hands-on robotics and STEM education. AAAI Spring Symposium, Resources for AI Education (2007)
5. Majgaard, G., Nielsen, J., Misfeldt, M.: Robot Technology and Numbers in the Classroom. In: CELDA 2010, International Association for Development, pp. 231–234 (2010)
6. Han, J., Dongho, K.: r-Learning services for elementary school students with a teaching assistant robot. In: 4th International Conference on Human-Robot Interaction (2009)
7. Diehl, J.J., Schmitt, L.M., Villano, M., Crowell, C.R.: The Clinical Use of Robots for Individuals with Autism Spectrum Disorders: A Critical Review. Research in Autism Spectrum Disorders 6(1) (2012)
8. Lee, C.J., Kim, K., Breazeal, C., Picard, R.: Shybot: Friend-Stranger Interaction for Children Living with Autism. In: CHI 2008 (2008)
9. Robins, B., et al.: From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot. In: 2nd IEEE International Conference on Advances in Computer-Human Interactions (2009)
10. Kozima, H., Nakagawa, C., Yasuda, Y.: Interactive Robots for Communication-care: A Case Study in Autism Therapy. In: IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005 (2005)
11. Hendler, J.: Robots for the rest of us: designing systems "out of the box", Robots for Kids: Exploring new technologies for learning. The Morgan Kaufmann Publishers (2000)
12. Goodrich, M.A., Schultz, A.C.: Human-Robot Interaction: a survey. Foundations and Trends in Human-Computer Interaction 1(3) (2007)
13. Papert, S.: Mindstorms: Children, Computers, and Powerful Ideas. Basic Books (1980)
14. Bers, M.U., et al.: Teachers as Designers: Integrating Robotics in Early Childhood Education. Information Technology in Childhood Education Annual (2002)
15. Hamner, E., et al.: Robot Diaries: Broadening Participation in the Computer Science Pipelin through Social Technical Exploration. In: AAAI Spring Symposium on Using AI to Motivate Greater Participation in Computer Science (2008)
16. Salter, T., Werry, I., Michaud, F.: Going Into the Wild in Child-Robot Interaction Studies: Issues in Social Robotic Development. Intelligent Service Robots 1(2) (2008)
17. Riek, L.D.: Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. Journal of Human-Robot Interaction 1(1), 119–136 (2012)
18. Leite, I., et al.: Long-term Interaction with Empathic Robots: Evaluating Perceived Support in Children. International Conference on Social Robotics (2012)
19. Shimada, M., Kanda, T., Koizumi, S.: How Can a Social Robot Facilitate Children's Collaboration? International Conference on Social Robotics (2012)
20. Von Hippel, E.: Democratizing Innovation: The Evolving Phenomenon of User Innovation. International Journal of Innovation Science 1(1), 29–40 (2009)

Computer-Supported Work in Partially Distributed and Co-located Teams: The Influence of Mood Feedback

Andreas Sonderegger[1], Denis Lalanne[2], Luisa Bergholz[3], Fabien Ringeval[2], and Juergen Sauer[1]

[1] Department of Psychology, University of Fribourg, Fribourg, Switzerland
[2] Department of Informatics, University of Fribourg, Fribourg, Switzerland
[3] ITB Consulting GmbH, Bonn, Germany
{Andreas.Sonderegger,Denis.Lalanne,Fabien.Ringeval, Juergen.Sauer}@unifr.ch, luisa.bergholz@itb-consulting.de

Abstract. This article examines the influence of mood feedback on different outcomes of teamwork in two different collaborative work environments. Employing a 2 x 2 between-subjects design, mood feedback (present vs. not present) and communication mode (face-to-face vs. video conferencing) were manipulated experimentally. We used a newly developed collaborative communication environment, called EmotiBoard, which is a large vertical interactive screen, with which team members can interact in a face-to-face discussion or as a spatially distributed team. To support teamwork, this tool provides visual feedback of each team member's emotional state. Thirty-five teams comprising 3 persons each (with a confederate in each team) completed three different tasks, measuring mood, performance, subjective workload, and team satisfaction. Results indicated that the evaluation of the other team members' emotional state was more accurate when the mood feedback was presented. In addition, mood feedback influenced team performance positively in the video conference condition and negatively in the face-to-face condition. Furthermore, participants in the video conference condition were more satisfied after task completion than participants in the face-to-face condition. Findings indicate that the mood feedback tool is helpful for teams to gain a more accurate understanding of team members' emotional states in different work situations.

Keywords: virtual teamwork, videoconference, face-to-face, mood, computer-supported cooperative work.

1 Introduction

Over the past 20 years, a growing body of literature addressed the role affective states such as mood and emotions play in the workplace [1]. Although the interest in mood and emotion in organizational research is rather young, there is already a considerable base of knowledge indicating the importance of affect in organizations and work teams. The accurate perception and understanding of team members' affective states for example has proved to have a positive influence on team processes and team

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 445–460, 2013.

performance [2]. In consideration of the importance of an accurate understanding of fellow team members' affective states for an effective and satisfactory collaboration in teams, a tool was developed that represents fellow team members' current mood by means of a visual feedback. The visual feedback consists of an avatar that can be presented on a computer screen (for teamwork in distributed teams) or on a large interactive wall (for teamwork in collocated teams). With this study we aimed to evaluate the utility of such a mood feedback in two different teamwork settings – face-to-face teams (FTFT) and partially distributed teams (PDT).

2 Related Work

2.1 Mood and Emotions in Teamwork

Mood and emotions can be distinguished with regard to their intensity. While moods are feelings of relatively low intensity, emotions are high intensity feelings triggered by certain stimuli [3]. Representing an important factor in every aspect of social communication, the important role of mood and emotions in the domain of organizational teamwork is generally not contested [4]. Various studies have already shown that affect influences human cognition and behavior in problem solving [5], motivation [6], and social behavior [7], and as such also plays a critical role in teamwork. Team members in a positive mood are for example more likely to be helpful, generous and to act with a sense of social responsibility [8]. Furthermore, positive emotions lead to positive relationships and sense of community in teams and hence have an important impact on team processes and team effectiveness [9]. Furthermore, emotional intelligence, defined as the specific ability to understand and manage moods in the self and others [10, 11], was shown to be a central factor for effective leadership in organizations [12], correlating highly with transformational leadership behavior, which is considered as being beneficial for team effectiveness compared to other leadership styles [13]. Emotionally intelligent leaders are leaders who perceive emotions accurately, understand emotions and manage emotions accurately [12]. The accurate perception and understanding of other team members' affective state is hence an important factor for successful leadership behavior. Understanding fellow team member's affective state is however not only important for group leaders but also for members of a team in general [14,2]. Awareness of fellow team members' affective states helps to maintain effective relationships, contributes to better information exchange and decision making in teams and facilitates conflict resolution [2].

2.2 The Influence of Videoconferencing in (Partially) Distributed Teams

In the context of increasing de-centralization and globalization of work processes, there is a rising demand for organizations to use technologies that enable employees to communicate and work across long distances [15,16]. The following factors have also contributed significantly to the increasing role that videoconferencing (VC) plays in today's corporate business: time constraints, high travel costs as well as the scarce availability and high cost of specialized human expertise [17,18].

VC is often considered as being equal to face-to-face communication (FTFC) with regard to the outcome of the communication process [19], for example by ensuring the use of meta-communication such as tone of voice and facial expressions. However, there are various aspects in which the two communication modes differ (e.g. body language, distancing, touch etc.). These aspects might become crucial in situations where the understanding of feelings and emotions of the other is important.

Theoretical concepts such as media richness theory [20], social presence theory [21] or telepresence theory [22] have tried to explain advantages and disadvantages of different forms of computer mediated communication compared to FTFC. Previous research has shown that collaboration within distributed teams may have some disadvantage compared to co-located teams [15,16]. This might be due to the loss of specific communicational cues based on the media that is utilized [19]. Eye contact is one example for an important cue for effective interpersonal communication [23] which is difficult to obtain in VC due to the vertical disparity between the camera mounted on the top of the screen and the position of the other person's eyes on the screen [24,25]. Other nonverbal cues such as body language, interpersonal distance or touch [26,27] as well as subtleties of affect expressions and personality appearances [19] are harder to discern in VC compared to FTFC. In this respect, missing nonverbal cues in VC play an especially important role with regard to social interaction, development of relationships and intimate communication [27]. According to Zajonc [28], emotions are a vital part of everyday social communication and are not only transmitted by the verbal channel but by nonverbal cues as well – nonverbal cues might in fact carry the main affective information. This indicates that the VC may impinge on recognition and interpretation of mood and emotions compared to FTFC. Since mood and emotions play an important role in team processes and team functioning [29], VC may not only influence mood detection and recognition among members of a team but also impinge on other factors such as team satisfaction and team performance.

3 The Present Study

The primary research question of this study addressed the influence of a tool providing visual feedback of each team-member's actual mood on the process and outcome of group work. It was expected that such a mood feedback tool would alleviate the loss of information richness in partially distributed teams (PDT) with regard to emotional aspects of team functioning. It was of particular interest whether such visual feedback would be beneficial in detecting emotional states of other team members and whether this would influence team satisfaction and team performance. In the experimental setup, one of the team members was a confederate who expressed a negative mood throughout task completion. The use of a confederate is a methodological approach of particular interest. The confederate is a specially trained actor who adopts a certain role in the study (e.g. expressing bad mood and showing withdrawal behavior during a meeting), based on the instruction of the researcher [30,31]. This allows the researcher to manipulate specific experimental conditions, such as to investigate how withdrawal behavior of one team member influences the functioning of a team. In addition, it also reduces the variance of team behavior since confederates will display only trained and fixed behavioral patterns during the testing procedure (e.g., only talking when being directly asked) rather than showing a wide range of behaviors as one would find for randomly recruited team members (e.g. ranging from dominating the group to being silent). A negative mood was chosen for this study because its influence on teamwork was expected to be stronger compared to a positive one. To answer our research question, an experiment was conducted in which 3-person teams (with a confederate in each team) completed three different tasks, either in a FTFT or in a PDT (video conference) situation. During task completion, half of the teams received feedback about the other team members' mood whereas the other half did not. The following hypotheses were formulated: it was expected that teams receiving mood feedback would be more accurate in detecting other team members' mood, that their performance would be higher and that subjective evaluations of team processes (i.e. team climate and team satisfaction) would be more positive compared to teams not receiving a mood feedback. Furthermore, it was also expected that the effect of mood feedback would be more pronounced in the PDT condition than in the FTFT condition.

4 Method

4.1 Participants

Thirty-five teams (comprising three members each) took part in the study. Since one person in each team was a (female) confederate, a total of 70 participants (80% female), aged between 18 and 35 years (M = 21.56, SD = 2.91), were recruited for this study. All participants were students and did not know each other. The gender composition of each team member was at random.

4.2 Experimental Design

Employing a 2 x 2 between-subjects design, mood feedback (feedback vs. no feedback) and communication mode (PDT vs. FTFT) were manipulated as independent variables. In the PDT condition, one person of the team (the confederate) was situated in a separate room and could interact with her teammates located in the other room in the form of a video conference setup. In the latter room, the other two team members worked together on a large screen, upon which the image of the third team member was projected using the EmotiBoard (c.f. description below).

4.3 Instruments

Participants' mood was measured twice during the experiment (at the beginning and after task completion) using the Self-Assessment-Manikins (SAM) [32], a non-verbal instrument measuring two distinct dimensions of emotions (valence and arousal) by means of graphic representations of mood in the form of manikins, based on the circumplex model of affect [33]. While arousal refers to the degree of physiological activation of the mood state (e.g. aggression vs. despair), valence is concerned with the degree to which the mood is positive or negative. For the pleasure-displeasure dimension (valence), the five depictions range from a smiling manikin to a frowning manikin. For arousal, the five depictions range from a calm manikin with closed eyes to a wide awake and highly aroused manikin. In addition, participants were asked to rate the mood of their two teammates once after task completion by means of a slightly adapted SAM scale. The instruction was changed from "How much do you feel emotionally aroused at the moment?" to "How much does the other person feel emotionally aroused at the moment?". As indicators of team performance, user behaviour was recorded with an event logger and different aspects of performance, such as task completion time, numbers of user interactions, and error rate, were calculated. Subjective workload was measured by means of the well-established NASA task load index (TLX) [34]. Team satisfaction was measured by five items of the team effectiveness scale [35].

4.4 Materials

A large plexiglass display (1.6 x 1.2 m), suitable for back-projection, served as the main interface for the EmotiBoard, with which users can interact simultaneously, sharing the same application. In our experiment, we used the interactive screen with a Wii-mote for each participant and a regular PC for the remote participant as an input device for task completion. The system provides a visual surface for collaboration by capturing and transmitting pointing device positions and events (i.e. clicking, drag and drop, deleting) between different machines. This setting was inspired by work from Ishii and Kobayashi on the ClearBoard [36]. In addition, a Java library supports the creation of multi-user applications that can be accessed through multiple remote machines at the same time, using multiple types of input devices. In the PDT condition, the video stream of the remote team members was presented on the screen in half transparency, in combination with the application surface for task completion (c.f. fig. 1). In the FTF condition, the application surface for task completion was presented on the screen.

For the mood feedback, an avatar was created and displayed in each team member's toolbox on the screen throughout the experiment, allowing the other participants to be aware of their co-workers' emotional state (valence and arousal). The mood feedback was based on participant's initial mood rating with the SAM

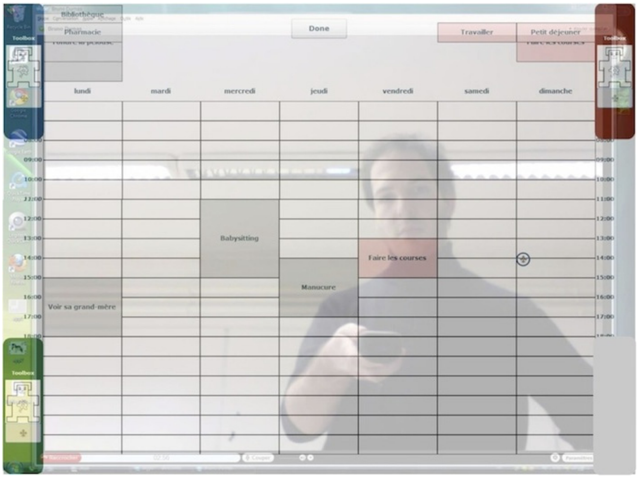


Fig. 1. EmotiBoard screen in PDT condition with remote participant in half transparency

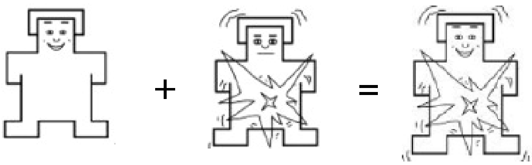


Fig. 2. Combination of SAM-measures of valence and arousal into one feedback image (here with the example of high positive valence and high arousal)

scale. For the design of the mood feedback avatar, SAM-manikins that were used to measure valence (smiling) and arousal (excitation lines) were integrated into one image (c.f. fig. 2). In the no-feedback condition, the toolbox was blank.

4.5 Team Tasks

Three tasks were used in this study, which differed with regard to cognitive demand and communication requirements: (a) a sensori-motor task required the team to connect 100 numbered dots by drawing lines between the dots (1-2, 2-3, etc., c.f. fig. 2), (b) a spatial reasoning task involved placing 12 jigsaw pieces into a figure (c.f. fig. 3), (c) a coordination and planning task required from the team members to organize the week of a student by placing 126 specific activities into their weekly schedule (planning a week, c.f. fig. 4). For each task, task completion time, number of errors and number of user interactions were recorded.

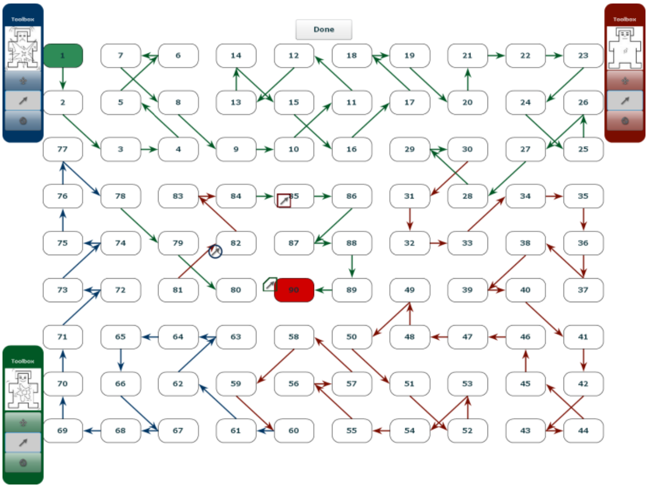


Fig. 3. Screenshot of the EmotiBoard screen for task a)

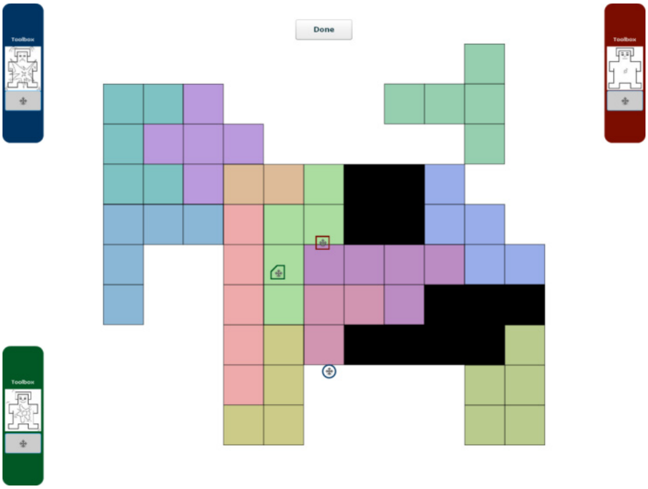


Fig. 4. Screenshot of the EmotiBoard screen for task b)

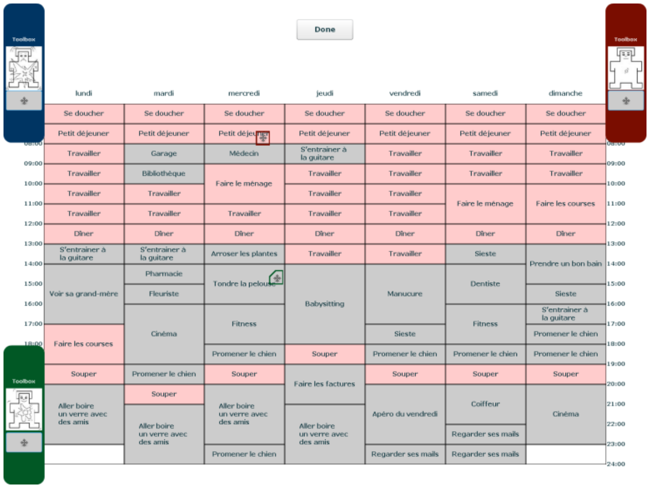


Fig. 5. Screenshot of the EmotiBoard screen for task c)

4.6 Procedure

A (female) confederate was recruited and trained to play a specific role during the completion of group-based tasks. The training was to ensure that the negative mood was expressed convincingly without appearing to be unnatural. The behavior of the confederate involved the refusal to smile throughout the session and to make negative comments about personal feelings and motivation at regular intervals. After introducing the participants to each other, a fake group drawing was organized (for the PDT condition), in which the confederate was always chosen to work in the other room. As a baseline measure, participants were then asked to rate their current mood by completing the PANAS and SAM. After a short explanation of the EmotiBoard, participants were asked to complete an initial training task to become familiar with the system and then completed the three tasks (a), (b) and (c). After that, participants were asked to rate their mood (SAM and PANAS) and the mood of the other team members (SAM). Finally, they completed the NASA TLX, TCI and team satisfaction scale.

4.7 Data Analysis

Analysis of co-variance was used to analyse the data. The influence of experience with Wii-motes, age and gender-composition of the group was examined by entering the factors as covariates. Due to the small interclass correlations (all ICC(1) < .05) and since the data was available for each team member, the data was analysed at an individual level [c.f. 37].

5 Results

5.1 Accuracy of Mood Appraisal of Other Team-Members

To evaluate the influence of the mood display on the accuracy of the rating of other team-member's mood, a difference value was calculated by subtracting the mood-assessment of the other team-members from participants' self-rating of their mood. Figure 6 shows the summarized differences between self-assessment and assessment by others for valence and arousal. Participants in the mood feedback condition were more accurate (mean difference is smaller) with regard to the assessment of others' ratings of valence ($F = 6.28$, $df = 1, 64$, $p < .05$,) and arousal ($F = 24.25$, $df = 1, 64$, $p < .001$) than participants not having mood feedback available.

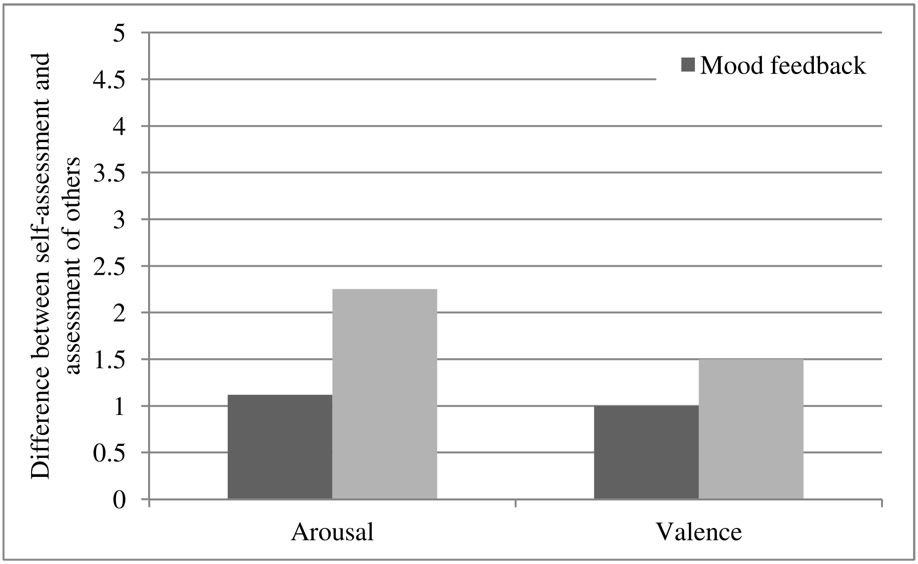


Fig. 6. Difference scores (self-rating – rating of others) for valence and arousal as a function of mood representation

Although these results indicate that mood feedback had an important influence on the accuracy of the evaluation of others' mood, test participants did underestimate this additional information. Interestingly, 38% of the participants in the mood feedback condition indicated that they did not consider the emotion representation for their assessment of other team members' mood. Only 8% reported to have exclusively considered the emotion representation whereas 54% indicated that they used both, behavioral information (gestures, speech, and facial expressions) as well as the emotion representation to assess the other team members' mood.

With regard to communication mode (c.f. fig. 7), the data indicates a less accurate appraisal of the confederate's mood in the PDT condition for valence and arousal compared to the FTFT condition ($F_{valence} = 5.24$; $df = 1, 64$; $p < .05$; $F_{arousal} = 7.16$, $df = 1, 64$; $p < .01$). Because only the confederate was in remote in the PDT condition, a difference value was calculated exclusively with regard to her self-rating; the two other participants were in the same room and hence communication mode had no influence on the accuracy of their mutual mood ratings. No significant interaction between mood display and communication mode was discovered (all Fs < 1).

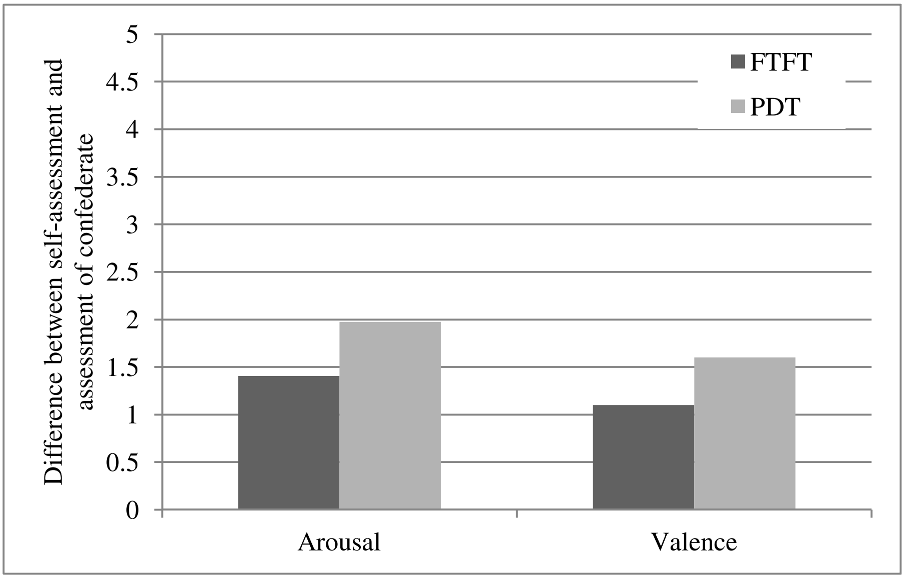


Fig. 7. Difference scores (self-rating of confederate – rating of others) for valence and arousal as a function of communication mode (FTFT: face-to-face team; PDT: partially distributed team)

5.2 Team Performance and System Management Behavior

A marginally significant interaction (communication mode x mood feedback) was found for number of errors ($F = 2.89$, $df = 1, 32$, $p < .1$), indicating an increased error rate in the PDT condition without mood feedback compared to the same condition with mood feedback. In FTFT condition, effect of mood feedback was inversed: more errors occurred with mood feedback compared to teams not receiving a mood feedback (c.f. fig.8). No main effect of mood feedback on measures of performance was found (all Fs < 1).

With regard to communication mode, analysis of the data indicated that participants in the PDT condition committed more errors compared to participants in the FTFT condition ($M_{PDT} = 4.88$, $SD = 4.47$; $M_{FTFT} = 3.56$, $SD = 2.44$; $F = 4.78$, $df = 1, 32$, $p < .05$). No further effects of communication mode and mood feedback on performance measures (task completion time and number of user interactions) were found to be significant (all Fs < 1).

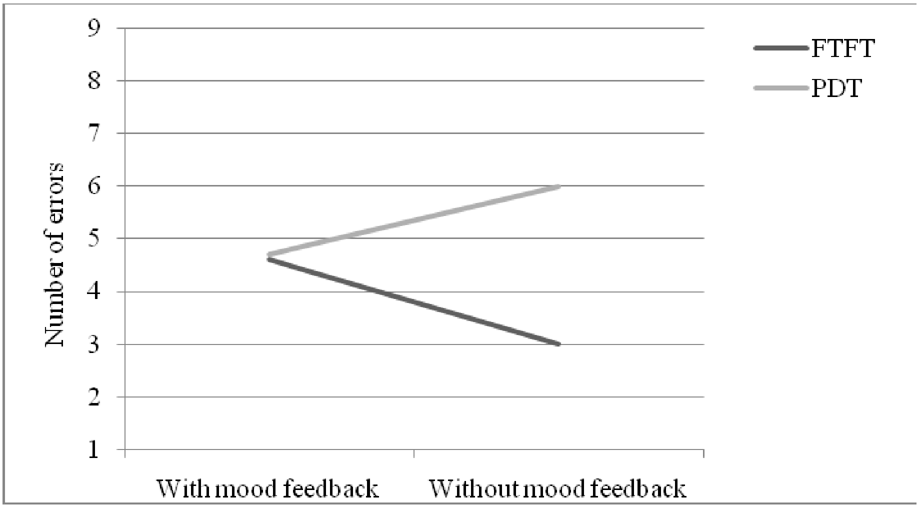


Fig. 8. Interaction between communication mode and mood feedback for number of errors committed during task completion (FTFT: face-to-face team; PDT: partially distributed team)

5.3 Workload

With regard to the different ratings of subjective workload, the analysis revealed a significant interaction (communication mode x mood feedback) for temporal demand (c.f. fig.9). Temporal demand was higher for participants in the PDT

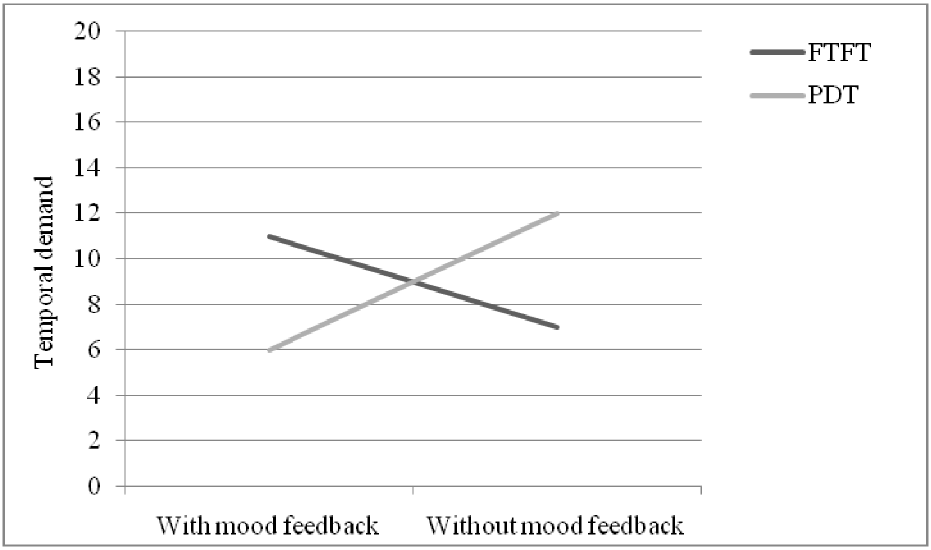


Fig. 9. Interaction between communication mode and mood feedback for temporal demand (FTFT: face-to-face team; PDT: partially distributed team)

condition when no mood feedback was presented, whereas in the FTFT condition, temporal demand was higher when mood representation was available ($F = 9.26$, df = 1, 64, $p < .01$). No significant main effect of communication mode and mood feedback and no further interaction were discovered on the other sub dimensions of the NASA TLX (all Fs < 1).

5.4 Team Climate and Team Satisfaction

Participants in the PDT condition were more satisfied with teamwork and team processes than participants in the FTFT condition ($M_{PDT} = 20.38$, $SD = 2.60$; $M_{FTFT} = 18.22$, $SD = 4.57$; $F = 3.98$, df = 1, 64, $p < .05$). Furthermore, participants in the PDT condition expressed a stronger preference for working again with the other team members compared to participants in the FTFT condition ($M_{PDT} = 3.94$, $SD = 0.74$; $M_{FTFT} = 3.25$, $SD = 1.18$; $F = 6.95$, df = 1, 64, $p < .05$). Mood feedback showed no effect on team satisfaction ($F < 1$). Furthermore, no significant interaction was found ($F < 1$).

6 Discussion

The main objective of this study was to evaluate the influence of an emotion representation tool in different collaborative work environments. The results indicated a more accurate appraisal of other team members' emotional state when mood feedback was available. This implies that this tool supports teams to gain a more accurate understanding of team members' emotional states in different work situations. Furthermore, appraisal of other team members' mood was less accurate in the PDT condition, indicating that it is more difficult to discern emotions of others in PDT, when less information is available due to the lack of social context cues compared to FTF communication.

Rather unexpected was the finding that almost 40% of test participants indicated that they did not consider the mood feedback for the evaluation of their fellow team members' mood. This is astonishing since those participants were still more accurate in their mood ratings of others compared to the participants in the condition without mood feedback. It can be assumed hence that they perceived the mood feedback in some unconscious way. It is planned for future research to evaluate whether team members are really not looking at the emotion representation (by means of an eye-tracking study) and why team members think or pretend that they do not consider the information provided by the EmotiBoard.

With regard to measures of performance, the results reported in this study are less clear and caution is advisable when interpreting the results. Although only marginally significant, a statistical trend indicated that teams in the PDT condition committed more errors than teams in the FTFT condition. This might be due to the lack of social context cues in PDT and corroborates previous findings indicating a decrease in performance in teams working remotely compared to FTF teams [15]. However, this difference occurred only if no mood feedback was available. When mood feedback was provided, error rates of teams working together in FTFT and PDT were very similar. This indicates that mood feedback is beneficial in PDT and leads to better performance. In contrast, teams working in the FTFT condition did not benefit from the mood feedback but committed even more errors when a mood feedback was available. This might be due to information overload or distraction, because team members in the FTFT condition, have already plenty of behavioral and non-verbal information about other team members' mood. The additional information on team members' mood provided by the system is largely redundant but commits additional attentional resources. Findings on subjective workload might be helpful to interpret this result. The interaction between mood feedback and communication mode for perceived temporal demand indicates a similar effect pattern as for the error rate: team members in the PDT condition felt more time pressure during task completion when no mood feedback was available whereas team members in the FTFT condition felt more temporal demand when the mood feedback was provided.

Data on team satisfaction indicated that team members working together in the PDT condition were more satisfied with teamwork and expressed a stronger preference for being in this group than team members in the FTF condition. This might be due to the fact that the confederate expressing a highly aroused bad mood was more distant in the PDT condition and hence had a smaller negative influence on measures of team satisfaction. Mood feedback however showed no influence on subjective measures of team satisfaction and team climate. This is somewhat astonishing since it was expected that knowing about other team members' mood would help to build and maintain positive relationships and facilitate conflict resolution [2]. There may be a number of reasons why the anticipated effect did not occur. The study made use of ad-hoc teams (i.e. team members had not known each other), which need some time to go through the typical processes of team building, such as forming relationships and mutual trust. Furthermore, the teams worked together on the tasks for rather a short period of time (M = 21min, SD = 12min). Finally, the team tasks in this study did not have a high potential for conflict. Since it seems that there have not been any conflicts during task completion, mood feedback did not facilitate conflict resolution and therefore could not have had a positive influence on team satisfaction and team climate. Future research employing more conflict-laden tasks may be needed to demonstrate that mood feedback has an influence on measures of team satisfaction and team climate.

Some limitations with regard to the generalization and interpretation of the results are acknowledged. It is important to note that the results of this study were obtained in a specific experimental setup, in which a confederate was expressing explicitly a negative, highly aroused mood. Although this did not lead, as expected, to high levels of conflict within the teams, the use of a confederate might still have had some influence on the results of the study, e.g. with regard to subjective measures of satisfaction, c.f. section 5.4. It would have been desirable to include also a confederate expressing an explicitly positive mood as well as a control group with no confederate in this study to have a more complete experimental design. Due to time and financial constraints, this was however not possible.

7 Conclusion

Overall, the results of this study indicate the usefulness of the EmotiBoard as a mood feedback tool, because it helped better understand other team members' mood and improved other outcome measures of team work. This is a very promising first result obtained with a tool that is still under development. A new version of the EmotiBoard is currently developed, which will automatically assess team members' mood, based on speech prosody and physiological data (skin conductance, heart rate variability). Future research still needs to determine whether such a tool would also work in a different cultural setting and different application areas (e.g. virtual teamwork, e-learning or online psychotherapy), however studies using a similar tool for self-feedback of affective states (AffectAura, c.f. [38]) or for honest signals in video conferencing [39] hinted already at the usefulness of such an instrument in similar application areas. The findings of this study are encouraging to continue the enhancement of the EmotiBoard to a team support system that automatically detects and represents moods in team work. This is because understanding mood and emotion is especially important within efficient teams, in particular with regard to difficult work situations such as intercultural teamwork [2,4].

Acknowledgements. This work was supported by the IM2 Swiss National Center of Competence in Research, a research instrument managed by the Swiss National Science Foundation on behalf of the Federal Authorities.

References

1. Iyer, A., Leach, C.W.: Emotion in inter-group relations. European Review of Social Psychology 19, 86–125 (2008)
2. Jordan, P.J., Troth, A.C.: Managing emotions during team problem solving: emotional intelligence and conflict resolution. Human Performance 17(2), 195–218 (2004)
3. Forgas, J.P.: Affect in social judgments and decisions: a multi-process model. In: Zanna, M. (ed.) Advances in Experimental and Social Psychology, vol. 25, pp. 227–275. Academic Press, San Diego (1992)
4. Ashkanasy, N.M., Zerbe, W.J., Härtel, C.E.J.: Managing emotions in the workplace. Sharpe, New York (2002)
5. Isen, A.M.: Positive affect and decision making. In: Lewis, M., Haviland-Jones, J.M. (eds.) Handbook of Emotions, 2nd edn., pp. 417–435. The Guildford Press, New York (2000)
6. Erez, A., Isen, A.M.: The Influence of Positive Affect on Components of Expectancy Motivation. Journal of Applied Psychology 87(6), 1055–1067 (2002)
7. Berkowitz, L.: Aggression: Its causes, consequences, and control. McGraw-Hill, New York (1993)
8. Fredrickson, B.L.: Positivity. Crown, New York (2009)
9. West, M.A.: Effective Teamwork: Practical Lessons from Organizational Research. John Wiley, Chichester (2012)
10. Goleman, D.: Emotional intelligence. Bantam Books, New York (1995)

11. Mayer, J.D., Salovey, P.: Emotional intelligence and the construction and regulation of feelings. Applied and Preventive Psychology 4, 197–208 (1995)
12. George, J.M.: Emotions and Leadership: The Role of Emotional Intelligence. Human Relations 53(8), 1027–1055 (2000)
13. Gardner, L., Stough, C.: Examining the relationship between leadership and emotional intelligence in senior level managers. Emerald 23 (2002)
14. Jordan, P.J., Ashkanasy, N.M., Härtel, C.E.J., Hooper, G.: Workgroup emotional intelligence: Scale development and relationship to team process effectiveness and goal focus. Human Resource Management Review 12, 195–214 (2002)
15. Hertel, G., Geister, S., Konradt, U.: Managing virtual teams: A review of current empirical research. Human Res. Mgmt. Rev. 15(1), 69–95 (2005)
16. Townsend, A.M., Demarie, S.M., Hendrickson, A.R.: Desktop video conferencing in virtual workgroups: anticipation, system evaluation and performance. Information Systems Journal 11, 213–227 (2001)
17. Press, L.: Low-cost estimation of travel trade-offs. Communications of the ACM 41, 17–20 (1998)
18. Bitti, P.E.R., Garotti, P.L.: Nonverbal communication and cultural differences: issues for face-to-face communication over the internet. In: Kappas, A., Krämer, N.C. (eds.) Face-to-Face Communication Over the Internet, pp. 81–99. Cambridge University Press, New York (2011)
19. Walther, J.B.: Visual cues in computer-mediated communication: sometimes less is more. In: Kappas, A., Krämer, N.C. (eds.) Face-to-Face Communication Over the Internet, pp. 17–38. Cambridge University Press, New York (2011)
20. Daft, R.L., Lengel, R.H.: Information richness: a new approach to managerial behavior and organization design. Research in Organizational Behavior 6, 191–233 (1984)
21. Short, J., Williams, E., Christie, B.: The social psychology of telecommunication. Wiley, London (1976)
22. Buxton, W.: Telepresence: integrating shared task and person spaces. In: Proceedings of Groupware 1991, Amsterdam, October 29, pp. 27–36 (1991)
23. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, London (1976)
24. Chen, M.: Leveraging the asymmetric sensitivity of eye Contact for videoconference. CHI Letters 4(1), 49–56 (2002)
25. Grayson, D.M., Monk, A.F.: Are you looking at me? Eye contact and desktop video conferencing. ACM Transactions on Computer-Human Interaction 10(3), 221–243 (2003)
26. Benford, S., Bowers, J., Fahlén, L.E., Greenhalgh, C., Snowdon, D.: User embodiment in collaborative virtual environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 1995, pp. 242–249 (1995)
27. Manstead, A.S.R., Lea, M., Goh, J.: Facing the future: emotion communication and the presence of others in the age of video-mediated communication. In: Kappas, A., Krämer, N.C. (eds.) Face-to-Face Communication Over the Internet, pp. 144–175. Cambridge University Press, New York (2011)
28. Zayonc, R.B.: Feeling and thinking: preferences need no inferences. American Psychologist 35, 151–175 (1980)
29. Kelly, J.R., Barsade, S.G.: Mood and emotions in small groups and work teams. Organizational Behavior and Human Decision Processes 86(1), 99–130 (2001)
30. Milgram, S.: Behavioral Study of obedience. The Journal of Abnormal and Social Psychology 67(4), 371–378 (1963)

31. Haney, C., Banks, C., Zimbardo, P.G.: Interpersonal Dynamics in a Simulated Prison. International Journal of Criminology and Penology 1(1), 69–97 (1973)
32. Morris, J.D.: SAM: The Self-Assessment Manikin – An Efficient Cross-Cultural Measurement of Emotional Response. J. Advert. Res. 35, 63–68 (1995)
33. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
34. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshakti, N. (eds.) Human Mental Workload, pp. 139–183. North Holland, Amsterdam (1988)
35. Hackman, J.R., Morris, C.G.: Group tasks, group interaction process, and group performance effectiveness. In: Blumberg, H.H., Hare, A.P., Kent, V., Davies, M. (eds.) Small Groups and Social Interaction, vol. 1, pp. 331–345. John Wiley, Chichester (1983)
36. Ishii, H., Kobayashi, M.: ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In: Bauersfeld, P., Bennett, J., Lynch, G. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1992), pp. 525–532. ACM, New York (1992)
37. Kenny, D.A., Kashy, D.A., Cook, W.L.: Dyadic data analysis. Guilford, New York (2006)
38. McDuff, D., Karlson, A., Kapoor, A., Roseway, A., Czerwinski, M.: AffectAura: an intelligent system for emotional memory. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems (CHI 2012), pp. 849–858. ACM, New York (2012)
39. Byun, B., Awasthi, A., Chou, P.A., Kapoor, A., Lee, B., Czerwinski, M.: Honest signals in video conferencing. In: Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, pp. 1–6 (2011)

Do Usability Professionals Think about User Experience in the Same Way as Users and Developers Do?

Torkil Clemmensen[1], Morten Hertzum[2], Jiaoyan Yang[3], and Yanan Chen[3]

[1] Department of IT Management, Copenhagen Business School, Denmark
tc.itm@cbs.dk
[2] Computer Science and Informatics, Roskilde University, Denmark
mhz@ruc.dk
[3] Institute of Psychology, Chinese Academy of Science, Beijing, China
{chenyn,yangjy}@psych.ac.cn

Abstract. In this paper, we study how usability professionals' thinking about system use relates to that of system developers and end users. We conducted 72 repertory-grid interviews to capture how usability professionals, developers, and users describe their system use. The participants in each stakeholder group were from China, Denmark, and India. Our results indicate that usability professionals focus on emotion-related aspects of system use, while users focus more on context in terms of utility and degree of usage. There are no interactions between stakeholder group and nationality, although both stakeholder group and nationality independently influence how participants think about usability and user experience. We recommend that to understand users' concerns, researchers should study context more.

Keywords: Usability professionals, UX professionals, user experience, stakeholder differences, cultural differences.

1 Introduction

In this paper, we focus on how usability professionals' thinking about system use is different from other stakeholder groups, in particular system developers and end users. By usability professionals, we mean people with titles such as usability professional, user experience manager, user experience specialist, user researcher, customer-experience architect, human factors specialist, interaction designer, usability engineer, and more. Differences between usability professionals and other stakeholders in their thinking about system use can be expected for several reasons. First, usability professionals' and system developers' thinking about system use may differ from that of end users, because professionals have systems as the objects of their professional work, whereas users apply systems to do their activities. That is, the "computer application, from the user's perspective, is not something that the user operates *on* but something that the user operates *through* on other objects or subjects." [1, p. 1]. Second, usability professionals' thinking about system use may differ from that of developers and users, because usability professionals will tend to be in an evaluation

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 461–478, 2013.

mode (trained to reflect explicitly about use), in contrast to system developers and users. Third, there are cultural differences in how people experience the use of interactive systems and products, and also in the way people respond to usability evaluation techniques, and in how they share their experiences of using a system [22].

The fit between the different stakeholder groups' view of system use is a basic raison d'être for the usability profession. A better fit between users and usability professionals in their thinking about system use, compared to users and system developers, has been the main argument for a usability profession since the beginning of the profession [20]. According to Putnam and Kolko [17, p. 2021] a HCI professional [their term] should be able to "walk in the users' shoes". The usability profession emerged because design-oriented developers did not have such empathy with the users.

Recently, two developments have made it imperative to study in depth how usability professionals think differently than users and developers about the use of systems. First, the usability professionals' identity has expanded into a much broader User Experience (UX) professional identity [3]. According to Dumas [3], some usability professionals with backgrounds in psychology and human factors perceived this development as a degradation of the skills of the profession, while other usability professionals accepted it as an expansion of the profession. To reconcile such incompatible beliefs, more knowledge about what distinguishes the thinking of a usability professional from that of developers and users is much needed.

Second, the globalization of the interactive systems industry has led to a worldwide quest for locally available usability and UX skills. The tasks of a global UX team can include designing a new interactive system for global users, localizing an existing system for national users, designing completely new system concepts for the local users, providing UX management for local developers, and more [19]. In such projects, the geographical location of stakeholders and the availability of specialized UX skills in a particular geographical location are important. UX skills are not distributed evenly across geographical locations but depend, among other things, on the mix of multidisciplinary university and professional training available locally [19]. According to Rosenberg and Kumar [19], local UX professionals may be needed to study domain specific requirements, while general issues like UI patterns and guidelines might be shared across countries. It appears to be a best practice in the computing industry to hire local professionals because they are assumed to be in some way different from professionals in other countries.

In this paper, we investigate differences between usability professionals, users, and developers in how they think about system use. We do this based on 72 repertory-grid interviews. We also examine whether nationality (China, Denmark or India) affects the differences between the three stakeholder groups. We use four UX classifications to make a content analysis of the interviews. This paper contributes with insights into the unique operational constructs held and used by usability professionals, compared to the operational constructs used by other stakeholder groups in systems development. These insights are important to usability professionals' practices when they design and evaluate UX. The insights should also improve communication about UX between stakeholder groups in systems development.

2 Background

Redish and Barnum [18] stated that from the late 1990s the usability profession opened up to technical communicators, information architects, interaction designers, content strategists, and others who were not part of the human factors or software engineering traditions from the beginning of the usability profession. As a part of this evolution, the usability profession gradually became a broader UX profession, culminating with Usability Professionals' Association (UPA) changing its name to User Experience Professionals Association (UXPA). Dray [2, p. 5] described how it became more difficult to identify a usability professional as one who practiced the usability test methodology: "In the past, usability could lay claim to a professional identity based on a particular set of practices or methodology—usability testing.". So what is it that distinguishes the usability professional from other IT professionals? Putnam and Kolko [17] proposed that we should study the differences between professionals who bear user-oriented job titles such as usability engineer and those who have design-oriented job titles such as developer. In a pilot study, they identified three professional groups: (1) design-oriented such as developers and designers, (2) UX oriented such as UX architects and usability professionals, and (3) management oriented. They found significant differences between the groups in their use of a range of methods to study and communicate about end-users, for example, the UX group reported a higher average use of contextual inquiry methods. They also found that the design-oriented group showed more empathy in how they considered users, compared to the UX oriented group, as measured by scales of emphatic concerns and personal distress [17]. However, the differences in empathy were not significant. Putnam and Kolko [17] concluded that we need to study differences, including regional differences, among the professionals who create products and those who conduct user research.

According to Lindgaard [15], usability professionals traditionally had their basic training in the behavioural and social sciences. Shackel [20] described the usability profession as being particularly focused on the evaluation of users' task completion with interactive tools in specified environments. Thus, while being a part of the design process, usability professionals' work was from the beginning of the profession defined as having to do with the analysis of user tasks and needs, and the forwarding of the resulting knowledge to developers. According to Lindgaard [15], the focus was on performance measures all the way from usability metrics of job task effectiveness and efficiency to broad business-level concerns. To the extent that the users' emotions have been in focus, the efforts have been directed toward helping the user avoid frustration, see for example [14]. Among others, Dumas [3] and Lindgaard [15] point out that although the concepts of attitude and satisfaction toward an interactive system were part of the classic understanding of usability, the broader field of user experience was still to come.

Early in the history of the profession, 'user' was the term used to denote the stakeholder group that interacted with the system to obtain some goal, and who did not have much interest in, or knowledge about, the technology itself. As Shackel [20, p. 339] put it: "For many users the informatics system is essentially the terminal or workstation which they are using, and that is the central computer as they see it". Thus, although the user was not attending to the system as such, s/he was acutely aware that she or he was using a system. In contrast, Shackel [20] claimed that developers tended to forget about the user and focus on the technological artefact. Shackel [20, p. 342] describe this: “Too easily this [setting goals and specifying attributes of the design] may be seen solely as the work of the designers… [but] to achieve usability in the ultimate design it is essential that users and managers are fully involved”. The stakeholder groups of user, developer, and usability professional were thus defined as being different in their relation to each other and to the technology.

Irrespective of the recent efforts of the HCI community to define usability and institute UX standards [10], people relate to systems through their personal experience and concepts. Hertzum, Clemmensen, Hornbæk, Kumar, Shi and Yammiyavar [9] analysed 48 repertory-grid interviews with developers and users to see how their personal usability constructs were affected by nationality (Chinese, Danish, or Indian) and stakeholder group (developer or user). For stakeholder groups, a significant overall difference between developers and users was found. The differences included that users associated ease of use with leisure and, conversely, difficulty in use with work-relatedness, while developers expressed the reverse associations. There was no significant overall difference across nationalities in participants’ thinking about usability.

Hertzum and Clemmensen [8] analysed how 24 usability professionals’ thinking about usability fitted with analytic usability definitions, and found that usability professionals made use of more utilitarian than experiential, i.e. UX related, constructs. This indicates that goal-related performance was central to usability professionals’ thinking about usability, whereas they had less elaborate sets of experiential constructs. The usability professionals tended to construe usability at an individual level, rather than at organizational and environmental levels. However, the usability professionals’ constructs went considerably beyond what was captured by the ISO 9241 definition of usability, thereby indicating a discrepancy between this definition of usability and the thinking of the professionals concerned with delivering usability.

We are unaware of studies in which usability professionals are compared systematically with other stakeholder groups. Thus, it remains unknown whether usability professionals think more like users than developers do. We need to know more about the usability constructs of usability professionals and other stakeholder groups, and in particular, how they may differ in the way that they think about user experience.

3 Method

In this paper, we merge the data from our two previous studies [8, 9] and, for the first time, provide an analysis of how usability professionals’ thinking about user experience compare to that of developers and users. As described above our previous studies have analysed usability professionals in isolation [8] and compared developers and users [9]. While the data in the current paper are re-used from the two previous studies, the analysis of the data is completely new.

We analyse repertory-grid interviews with 72 people to capture how they described their system use. The repertory-grid technique was developed by Kelly [11] as part of his personal construct theory. It has been successfully used in many studies of systems development, design, and use [7, 21]. We used local interviewers in the three involved countries to help make the data collection cross culturally valid.

3.1 Participants

The 72 participants were 24 usability professionals, 24 system developers, and 24 end users. Within each of these three stakeholder groups there were eight participants from China, eight from Denmark, and eight from India (see Table 1). To ensure that the participants were representative of the general cultural context of computer use in their country, participants from each country had to be a resident of the specific country and have been raised in the country. We recruited through our personal networks, and the participants were thus a convenience sample. All participants had good English skills. All participants used text-processing, web applications, and email every day, or nearly every day. They all stated that their computer hardware was sufficient, that they were comfortable with computers, and that they felt neither anxious nor relaxed/indifferent when they ran into a problem with a computer or application.

Table 1. Demographic information about the 72 participants

Group	**Gender**		**Age (years)**		**IT experience (years)**	
	Male	*Female*	*M*	*SD*	*M*	*SD*
Chinese usability prof.	5	3	26.6	±3.2	9.8	±2.4
Chinese developers	5	3	31.5	±1.9	10.6	±1.7
Chinese users	5	3	27.3	±1.9	8.4	±1.9
Danish usability prof.	3	5	32.5	±4.6	18.6	±6.1
Danish developers	5	3	36.6	±5.8	19.3	±5.8
Danish users	5	3	36.8	±6.2	16.9	±3.6
Indian usability prof.	7	1	29.9	±1.6	9.5	±2.7
Indian developers	8	0	29.6	±1.7	9.9	±2.5
Indian users	5	3	29.0	±4.0	7.0	±2.1

3.2 Procedure

The procedure was similar to the procedure proposed by Kelly [11]. Participants were interviewed individually at their workplace by a local interviewer; that is, a person with the same nationality as the participant. Participants filled out a questionnaire about their background and tried to elicit constructs with the repertory-grid technique on a couple of training tasks. Then the actual repertory-grid interview was conducted. It consisted of two steps: selection of systems and elicitation of constructs.

In selecting systems, the participant was asked to consider “the array of computer applications you use for creating, obtaining, revising, managing, and communicating information and documents in the course of your day-to-day activities.” We maintained a focus on the participants’ work by interviewing them at their workplace and by encouraging them to look for candidate systems at their workplace computer (in the start menu). Participants had to select a system within each of six categories:

my text processing system, my email, a useful system, an easy-to-use system, a fun system, and a frustrating system.

To elicit construct-contrast words, the participant was given groups of three of the selected systems and asked: “Can you think of some important way in which your personal experience using these three systems makes two of the systems alike and different from the third system?” Having indicated the two similar systems, the participant wrote down a short phrase that explained how these two systems were alike – the construct – and another short phrase that explained how the third system differed – the contrast. Then, a seven-point rating scale was defined with this construct/contrast pair as its end points, and the participant rated all six systems according to this rating scale. The construct-elicitation step was repeated for all twenty combinations of three systems, in random order, or until the participant was unable to come up with a new construct for two successive combinations. In this way, each participant elicited his or her personal grid of constructs.

Following local customs, Danish and Indian participants received no compensation for their participation in the study while Chinese participants were paid a small amount in RMB for their participation. Each interview lasted about 1.5 hours.

The repertory-grid interviews were conducted by three local interviewers: a Chinese, a Dane, and an Indian. The interviewers were all HCI researchers. We followed closely the prescriptions from Kelly [11] for doing repertory grid interviews. After a pilot interview, we met for sharing experiences and creating consensus about how to do the interviews, while maintaining the benefits of having local interviewers.

3.3 Data Analysis

Microsoft Word was the dominant choice of text-processing system among the participants, and Microsoft Outlook was the most frequently selected email system. For the four other types of system, there was more diversity among the selected systems. Most of the selected systems were software that is used all around the world. The selection of these systems did not appear to be biased by the participants’ nationality.

We analysed the elicited constructs by categorizing them according to four UX classifications, Table 2. The first classification distinguished between system, user, and context of use. The ISO standard for UX [10] lists these three factors as those that influence the subjective UX. The second classification contrasted subjective and objective experiences, by following the suggestion in Han [5, 6] that objective performance and subjective image/impression are both important dimensions. This classification was chosen because it gave detailed definitions of the two categories and could be expected to capture broadly any UX related constructs [5, 6]. In addition, the classification seemed to reflect the utilitarian/experiential dimension used by other researchers. The third and the fourth classifications were chosen to maximize the differences among the four classifications. We included Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, and Sinnelä [12]’s classification of UX into general relation to the system, attractiveness of the system, ease of use, utility, and degree of usage because all of these categories were defined with a focus on the IT system. In contrast to this system-oriented UX definition, we chose McCarthy and Wright [16]'s four threads of experience: compositional, sensual, emotional, and spatio-temporal, because they focus on human experience in general.

Table 2. The classifications used in categorizing the constructs

Classification	Category definitions
ISO 9241-210 user experience	
System-related	All components of an interactive system that provide information and controls for the user to accomplish specific tasks with the interactive system
User-related	Emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use.
Context of use-related	Other people, tasks, equipment, and the physical, social, temporal, organizational and cultural environments in which a product is used
Objective vs. subjective UX	
Performance	How well users perceive and interpret the interface of a system, how fast the users get used to the product and how well they remember it, and the users control activity and its results;
Image/Impression	Basic sense (the primitive and direct image/impression stemming from the design characteristics), description of image of a system, and evaluative feeling
System-oriented UX	
General relationship and user experience with the system	Any experience with systems that users find meaningful and important
Attractiveness of the system	General attractiveness (appeal) of the product in the users' own eyes and those of their friends; more than users' rational or practical experiences
Ease of use of the system	Product is easy and effortless to use
Utility of the system	Product serves an important function for the user
Degree of usage of the system	Degree of usage which affects user experience over time, related to quality of experience
Human experience of technology	
Compositional	The way that different elements of experience form a coherent whole; the narrative structure, action possibility, plausibility, consequences and explanations of actions
Sensual	The concrete, palpable (to take in your hand), and visceral (in flesh and blood) character of experience that is grasped pre-reflectively in the instant situation
Emotional	Value judgments (e.g., frustration and satisfaction) that ascribe importance to other people and things with respect to the user's needs and desires
Spatio-temporal	What draws attention to the quality and sense of space-time that pervades experience

We categorized all constructs according to the first classification before we proceeded to the second classification, then categorized all constructs according the second classification, and so forth. For each classification, the categorization of the constructs involved five steps. First, the first author thoroughly discussed the classification with to the two coders, authors three and four, in order to ensure a common understanding of the classification. Second, a randomly selected training set, consisting of 20% of the constructs, was categorized by each coder independently. Each construct was assigned either to one of the categories of the classification or to an 'other' category. Third, all disagreements in the coders' categorizations of the training set were discussed to reach consensus about the categorization of the constructs. The rationale for the second and third step was to create a shared understanding of how the classification was used. Fourth, the remaining 80% of the constructs were categorized by the two coders independently to be able to assess the quality of the categorization. Fifth, all disagreements in the coders' categorizations of these 783 constructs were discussed and a consensus was reached, in order to arrive at the final classification. used in the analysis.

The Kappa values of the agreement between the coders in their coding of the 783 non-training constructs were 0.55, 0.62, 0.61, and 0.68 for the ISO 9241-210 UX, objective vs. subjective UX, system-oriented UX, and human experience of technology classifications, respectively. Whereas all four Kappa values indicate statistically significant agreement, the value of 0.55 for the ISO 9241-210 UX classification is below the minimum threshold of 0.60 recommended by Lazar et al. [13]. Since this classification was important, and the agreement was approaching the threshold, we decided to keep the classification in the analysis.

4 Results

The 72 participants elicited a total of 977 construct/contrast pairs, corresponding to an average of 13.57 pairs per participant. The minimum number of construct/contrast pairs elicited by a single participant was three, the maximum 20. Below, we first analyse the constructs for effects of stakeholder group, then look into effects of the participants' nationality, and finally investigate interrelations across the classifications.

4.1 Differences between Stakeholder Groups

To analyse whether usability professionals had different user experiences than developers and users Table 3 gives, for each stakeholder group, the average percentage of constructs in a category across the participants in that stakeholder group. We used multivariate ANOVAs to test for stakeholder differences in the distribution of constructs across all categories in a classification and univariate ANOVAs for the individual categories. The statistical analysis was performed on the percentage distributions of the constructs (columns 2 to 4 of Table 3); this was done to assign equal weight to participants, irrespective of the number of constructs elicited by a participant. Before conducting the statistical analyses, the percentage values for each participant were arcsine transformed because percentages cannot be assumed normally distributed [4]. All pair-wise comparisons reported below were Bonferroni adjusted to compensate for multiple comparisons.

For the classification of the constructs into system-, user- and context-related ISO 9241-210 UX there was a significant effect of stakeholder group, Wilks' $\lambda = 0.77$, $F(8, 132) = 2.35$, $p < 0.05$. While there were no differences across stakeholder groups for the percentage of system-related constructs (e.g., "Address register/No address register"), $F(2, 70) = 0.99$, $p = 0.38$, and other constructs (e.g., "Copyright protected (license needed)/Download from the internet for free"), $F(2, 70) = 0.19$, $p = 0.83$, there were significant differences across stakeholder group for user-related constructs (e.g., "Lot of fun to use/Hard job"), $F(2, 70) = 5.55$, $p < 0.01$, and for context- related constructs (e.g., "Communication tool/Work tool"), $F(2, 70) = 6.36$, $p < 0.05$. Pair-wise comparisons showed that usability professionals had a higher percentage of user-related constructs than both developers ($p < 0.05$) and users ($p < 0.01$); developers and users did not differ significantly from each other. In contrast, usability professionals had a *lower* percentage of context-related constructs than users ($p < 0.001$).

For the classification of the constructs according to objective vs. subjective UX there was a significant effect of stakeholder group, Wilks' $\lambda = 0.81$, $F(6, 134) = 2.47$, $p < 0.05$. While there were no differences across stakeholder groups for the percentage of performance constructs (e.g., ""Efficiency/Productivity"), $F(2, 70) = 1.79$, $p = 0.18$, and other constructs (e.g., "Environment driven...peer group team members involved in its usage/Personal, self-driven"), $F(2, 70) = 2.06$, $p = 0.14$, there was a significant difference across stakeholder group for image/impression constructs (e.g., "Overwhelming interface [means that the interface has too many features]/Enjoyable, rich interface"), $F(2, 70) = 5.60$, $p < 0.01$. Pair-wise comparisons showed that usability professionals had a higher percentage of image/impression constructs than developers ($p < 0.05$) and users ($p < 0.01$); developers and users were not significantly different from each other.

The classification of the constructs according to system-oriented UX showed a significant effect of stakeholder group, Wilks' $\lambda = 0.63$, $F(12, 128) = 2.82$, $p < 0.01$. We found significant differences for the categories of attractiveness (e.g., "Nice-looking UI/Ugly UI"), $F(2, 70) = 7.83$, $p < 0.01$, and utility (e.g., "For playing movies/Do scientific analysis"), $F(2, 70) = 6.87$, $p < 0.01$. Pair-wise comparisons showed that usability professionals had a higher percentage of attractiveness constructs than developers ($p < 0.05$) and users ($p < 0.01$); developers and users were not significantly different from each other. In contrast, usability professionals had a *lower* percentage of utility constructs than users ($p < 0.01$). There were no differences across stakeholder groups for general relationship and user experience (e.g., "Bright and fluid [means good to look at…organic]/Static [means very rigid in structure…not organic]"), ease of use (e.g., "Demands more clicks/Demands fewer clicks"), degree of usage (e.g., "Frequent use/Seldom use"), and other (e.g., "Install package is small/Size is too large"), Fs$(2, 70) = 2.35$, 1.50, 0.94, and 0.94 respectively (all ps > 0.1).

For the classification of the constructs according to human experience of technology there was a significant effect of stakeholder group, Wilks' $\lambda = 0.76$, $F(10, 130) = 1.95$, $p < 0.05$. While there were no differences across stakeholder groups for the categories compositional (e.g., "Can be used independent of other applications/Have to be used with other applications"), sensual (e.g., "Complex/Simple"), spatio-temporal (e.g.

"Available everywhere/May not be available everywhere"), and other (e.g., "On cd/Downloaded"), Fs(2, 70) = 1.15, 1.33, 0.65, and 1.01, respectively (all ps > 0.2), there was a significant difference in the percentage of emotional constructs (e.g., "Live, active/Dead"), $F(2, 70) = 5.80$, $p < 0.01$. Pair-wise comparisons showed that usability professionals had a higher percentage of emotional constructs than developers and users (both ps < 0.05).

Table 3. Frequency (N = 977 constructs) and percentage (N = 72 participants) of constructs within each classification, averaged across participants from the same stakeholder group

Category	Frequency	Usability profs		Developers		Users		Total	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ISO 9241-210 UX *									
System-related	306	33	±18	34	±14	27	±20	31	±17
User-related**	204	30	±19	19	±15	16	±12	22	±17
Context-related**	427	33	±18	44	±20	52	±19	43	±20
Other	40	4	±5	3	±6	4	±6	4	±6
Objective vs. subjective UX *									
Performance	414	38	±14	47	±21	44	±13	43	±*17*
Image/impression**	353	44	±13	34	±16	31	±14	36	±15
Other	210	18	±10	20	±15	25	±14	21	±13
System-oriented UX **									
General relationship and user experience with the system	77	11	±8	7	±9	6	±8	8	±9
Attractiveness of the system**	75	14	±13	6	±6	4	±6	8	±10
Ease of use of the system	178	22	±11	17	±16	16	±13	18	±14
Utility of the system**	405	30	±16	41	±23	54	±20	42	±22
Degree of usage of the system	46	3	±4	7	±7	4	±4	5	±5
Other	196	20	±14	22	±17	16	±15	20	±16
Human experience of technology *									
Compositional	547	53	±14	59	±17	57	±14	57	±*15*
Sensual	95	12	±8	10	±9	8	±7	10	±8
Emotional**	94	15	±11	7	±6	7	±9	10	±9
Spatio-temporal	178	15	±11	18	±14	20	±11	18	±12
Other	63	5	±6	5	±8	7	±9	6	±8

$^{*}p < 0.05$, $^{**}p < 0.01$, indicate significant differences between stakeholder groups.

4.2 Nationality Differences

We tested for interactions between nationality and stakeholder group using multivariate ANOVAs, but found no overall interactions between nationality and stakeholder group for any of the four classifications. Stakeholder group and nationality appeared to be two independent factors, each influencing how participants think about UX. Regarding nationality, Table 4 gives, for each nationality, the average percentage of constructs in a category for a single participant.

The international standard for UX lists three factors as those that influence subjective UX (system-, user-, and context-related). For the classification of the constructs into these three ISO 9241-210 UX categories there was a significant effect of nationality, Wilks' $\lambda = 0.54$, $F(8, 132) = 5.88$, $p < 0.001$. There were significant differences across nationality in the percentage of system-related constructs, $F(2, 70) = 12.44$, $p < 0.001$; user-related constructs, $F(2, 70) = 3.63$, $p < 0.05$; context-related constructs, $F(2, 70) = 3.30$, $p < 0.05$; and other constructs, $F(2, 70) = 5.68$, $p < 0.01$. Pair-wise comparisons showed that Chinese participants had a higher percentage of system-related constructs than Indian and Danish participants (both *p*s < 0.001) and a lower percentage of user-related ($p < 0.05$) and, tentatively, context-related ($p = 0.05$) constructs than Indian participants. Chinese participants had a higher percentage of other constructs than Indian participants ($p < 0.01$).

Table 4. Percentage of constructs in the categories of each classification, averaged across participants with the same nationality, $N = 72$ participants

Category	Chinese		Danish		Indian	
	M	*SD*	*M*	*SD*	*M*	*SD*
ISO 9241-210 UX ***						
System-related***	44	±16	26	±12	25	±17
User-related*	15	±11	24	±13	27	±22
Context-related*	35	±17	46	±14	48	±26
Other**	6	±6	4	±6	1	±3
Objective vs. subjective UX						
Performance	44	±17	38	±14	47	±18
Image/impression	40	±16	39	±11	30	±17
Other	16	±8	24	±12	23	±18
System-oriented UX ***						
General relationship and user experience	6	±7	11	±9	8	±9
Attractiveness	7	±6	8	±9	10	±13
Ease of use	22	±12	13	±10	20	±16
Utility*	32	±16	44	±19	50	±27
Degree of usage***	5	±6	7	±5	2	±4
Other***	29	±15	17	±14	12	±12
Human experience of technology*						
Compositional***	59	±14	49	±13	62	±16
Sensual	12	±10	11	±6	7	±7
Emotional	8	±6	11	±9	11	±12
Spatio-temporal**	13	±9	24	±13	16	±12
Other	9	±9	5	±7	4	±7

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The classification of the constructs according to objective vs. subjective UX showed no effect of nationality, Wilks' $\lambda = 0.85$, $F(6, 134) = 1.84$, $p > 0.05$. Thus, Chinese, Danish, and Indian participants did not display a different distribution of their constructs. The percentages of performance and image/impression constructs tended to be roughly similar.

The classification of the constructs according to system-oriented UX showed a significant effect of nationality, Wilks' $\lambda = 0.55$, $F(12, 128) = 3.73$, $p < 0.001$. Significant effects were found for three categories: utility, $F(2, 70) = 4.36$, $p < 0.05$, degree of usage, $F(2, 70) = 6.88$, $p < 0.01$, and other $F(2, 70) = 9.71$, $p < 0.001$. Pair-wise comparisons showed that Indian participants had a higher percentage of utility constructs than Chinese participants ($p < 0.05$), and that Danish participants had a higher percentage of degree-of-usage constructs than Indian participants ($p < 0.01$). Chinese participants had a higher percentage of other constructs than Danish ($p < 0.05$) and Indian ($p < 0.001$) participants. As much as 29% of the Chinese participants' constructs were not covered by the system-oriented UX classification.

For the classification of the constructs according to human experience of technology there was a significant effect of nationality, Wilks' $\lambda = 0.72$, $F(10, 130) = 2.38$, $p < 0.05$. There were significant differences in the percentages of compositional constructs, $F(2, 70) = 5.00$, $p < 0.01$, and spatio-temporal constructs, $F(2, 70) = 5.32$, $p < 0.01$. Pair-wise comparisons showed that Indian participants had a higher percentage of compositional constructs than Danish participants ($p < 0.01$), and that Danish participants had a higher percentage of spatio-temporal constructs than Chinese participants ($p < 0.01$).

As stated above, we did not find any overall interaction effects between stakeholder group and nationality. Since this was an interesting and surprising result, we decided to analyse whether the influence on UX of stakeholder group and nationality were also independent for each category of the classifications. We found only two interaction effects when looking at the individual categories of the UX classifications. First, for the ISO 9241-210 classification, there was a significant interaction effect between nationality and stakeholder group for system-related constructs $F(4, 68) = 2.95$, $p < 0.05$. Chinese users and Chinese usability professionals used more system-related constructs than the same stakeholder groups from India and Denmark, while for developers there was no difference. Second, for the objective vs. subjective UX classification, there was a significant interaction effect between nationality and stakeholder group for *other* constructs $F(4, 68) = 4.16$, $p < 0.01$. Users from India were quite different from developers and usability professionals in the degree to which they construed UX in terms categorized as "other". The objective vs. subjective UX classification was more exhaustive for all Chinese and Danish stakeholder groups.

4.3 Interrelations of Constructs across Classifications

To analyse the interrelations between categories in different classifications, Table 5 shows the ISO 9241-210 classification crossed with the three other classifications.

Comparing the ISO 9241-210 classification with the Objective vs. subjective UX classification, the differences across stakeholder groups were unevenly distributed across the Objective vs. subjective categories. The usability professionals had overall most subjective *image/impression* constructs (**139**/105/109), and had in particular more *user related* (**55**/26/32) subjective *image/impression* constructs, compared to the other stakeholder groups. They had fewer *context-related* constructs in the *performance* (**35**/73/58) and *other* (**46**/78/61) categories, compared to users and developers.

Comparing the ISO 9241-210 classification with the system-oriented UX classification, the usability professionals had more *user-related* constructs, in particular about *attractiveness* (**36**/13/17) and *ease of use* (**32**/20/26). The usability professionals also had fewer *context-related* constructs, in particular about *utility* and *degree of usage*.

Comparing the ISO 9241-210 classification with the human experience of technology classification, the usability professionals had more *user-related emotional* constructs. Again, the usability professionals had fewer *context-related* constructs, in particular *compositional* constructs, compared to the other stakeholder groups.

Table 5. Cross-tabulation of stakeholder groups' ISO 9241-210 and the other UX classifications

	Subjective vs. Objective UX			**System-oriented UX**						**Human experience of technology**					**Total**
	Objective performance	Subjective image/impression	Other	General relationship	Attractiveness of system	Ease of use	Utility	Degree of usage	Other	Compositional	Sensual	Emotional	Spatio-temporal	Other	
ISO 9241-210 UX															
Usability professional															
System	48	55	3	16	2	24	29	0	35	76	24	0	3	3	106
User	31	55	4	14	36	32	5	0	3	35	9	42	2	2	90
Context	35	26	46	4	2	8	63	11	19	55	3	2	45	2	107
Other	5	3	5	0	0	4	1	0	8	2	1	0	0	10	13
Total	119	139	58	34	40	68	98	11	65	168	37	44	50	17	316
User															
System	44	49	1	5	0	24	33	0	32	63	18	0	2	11	94
User	27	26	2	11	13	20	8	1	2	22	3	22	2	6	55
Context	73	27	78	3	1	5	138	11	20	106	1	3	61	7	178
Other	10	3	3	1	0	6	4	1	4	7	2	0	2	5	16
Total	154	105	84	20	14	55	183	13	58	198	24	25	67	29	343
Developer															
System	57	49	0	10	4	21	24	1	46	74	22	1	6	3	106
User	23	32	4	8	17	26	5	1	2	25	6	21	4	3	59
Context	58	23	61	4	0	5	94	19	20	81	1	3	50	7	142
Other	3	5	3	1	0	3	1	1	5	1	5	0	1	4	11
Total	141	109	68	23	21	55	124	22	73	181	34	25	61	17	318
Total	414	353	210	77	75	178	405	46	196	547	95	94	178	63	977

5 Discussion

5.1 Usability Professionals Differ from Other Stakeholder Groups

We find three main differences between usability professionals and the other stakeholder groups in their thinking about usability and UX: Usability professionals focus more on user-related constructs and subjective UX than developers and users, and they focus less on context-related constructs than users, as illustrated in Fig. 1 which shows selected information from Table 3. Below we discuss these differences.

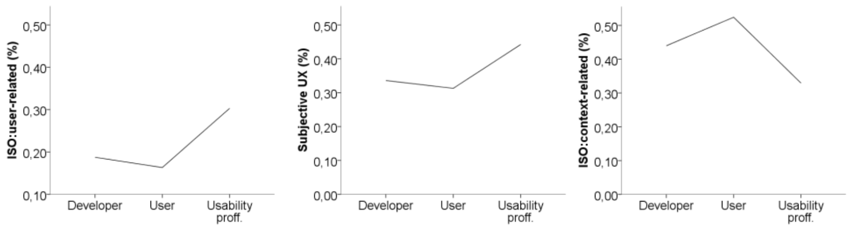


Fig. 1. Usability professionals focus more on users and subjective UX, less on context

Usability professionals focus more on user-relatedness and subjective UX than developers. This supports the main rationale behind the profession, which is that usability professionals compared to developers have more empathy with users and their situation [20]. The implication of this is that while developers might need to be convinced to become more user centred, usability professionals are ready for more advanced theory and techniques that go beyond simply being user centred.

Usability professionals focus more on user-relatedness and subjective UX than users. This finding is somewhat against the idea of usability professionals having the same view on system use as users. Usability professionals may focus more on subjective UX, emotions, attractiveness and so forth than users, because they as professionals have a greater range of constructs to describe and distinguish between UX-related emotions. The distinction between objective performance and subjective image/impression by Han [5] appears on the surface to be similar to a distinction between utilitarian and experiential constructs. However, in this study usability professionals attended a lot to subjective aspects of UX, whereas Hertzum and Clemmensen [8] found that usability professionals made use of more utilitarian than experiential constructs. A utilitarian/experiential distinction divides usability into a utilitarian factor concerned with goal-related performance and an experiential factor concerned with satisfaction [8]. Satisfaction is, partly, about freedom from discomfort, that is, when there is no frustration there may be satisfaction [14]. In contrast, in the objective/subjective classification, the subjective dimension was inspired by Kansei-engineering and focused on the emotions actually experienced by the user: "…collecting the subjective feelings of a product in words, extracting the key feelings…" [5, pp. 478-479]. Furthermore, the subjective dimension covers both the immediate image/impression that the user gets from interacting with the product in that moment, the user's description of impressions of an interaction, and their judgment and attitude toward this interaction [6]. Thus, a different set and range of emotional processes involved in interacting with computers may be covered by the subjective image/impression category, compared to an experiential dimension.

Usability professionals focus less on context-related constructs than users. We speculate that the difference in focus on context-relatedness may be related to the difference in focus on subjective UX. Users may be concerned with getting the work done, a context-related concern, to a larger extent than usability professionals, who in contrast focus more on the emotions experienced during use, a less context-related concern. According to Hertzum et al. [9], users associate ease of use with leisure and difficulty in use with work-relatedness, while developers express the reverse associations. The usability professionals in the present study rarely distinguished between work use and other use (e.g., "Working Tools/For Fun"). Only three of the 107 context-related construct/contrast pairs elicited by usability professionals concerned a work-leisure distinction. A possible explanation for the near absence of a distinction between work and leisure in the usability professionals' constructs may be that they consider UX-related emotions relevant to leisure as well as work. If a system is sufficiently mature in that most performance and utilitarian issues have been taken care of, the way to improve the system further may be to focus on the emotional aspects of the user experience, irrespective of whether the system is for work or leisure. This explanation may apply to our study because most of the systems selected by the participants were mature products, such as MS Word and Outlook. In addition, usability professionals may be more business-oriented than users in their thinking about usability and UX. Though we did not find any direct evidence of this, usability professionals may tend to think about how the user experience can give a product a competitive edge on the market, and therefore value emotional over context-related constructs.

5.2 Global Software Development and Inclusive UX Definitions

Stakeholder group and nationality independently influence how people think about UX, as illustrated in Fig. 2. This means that usability professionals are different from the other stakeholder groups, irrespective of nationality. The lack of significant interaction effects between stakeholder group and nationality suggests that the ways in which usability professionals differ from local user groups will be somewhat similar across national borders. This result is promising for the global IT industry because it indicates that usability professionals' competences may be globally applicable, thereby meeting a need pointed out by the global IT industry, see [19]. However, since we found clear main effects of nationality – people from different countries think differently about their system use – it seems to vary from one cultural context and country to another what exactly usability professionals do when they, paraphrasing Putnam and Kolko [17, p. 2021], "…walk in the [local] users' shoes".

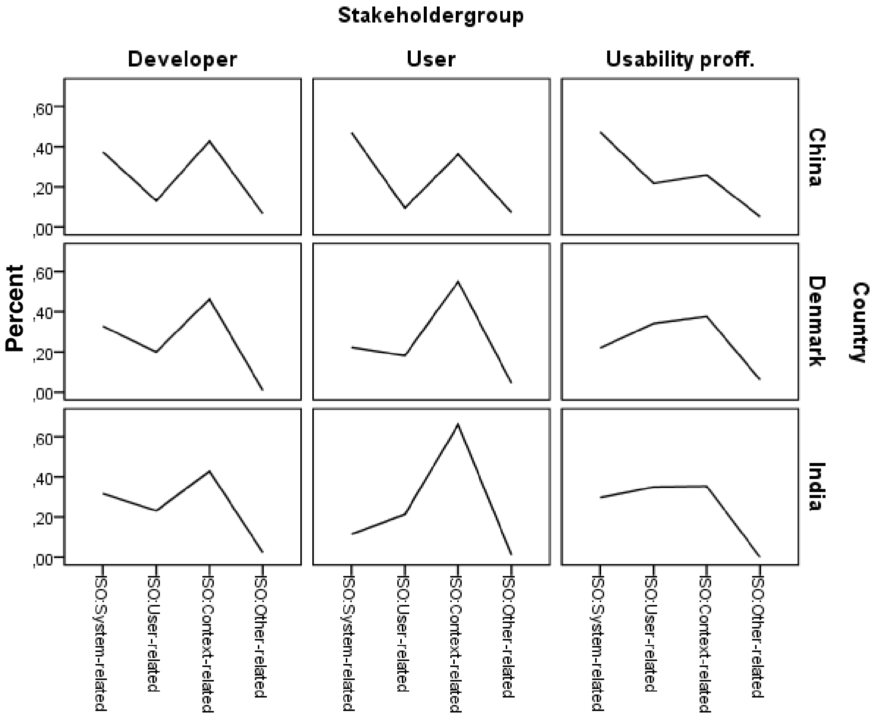


Fig. 2. Effects of stakeholder group and nationality on the ISO UX classification (%)

The present study confirms that UX is an inclusive concept. In the ISO definition, UX is a broad concept that aims to capture most, if not all, of what people experience when interacting with computer systems. ISO [10, p.3] defines the user experience as a "person's perceptions and responses resulting from the use and/or anticipated use of a product". The four UX classifications were all good or very good at capturing the 977 constructs used by the 72 participants to describe their everyday system use. All four UX classifications had better coverage (average 87%) compared to the coverage found in an earlier study [8] for the ISO usability definition (average 53%). This study provides empirical support that selected UX classifications, including the ISO definition of UX, match the variation in how core stakeholder groups think about their system use.

5.3 Limitations

Three limitations should be remembered in interpreting the results of this study. First, the stakeholder groups of developer, user, and usability professional evolve over time. In particular, the usability profession may be evolving into a UX profession with a broader or different focus. Hence, we characterize usability professionals' thinking about system use at a time when their profession is in a process of discussing its identity. Second, the Danish participants were older than the Chinese and Indian participants were, and the Indian participants were gender skewed toward males. The age difference may reflect that in Denmark people are relatively older when they finish their education. However, we acknowledge that these issues related to the sample and recruitment of participants could have affected our results. Third, with respect to the developers and usability professionals there may be a gap between their personal constructs and their professional knowledge. While we asked participants for their personal constructs, their descriptions of their user experiences might be influenced by explicit definitions of usability and UX, learned during their education rather than experienced during the use of the selected systems.

6 Conclusion

We find differences in how usability professionals think about their user experiences, compared to developers and users. Therefore, if the usability/UX profession in the future wants to continue to claim that they are the users' advocate, it may be worthwhile to review the profession's key constructs. To meet users' concerns, it may be time for HCI researchers to revisit the context of use and focus more on the physical, social, temporal, organizational, and cultural environments in which a product is used. In addition, people's nationality influences all stakeholder groups' thinking about system use, according to our study of a sample of Danish, Chinese, and Indian people. Finally, the four UX classifications used in this study have emerged as inclusive concepts that captured nearly all the ways in which our participants thought about system use. This inclusiveness is encouraging from the point of view of devising analytic UX conceptualizations that encompass most of the variability of different stakeholders' actual user experiences.

Future research may investigate how varying levels of professional experience influence usability professionals' thinking about UX. The usability professionals participating in this study were intermediate-level to experienced usability professionals. Their thinking about UX may gradually have moved away from that of users and developers. Novice usability professionals may think more like users, or they may more directly apply textbook definitions of usability and UX.

Acknowledgements. This study was co-funded by the Danish Council for Independent Research. Thanks to Institute of Psychology, Chinese Academy of Science, Beijing, for hosting the first author during data analysis in 2012. Thanks to Jyoti Kumar, Qingxin Shi, Xianghong Sun, Pradeep Yammiyavar, and the interviewees for taking part in this study.

References

1. Bødker, S.: Through the Interface: A Human Activity Approach to User Interface Design. Lawrence Erlbaum Associates Inc., Hills dale (1990)
2. Dray, S.M.: Engaged Scholars, Thoughtful Practitioners: The Interdependence of Academics and Practitioners in User-Centered Design and Usability. Journal of Usability Studies 5(1), 1–7 (2009)

3. Dumas, J.: The Great Leap Forward: The Birth of the Usability Profession (1988-1993). Journal of Usability Studies 2(2), 54–60 (2007)
4. Fleiss, J.: Statistical Methods for Rates and Proportions. Wiley, New York (1981)
5. Han, S.H., Hwan Yun, M., Kim, K.J., Kwahk, J.: Evaluation of Product Usability: Development and Validation of Usability Dimensions and Design Elements Based on Empirical Models. International Journal of Industrial Ergonomics 26(4), 477–488 (2000)
6. Han, S.H., Yun, M.H., Kwahk, J., Hong, S.W.: Usability of Consumer Electronic Products. International Journal of Industrial Ergonomics 28(3), 143–151 (2001)
7. Hassenzahl, M., Wessler, R.: Capturing Design Space from a User Perspective: The Repertory Grid Technique Revisited. In: International Journal of Human-Computer Interaction, pp. 441–459 (2000)
8. Hertzum, M., Clemmensen, T.: How Do Usability Professionals Construe Usability? International Journal of Human-Computer Studies 70(1), 26–42 (2012)
9. Hertzum, M., Clemmensen, T., Hornbæk, K., Kumar, J., Shi, Q., Yammiyavar, P.: Personal Usability Constructs: How People Construe Usability across Nationalities and Stakeholder Groups. International Journal of Human-Computer Interaction 27(8), 729–761 (2011)
10. ISO 9241-210: Ergonomics of Human System Interaction-Part 210: Human-Centred Design for Interactive Systems (Formerly Known as 13407). International Organization for Standardization (ISO). Geneva, Switzerland (2009)
11. Kelly, G.A.: The Psychology of Personal Constructs. Theory and Personality, vol. 1. Norton, New York (1955)
12. Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., Sinnelä, A.: UX Curve: A Method for Evaluating Long-Term User Experience. Interacting with Computers 23(5), 473–483 (2011)
13. Lazar, J., Feng, J.H., Hochheiser, H.: Research Methods in Human-Computer Interaction. Wiley, Chichester (2010)
14. Lazar, J., Jones, A., Shneiderman, B.: Workplace User Frustration with Computers: An Exploratory Investigation of the Causes and Severity. Behaviour & Information Technology 25(3), 239–251 (2006)
15. Lindgaard, G.: Early Traces of Usability as a Science and as a Profession. Interacting with Computers 21(5-6), 350–352 (2009)
16. McCarthy, J., Wright, P.: Technology as Experience. Interactions 11(5), 42–43 (2004)
17. Putnam, C., Kolko, B.: HCI Professions: Differences & Definitions. In: CHI 2012 Extended Abstracts, pp. 2021–2026. ACM, New York (2012)
18. Redish, G., Barnum, C.: Overlap, Influence, Intertwining: The Interplay of UX and Technical Communication. Journal of Usability Studies 6(3), 90–101 (2011)
19. Rosenberg, D., Kumar, J.: Leading Global UX Teams. Interactions 18(6), 36–39 (2011)
20. Shackel, B.: Usability-Context, Framework, Definition, Design and Evaluation. Interacting with Computers 21(5-6), 339–346 (2009)
21. Tan, F.B., Hunter, M.G.: The Repertory Grid Technique: A Method for the Study of Cognition in Information Systems. MIS Quarterly 26(1), 39–57 (2002)
22. Walsh, T., Nurkka, P., Walsh, R.: Cultural Differences in Smartphone User Experience Evaluation. In: MUM 2010, Limassol, Cyprus, (Article No. 24). ACM, New York (2010)

Here or There? How Configuration of Transnational Teams Impacts Social Capital

Julia Katherine Haines, Judith S. Olson, and Gary M. Olson

Donald Bren School of Information and Computer Sciences
University of California, Irvine, USA
{hainesj,jsolson,gary.olson}@uci.edu

Abstract. The many challenges of distributed communication and the many challenges of intercultural collaboration have been researched and discussed at length in the literature. What is lacking is a combined approach that looks at both issues of distance and diversity in collaboration. We conducted research in a large, multinational technology company to better understand team configurational factors in transnational work. In this case study, we found that the development of social capital is impacted by whether a person is in their home context or transplanted and their expectations based on that context. This has implications for the development of intellectual capital in the team. We highlight factors in the creation of social capital as well as some mechanisms that may mitigate cultural difference. In addition to bringing into focus the challenges that arise in various configurations, this study contributes to the transnational literature by highlighting the importance of local context in diverse collaborations.

Keywords: Transnational collaboration, team configuration, social capital.

1 Introduction

Many organizational structures today are truly global. Companies find talent all over the world, place employees in all sorts of geographic locations, and coordinate teams, both distributed and co-located [1]. This sort of transnational work requires crossing all sorts of important boundaries, including temporal, spatial, and cultural. Much has been written about the multitude of issues in working across spatial and temporal distances [2-6]. And a great deal of research has focused on transnational or cross-cultural differences as they related to the workplace [7-9]. Yet, given the prevalence and significance of transnational work in today's globalized world, there appears to be a lack of empirical studies that focus on the variety of ways such collaboration can be structured and the consequences of different configurations [10-13]. Global organizations can choose to configure and structure their teams in a wide variety of ways, and it is important to look at the costs and benefits associated with these choices.

First, why might you want to have diverse cultural configurations? Research that has focused on the value of diversity in groups illustrates that, for many tasks, diversity trumps homogeneity [14] by allowing for a diversity of perspectives, which aid in a

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 479–496, 2013.

number of ways, such as problem solving, information aggregation, and prediction. Other studies have indicated that cultural diversity leads to increased creativity and satisfaction of experience [15], and with the right mechanisms in place, can result in better cohesion, trust, and innovation [16]. Most work in this area suggests, however, that the net benefits of such diversity must account for the costs related to differences in cognitive, behavioral, and affective aspects [13]. Diversity can most certainly be an asset, but it can also a liability if not structured and managed well.

Second, from an organization's perspective, why might you choose to have a team distributed over continents and time zones as opposed to co-located in one area? As Castells and others have suggested, today's knowledge-based workforces are defined by their flexibility and adaptability [17], and it is quite possible to configure teams in many ways [18]. Benefits to globally distributed collaboration include garnering expertise in local markets, cost reductions, and a 24-hour development cycle [19, 20]. But, there are still many challenges for distributed teams in terms of processes, tools, and structuring [1, 21]. And, there are particular difficulties for distributed teams that cross cultures [13, 22].

With all the permutations possible, it is important to understand the implications of various transnational team configurations. In our research, we found that the most salient implications are related to interpersonal relationships. Thus, we use the theory of social capital in framing our findings. This theory highlights the importance of interpersonal relationships, which serve as a valuable resource for social action within a network [23]. Nahapiet and Ghoshal's work on social capital in the organization shows that social relationships within a firm, which are the foundations of social capital, are intricately tied to the firm's ability to create new intellectual capital, or knowledge [23]. This theory serves well as a lens to investigate the transnational configurations within firms, enabling insight into interpersonal relationships at the individual level as well as insight into overall team and organizational functioning.

In this paper, we contribute to the transnational literature by examining how the configuration of global teams and the context of an individual's location within that team impact social capital. We seek to examine how distributed and co-located configurations of transnational teams are similar and different in terms of the social relationships that are developed, and in turn, how that impacts the individual and the team. We recognize, as others do [11, 24], that one's culture is not static, and that although culture influences one's behavior, behavior is also influenced by the situation. Over time people adapt to their context, either converging toward the local culture or diverging. Thus, we focus on these central ideas of location and configuration.

2 Method

We conducted research within a large, multinational high-technology company, which we will refer to by the pseudonym PrimaTech. PrimaTech is headquartered in the US, but has offices in more than 40 countries worldwide. The structure of the organization is complex and dynamic, with many teams distributed all over the globe.

2.1 Participants

We specifically chose to focus on participants from India and China, as a significant portion of PrimaTech's employees come from these two countries. Participants were independently recruited at multiple work sites using a purposive sampling methodology [25]: we broadcast our study requirements on various listservs, and when people responded, we selected them according to their background, work situation, and availability for interview. Among the 22 we selected, nine of whom were female, all were of Indian (13) or Chinese (9) national origin currently working in either the US or offices in their home country, and all participants had previously or were currently collaborating in a transnational team. Thirteen were based currently in the US, nine in offices abroad (Bangalore, Beijing, and Hyderabad). Six had spent two or more years in the US offices, eight less than two years, and eight spent no time at US offices. Table 1 shows these participant aspects in more detail.

2.2 Interviews

Prior to data collection, we conducted extensive background research to better understand common cultural characteristics amongst the various national cultures we considered for the study. In all, we created a bibliography of 88 sources. From this meta-analysis, we created an informational table that included aspects of culture in the US, China, India, Japan, Korea, and Singapore along 29 dimensions, including behavioral elements, cognitive and perception differences, and macro-level social conventions. This table was used to inform the interview script. We conducted pilot interviews with contacts of Chinese and Indian origin and refined the scripts according to their feedback, which included areas for further inquiry as well as suggestions on how to approach certain delicate topics.

The interviews themselves were conducted in a semi-structured format to encourage participants to openly discuss their perspectives and reflect on experiences. We collected information on participants' backgrounds and their work at PrimaTech, including team collaboration methods and tools, their cross-cultural experiences, and perspectives on differences in communication and social conventions. We visited the company's US headquarters and conducted a portion of the recruiting there, but all interviews were conducted remotely via video-conferencing. Interviews lasted from around 45 minutes to over an hour. Participants were not compensated.

2.3 Teams

Teams at PrimaTech are often comprised of people of many nationalities, so participants were not necessarily working primarily with American-born colleagues, but truly transnational teams. We use the term transnational here to indicate that the teams were composed of two or more nationalities, indicating diversity that goes beyond just different cultures, which can occur within nations [13]. In this research, we look both at transnational teams with participants in their *home countries* of India and China

Table 1. Characteristics of Participants

T/R*	Gender	Nation	Loc.	Yrs at Co.	Yrs US Office	Role	# Offices work with	Time on team (yrs)	Team size**	Some Members Co-located
T1	M	India	US	3	0.5	Manager	4	< 1	L	Y
T2	F	China	US	8	8	Infrastructure	7	> 1	L	Y
T3	M	China	US	6	6	Engineer	1	> 1	S	Y
T4	F	China	US	6	0.25	Consumer Operations	5	< 1	L	Y
T5	M	China	US	6	6	Engineer	2	< 1	L	Y
T6	M	India	US	5	2	Engineer	3	> 1	M	Y
T7	M	China	US	6.5	6.5	Analyst	2	< 1	S	Y
T8	M	India	US	4	0.25	Engineer	3	< 1	L	Y
T9	M	India	US	6	1.5	Manager	6	> 1	L	Y
T10	F	China	US	1.25	1.25	Engineer	5	> 1	M	Y
T11	F	India	US	6.5	3.5	Manager	5	> 1	L	Y
T12	M	India	US	3.5	0.1	Manager	5	< 1	L	Y
T13	M	China	US	4	1.25	UX	10+	< 1	L	Y
R1	F	India	India	1	0	Planner	2	< 1	S	N
R2	F	India	India	6.5	0	Optimizer	2	< 1	S	N
R3	M	India	India	1.6	0	Planner	3	< 1	S	N
R4	F	India	India	1	0	Optimizer	3	< 1	S	N
R5	M	India	India	0.75	0	Planner	3	< 1	M	N
R6	M	India	India	5.5	0	Manager	4	< 1	L	Y
R7	M	India	India	1	0	Consumer Operations	3	< 1	S	Y
R8	F	China	China	6	0.25	Analyst	2	1	M	Y
R9	F	China	China	2.5	0	Engineer	2	< 1	S	Y

* T=Transplant, R=Remote; **Team size (L = 20+, M = 11-20, S = 4-10)

working remotely with those in the US and other countries and transnational teams in which Indian and Chinese participants *moved to* the US to be co-located with their teammates. We label these participants in two groups. Since the terminology of migration is often problematic and inaccurate [26], we use the term "Transplants" for the latter. Borrowed from the language of ecology, this term indicates a move from one environment to another, making it more contextually appropriate. By contrast, "Remotes" refer to participants working from their home country, in these cases, based in Bangalore, Beijing, or Hyderabad.

Some structural differences between these two groups should be noted at the ouset. First, the average number of team members was somewhat different between the two groups. Remotes reported around ten members being involved in their current team, whereas Transplants frequently reported much larger team sizes. However, there was not a discernible difference in the number of team members with whom participants worked directly. Most worked closely with five or fewer colleagues. Second, Transplants reported quite varied tenures with their current teams, ranging from two months to over five years, while Remotes had all worked with their current team for a year or less. This is representative of recent growth in PrimaTech's workforce. The number of offices with which they regularly collaborated was also variable. Transplants typically worked across more than four offices, but all had some co-located colleagues. Remotes, on the other hand, worked across three or fewer offices on average. Moreover, they were often the only member of their team based in their office. So, while Transplants were mostly part of the hub in a hub-and-spoke or hub-to-hub model, Remote participants were often spokes in this sort of model [27]. These structural differences further distinguish the two configurations, but we focus primarily on the context of location, as it had a more important impact on social relationships.

2.4 Analysis

Although we were interested in comparing Transplants with Remotes, we did not have strong hypotheses about what differences would emerge. Therefore, we used grounded methods [28] in our initial analyses. We used the software Dedoose to aid in data analysis while engaging in an abductive, iterative process of comparing data and relevant research literature and theory. We began with open-coding to discover patterns and recurring themes in our initial round of interviews, then created memos to refine our ideas. This guided data collection as well as our targeted sampling. We then conducted closed coding, applying lower-level codes to categories to identify emergent concepts. We found that much of what emerged related to differences in expectations involving social relationships, and thus we used Nahapiet and Ghoshal's theory of social and intellectual capital in framing our findings [23].

3 Social Capital and Configuration

Many points of interest emerged from our interviews, but most salient were interpersonal relationship differences among Transplants and Remotes. In our findings, we discuss the elements of social capital, which Nahapiet and Ghoshal decompose into three dimensions: the structural, the cognitive, and the relational dimensions. These dimensions of social capital in turn impact the conditions for exchange and combination of intellectual capital [23]. It was out of the scope of this research to investigate the structural dimension within PrimaTech. The structural dimension is concerned with network connections between actors, which would require social network analysis. Thus, we focus primarily on the cognitive and relational dimensions. First, we touch on the differences in expectation between participants in the two configurations. Then we highlight the factors that play a role in shaping cognitive and relational social capital, respectively, and discuss how these impact the ability of the team to create new intellectual capital. We also discuss implications of social capital at an individual, team, and organizational level. Finally, we introduce some mechanisms that appear to mitigate these challenges.

3.1 Configurations and Expectations

Transplants to the US noted very different experiences with team colleagues than Remotes. The ways in which the two different groups of participants perceived their interpersonal relationships with colleagues were, in part, predicated upon their expectations. In turn, those expectations were based upon their location and the team configuration. Those who had moved to the US generally noted disappointment in their lack of social bonds with teammates. On the other hand, remotes approached these relationships with a pragmatism that showed reserved expectations.

Transplants. Those who moved to the US came for a variety of reasons, personal and professional, but amongst the various expectations, social relationships were a priority.

But as other narratives of migration [29] have shown, lived experiences often do not meet expectations, especially when it comes to a sense of social belonging. One reason for this is that social relationships at work are often strong in India and China, but are much more compartmentalized in the US [30]. Thus Transplants often experience moving from a situation with close relationships and strong social capital to one in which there seems to be a void. One participant, who had worked first in her home country of India before coming to the US highlights this differentiation: "When I used to work in India, work relationships were much more friendly. The demarcation between work colleagues and friends is a lot fuzzier…Here, it's just that they are not friends- we don't meet up as families and don't socialize." (T11) While this is typically recognized beforehand, as one prepares to move, it still serves as a shock to many Transplants. They move from having quite rich social relationships with coworkers to having much weaker social capital, regardless of their tenure, as we discuss below. Though this difference in work/life compartmentalization is usually known before a Transplant arrives, there is still an expectation of building work-based social relationships. Expectations may be dimmed a bit, but they are still there. The transplants we interviewed did work to cultivate relationships, yet many noted tribulations with creating social relationships, even after years at the same location. The troubles of expatriates integrating into host communities is not a new concept [31]. But it is worthwhile to note that even with the knowledge of work/life difference and even after attempts at trying to create stronger social relations, expectations still led to disappointments.

Transplants desired stronger relationships, but were unsure of how to do it: "It's difficult to get along with the rest of the team as well as I could." (T11) And often, this led to feelings of distress or regret: "I should do more. I go out and have dinner, but sometimes it feels like it's really hard to fit in, no matter how hard I try." (T2) In many instances, this led to feelings beyond a desire for social engagement, into a deeper level of identity. One participant, who had Anglicized his Chinese name, felt as if colleagues did not know who he is: "They only see the way I work. They see 'working [name].' My Chinese name is hard to pronounce… So, they don't know my name. They don't see my social aspect, so they are not comprehending me. Work is just one slice of me." (T13) For those who have just moved, this struggle to create interpersonal relationships is especially salient. But, notably, this lack of social capital does not appear to improve with time spent at the US office. One participant, who has been in the US for over 20 years and working at PrimaTech's US headquarters for over eight (T11), indicated a high level of frustration with this aspect of her work.

Remotes. In contrast, Remote participants indicated that while they realize they cannot fully socialize with distributed colleagues, they felt that they do have close enough relationships. This highlights an adjustment in expectations; distance tempered expectations of social relationships. Employees working on distributed teams at PrimaTech typically meet all colleagues with whom they work closely within the first three to four months of being on a project, so most Remote participants had had the opportunity to meet their colleagues, either through travel to other offices, or through their colleagues visiting the office in India or China. Participants cited that these visits gave them a better sense of their colleagues. Beyond the bonds created through visits, information and communication technologies (ICTs) served as the way through which they maintained social connections. Several participants said they talk informally with distributed colleagues over collaborative technologies such as chat or in the first minutes of a video conference: “We do a lot of that- talking about movies, what we’ve been doing over the weekend. I feel like I know them as colleagues.” (R3) These participants generally felt that their distributed colleagues had accurate perceptions of them as colleagues, though nothing approaching deep connections as “friends.”

In sum, expectations based on team configuration and location within that configuration played a role in perceptions of social relationship. Transplants expected deeper friendships than they got; Remotes found a shallower but sufficient sense of friendship. Next, we take a look at factors that play a role in whether social capital is developed and how that impacts intellectual capital.

3.2 Social Capital Factors and Intellectual Capital in Transnational Teams

When working across cultures, some behaviors will converge, some will diverge. As Hinds et al. note, there must be some degree of convergence for collaboration to work [11]. In our research, a few factors stood out as limiting convergence, particularly for Transplants. We highlight these factors through the framework of social capital, first focusing on the cognitive dimension, then the relational, though both are interrelated to an extent. In delineating these factors, we investigate how they impact conditions necessary for exchange and combination of intellectual capital in the organization.

Cognitive Social Capital. Nahapiet and Ghoshal describe the cognitive dimension of social capital as shared cognition that is facilitated through shared codes and language as well as shared narratives [23]. Shared codes and language are important for the establishment of social relations and thus the ability to exchange information, but beyond that, they also influence perception, meaning the perceived value in sharing information and the ability to combine it meaningfully. Similarly, shared narratives impact interpretability of information.

Small Talk/Topics of Interest. “Small talk,” as Goffman explains, is a bracket to larger social affairs that has bearing on the overall relation of participants in an interaction [32]. Thus, it is an important mechanism for establishing stronger capital. Engaging in small talk is a ritualistic interaction that quite easily occurs within groups that have a shared cognitive base. However, engaging in small talk can be quite difficult across cultures due to lack of shared codes and narratives. In our interviews with PrimaTech employees, many Transplants reported difficulties with small talk, noting that they often don’t understand the topics at hand. One participant explained: “People go out to lunch every day and have a lot to talk about. It’s hard to get it all if it’s not work related.” (T10) She went on to explain she realizes that her different background influences the perception of what is interesting, so she finds it easier to chat with other Chinese, rather than struggle to find interesting topics to discuss with her American colleagues. For this participant, the lack of common topics was so salient that it influenced her to end lunchtime socializing, further isolating her from her teammates.

In contrast, many Remotes reported not only engaging in small talk in the beginning moments of a video conference, but finding enjoyment in it. So, while Transplants reported finding topics of conversation and small talk difficult, Remotes generally reported small talk as a great way to engage with colleagues for a few moments at the beginning or end of a meeting. This difference relates back to expectation issues, particularly with regard to the length of time one must socialize; having a short chat at the beginning of a video conference does not create a burden, whereas sustaining small talk for an hour of lunch may appear daunting. Similarly, Remotes frequently noted that they talked about movies, travel, and weekend plans as topics of conversation, which are fairly general and not highly location-specific, whereas Transplants had higher expectations placed upon them. Transplants noted pressure to know more about local sports or politics, both of which require more of a shared cognitive framework. It may also be the case that Transplants have to engage in this talk with many more people, potentially from many various backgrounds, feeling more pressure yet. One Transplant (T12) noted that so many people ask how his weekend was on Mondays that he plans out what he wants to say in advance, on his way to work. Thus, it appears that whether one only occasionally chats with teammates versus having regular, extended contact with them affects one's perception of whether small talk is a problem or a benefit in building social capital.

Cultural References. Whether part of small talk or part of general work discussions, culturally-specific references can impact one's ability to understand or engage in conversation. This can cause disengagement from social interactions, and, worse, can impact the ability to participate fully in meetings. Cultural references can create misunderstandings of the message itself, but beyond that, they can create confusion about tone. Several Transplants noted that technical conversations were typically fine until cultural references were applied within them, such as a person referring to an idea being "out in left field." This participant elaborates: "I think they tend to use examples from games or football references, so sometimes can't tell if they're being sarcastic or not. It becomes difficult." (T12)

On the other hand, Remotes generally used cultural references as a kicking-off point for understanding their colleagues better. One participant explained: "What is nice and what is fun for us —watching cricket here—is like baseball, basketball there. So watching sports connects people. When someone talks passionately about a match, I understand because I go equally bonkers over cricket." (R2) These Remotes reported that if they do not understand a cultural reference, this allows them to then ask questions to get to know each other better. One possibility for these different attitudes toward the use of cultural references is that there may be an implicit assumption that people will understand most references if they are living in the same location. On the other hand, when talking with someone who lives halfway around the world, individuals may hold fewer assumptions of what others do or do not know about another's culture. Remotes, it seems, are expected to not know everything; Transplants, however, are co-present and expected to know what other locals know.

This may be intensified by the fact that those in the US headquarters typically work with a higher number of distributed offices. So, as one participant points out, you must know about many more cultures: "I didn't socialize so much this past year just because I myself haven't fully plugged in to American culture. And my work and my mind's resources have limited me from doing that. I have to be distributed in many places with all the offices we work with, so I didn't spend much time to figure out local things here." (T13) This highlights that, beyond transnational context, the configuration of the team and the participant within the team really does matter in terms of creating social capital. Expectations from others and from oneself are shaped by location, and in turn, location influences the development of social capital. Lack of shared language, codes, or narrative can serve as a kicking-off point for sharing, which enable conditions for exchange, while assuming these are shared closes off the possibility for exchange.

Humor. Humor is an intensely social experience that provokes strong emotions [33]. As such, it can serve as a building block for social capital. But, humor is notoriously difficult to understand across cultures. One Transplant explained: "The most difficult thing here is to make a joke... It connects to many things out of the workplace…But a joke is very important for people's emotional connection. And I cannot connect that way. That is most difficult." (T13) Both Transplants and Remotes reported difficulties with humor, but Remotes had a more generally positive attitude about its impact. Several Remotes explained that joking around became easier for them to handle once they had met their colleagues in person: "Sometimes I don't know how to respond if someone is sarcastic… but everything is easier once you meet- like jokes. Because I've met him and know he's doing it for fun, that barrier gets broken." (R7) Thus, this lack of shared cognition impairs both configurations, but to a different degree. While humor is still difficult for both, it appears that Transplants, being located together, may attempt to understand humor on a deeper level, while Remotes seem satisfied to identify only when humor is being used, as it helps to understand context and tone.

Relational Social Capital. The relational dimension of social capital centers on the interactional and behavioral elements between parties. Nahapiet and Ghoshal describe this dimension in terms of trust, norms, obligations, and identification. They illustrate how these impact intellectual capital by influencing access to parties, perceived value and the motivation to exchange and combine information [23].

Trust. There has been much research on basic trust differences across cultures [34, 35]. Core values play a significant role in base levels of trust as a cognitive factor, but trust is also a dynamic property of relationships in context. Rocco et al., for instance, found a positive relationship between non-work communication and trust [36]. So, trust both shapes and is shaped by interpersonal relationships; when ties with work colleagues are not strong, a strong level of trust is not manifest and vice-versa. Thus, we found trust to be an issue both before and after a relationship has been established.

Despite trust being an oft-cited component of PrimaTech's organizational culture, trust among colleagues was complicated. One Transplant explained the differences in levels of trust with colleagues in the US versus India in the following way: "The meaning of a relationship here is different… In India you don't distinguish between social friends and professional ones. You trust all friends equally. Here it is very professional… I won't trust any friend here as I would in the Indian office." (T12) From this perspective, it appears trust can be gauged *a priori*. But others mentioned that trust requires rapport, thus it is also impacted *a posteriori.*

These complexities of trust fundamentally affect group work, whether co-located or distributed. However, trust issues were mentioned more frequently among Transplants than Remotes. One school of thought indicates that this might be due to "swift trust." Jarvenpaa and Leidner found a lack of cultural effects in looking at trust in global virtual teams [37], and they suggested these teams can rely upon trust that is based more on categorical expectations than interpersonal relationships, known as swift trust [38]. For Transplants, other factors impact interpersonal relationships, which then impacts trust *a posteriori*. For Remotes, though, trust is based more on team members' work effort and less on personal factors and is only expected to be sufficient, not deep. This is enhanced by visits in which brief contact allows for a sense of good rapport. While this does not mean that trust from Remotes is necessarily stronger in any sense, it may serve as less of an impediment as compared to co-located teams that may struggle with interpersonal issues.

Speaking Up: Norms of Openness. PrimaTech's organizational culture is often described as being very "bottom-up." Being "bottom-up" means that change is often initiated in a very democratic, communal way. As such, being involved in discussions and decisions means embracing the norms of openness and speaking up. This social norm of openness is noted as a key feature needed in knowledge-intensive firms [39].

"Speaking up" was the most frequently self-cited focus of behavioral change for participants in both groups and a repeated piece of advice they would give to others. One participant, coming from India noted: "It's difficult to get your head around it because if you've never worked in place like [PrimaTech], it's engrained into you to pay attention to social strata or not to speak up." (T11) This general lack of "speaking up" in meetings creates a disparity in being involved. While this happens for both co-located and remote due to different macro-level social conventions, Transplants have the opportunity to speak up more often. But, being in a new location and not having strong social relations in other ways impacts that. One Transplant even noted her move had a negative effect on speaking up, despite it being a norm at PrimaTech: "I have been less outspoken, more quiet since coming to [US office]. Part of that is because I'm transferring to a new location and I don't know people that well, so my confidence level isn't as high. I think more before I speak..." (T4)

For Remotes, speaking up is also an issue, but due to constraints, teams often structure formal conversations differently. For instance, in video conferences, distributed teams focus more on rotating participation and feedback from the various parties, as large video conferences with spontaneous input from simultaneous audio streams can become chaotic. In the same vein, there is an expectation in remote meetings that people may hold back a bit more; one manager noted that he makes a point of specifically asking what others think in making decisions, noting that often that person would have "a brilliant point of view that was missed." (R6) Remotes tended to recognize speaking up as a potential issue in advance and address, while Transplants' co-located teams perceived no barriers to speaking up, and thus didn't address it.

Notably, outside of formal conversations, those who are co-located also have the opportunity to communicate through a variety of informal channels that are beneficial to being involved in decisions, such as side conversations, hall talk, and "meetings over the water cooler" [8]. But when social relationships with colleagues are not strong, Transplants will miss out on such opportunities to hear and to have their opinions heard informally, which leads to less knowledge exchange.

Feedback. Within these transnational teams, lack of communication sometimes led to less access and motivation to exchange and combine knowledge. In particular, many participants noted how they often do not know when they have made mistakes because colleagues are hesitant to provide negative feedback. These behaviors are in contrast to the literature, which indicates that Americans are direct, while Chinese and Indians use indirect speech in order to promote harmony [40]. An Indian participant explained how this is impacted through social relationships, not just hierarchy issues: "This is confounded based on how close you are with colleagues. People get more direct negative feedback in India. Here, with a lot mistakes, you don't get direct feedback. I've had to seek it out. But one reason might be that I've only been here going on four months, so I don't have those relationships yet perhaps." (T1)

Additionally, some participants expressed an inability to understand when their colleagues needed help and feedback. Weak social relations with colleagues made it difficult to know when a team member needed help, unless it was directly requested: "If you want to help a coworker, it's not easy…I don't know how much they need help. . . I wish I had the conception to see when help can be appreciated or accepted." (T13) Again in this instance, there is not necessarily an advantage for Remotes, but communication among Remotes and their colleagues was more explicit, so there were fewer issues and less confusion related to implicit communication norms.

Meals. Breaking bread together is a meaningful social bonding activity. Yet this is a difference for those in the US offices in two ways. First, those who had worked in their home country before moving to the US cited that lunch is quite social in India or China, so it was a surprise to see colleagues in the US eating at their desks: "A lot of people eat lunch alone and keep working here. In China, I would always go out of the way to meet other people for lunch. It's a much more social experience there. You never eat at your desk." (T4) This highlights one of the unmet expectations of Transplants, who sought more social engagement with teammates, but did not always find it. Moreover, when these participants did go out to share meals with colleagues, some were disappointed with the food choices as well as the focus on drinking, which discouraged them from joining. One participant explained that team dinners were uncomfortable because he did not eat meat or drink alcohol. Yet, teammates would often choose restaurants for team dinners based on a good wine list. He noted this limited his participation in team dinners: "You want to give them company and build a relationship, build your network, but you're hungry." (T12)

By comparison, dinners are always part of the agenda when Remotes visit their teammates, or when teammates visit them. While less frequent, everyone typically attends, precisely since opportunities to do so are limited, and several Remotes cited good experiences getting to know their distributed counterparts this way.

Families. Finally, getting to know colleagues' families was a significant difference. Most Transplants cited their US-based colleagues' strong work/family divide, noting how different this was from their home culture. This supports the literature on this topic, which references how familial collectivism is more common in Asian cultures and is expressed in business relationships through socializing with families, among other things [11]. Participants based in the offices in India and China often mentioned socializing with colleagues' families. While it is not possible for this to be much of a part of distributed teammate bonding, several Remotes did cite introducing visiting colleagues to their own families, a point of difference based on distance limitations.

3.3 Impact on the Individual

When transnational teams work together, they face many issues that may impede success. As we have seen so far, lack of social capital impacts a team's information exchange and knowledge creation. But importantly, it also impacts the individual dramatically, which in turn can also affect the team.

Career Development. PrimaTech has a promotion process that is heavily influenced by one's peers, which is a unique aspect of the company that all employees encounter. Though most people appreciated the fundamental idea of this process, many participants expressed some level of concern about the importance of relationships. The idea of colleagues, rather than just managers, having a significant impact on career advancement is of much greater concern when one feels he or she does not have strong social capital with colleagues. This was understandably more of a concern for Transplants, who had difficulties building these relationships. One participant remarked: "Making friends is relevant to career growth. It's very difficult to do. You have to talk about children's soccer games and golfing, and I'm not interested… but in order to get people to help and evaluate you, you have to." (T2)

This policy is employed company-wide, so those at international offices review and are reviewed by not only their co-located peers, but their distributed teammates as well. Interestingly, Remotes did not indicate strong concerns with this process, indicating that anxiety about this process is perhaps induced by the local context of Transplants. Remotes may not have developed very strong social capital with colleagues abroad, but expectations for this remain low, and thus they feel their work, not their relationships, is what is being evaluated. Moreover, Remotes straddle different social worlds and arenas and thus have strong social capital in their home location. Even though their US-based colleagues may have more weight in reviews, these local relationships perhaps provide some reassurance in terms of feeling social cohesion, and thus less anxiety about the process. Transplants, in contrast, have major ties with their US-based colleagues, which lead to different expectations and different things expected of them.

General Satisfaction. Ostensibly, participants' enjoyment of their job was also influenced by local contexts. Previous research has shown mixed results in terms of satisfaction in diverse teams [15], but in general supports the claim that distributed, virtual teams have lower satisfaction [41]. Despite this, Remotes seemed more satisfied in certain ways, likely due to better social circumstances. Remotes, who unanimously noted close friendships with colleagues in their home country office, mentioned that they would hang out and socialize more in the office. They were more willing and interested in staying around the office for a while, even though they often worked long hours to accommodate distributed teammates on other continents.

In contrast, Transplants often cited the challenges of their local context. One participant gave this report of dissatisfaction through this solemn advice: "Be prepared that people will misunderstand you. And though you're in [PrimaTech] for five years in India, you have to prove yourself again over here… You don't think of how hard it can be. If you are not used to staying alone or fighting alone, don't join." (T12)

4 Mechanisms That May Aid in Convergence

What emerges from the findings of this study is the importance of focusing on the local in global work. Team configuration and the context of location play important roles in transnational collaboration and have significant implications for teams and individuals, particularly in how a team maintains social cohesion and a sense of community. This in turn impacts the conditions necessary for the exchange of intellectual capital. But, as Hinds et al. note, there must be some degree of convergence for collaboration to work at all [11]. Despite the many points of difference, there are mechanisms that did help teams collaborate. Here, we present ideas as to what might help develop social capital in these transnational collaborations. Emerging from our interviews, we find that an organizational "cultura franca" may help create stronger bonds, and that hiring practices may influence the possibility of convergence. In other words, organizational culture may mitigate some effects of different national cultures.

4.1 A Cultura Franca

The idea of a "cultura franca" is an inchoate idea that emerged in this study. It is based on the concept of a lingua franca, or bridge language, used to communicate between parties with different mother tongues. We use the term cultura franca, rather than organizational culture, to indicate that the strong, distinctive culture of PrimaTech serves as a bridge across diverse global cultures and creates a space for common goals and values. Like a lingua franca, it serves as a vehicle for mutual intelligibility--in this case, of practices and priorities, rather than language. Additionally, just as pidgin languages based on a lingua franca evolve in certain locales, the term cultura franca indicates that the culture is not completely uniform across all the geographies; it becomes imbued with attributes of the location.

Throughout PrimaTech, the cultural values of the organization are evident. All participants had very similar ideas about the organizational culture and frequently used the same, or similar, terms to describe it. The most oft-cited descriptors were "open" or "transparent," followed by "democratic," "independent," and "trusting." Many also described the culture in terms of things being "data-driven" and people being "smart" and "friendly." While these were considered to be universal qualities of the organization, there were mixed indications as to how similar the culture is across offices. Because PrimaTech's offices are very similar in physical environment, structure, and character across the globe, many referenced these as indications of uniformity. So, many believed that the culture was quite uniform: "They have done an amazing job of spreading the culture all over the globe…It's pretty much the same." (T9)

But on the other hand, many others indicated that, while the culture is promoted the same way across the globe, the way it is actually exhibited in different offices is not necessarily the same: "From people I've met, the [PrimaTech] culture is the same. Although the way we go about it is completely different." (R7) It appears that, though the values that permeate PrimaTech are fairly universal, the way in which they impact work relationships depends more on the location and local culture therein. The social environment is often adapted to fit better with the local culture. For instance, participant R2 noted that, unlike in the US, alcohol is not involved in office activities in

India. One participant describes these variations: “The basic things are similar across the geographies, but it’s a bit different because of other cultural differences… But in work ethic, it is the same... ” (R4) The cultura franca, then, serves as a mechanism to bridge what might otherwise be major differences in values. It seems that the cultura franca of the organization may perhaps ease some factors that negatively impact collaboration, while still enabling diverse perspectives, but to what extent it enables convergence remains a question.

4.2 Selection in Hiring

The development of this sort of organizational culture is in part predicated upon the population that is initiated into it. A factor that clearly emerged from our interviews was how the organization selects for those who embody the qualities that PrimaTech promotes. This is a fairly obvious factor in mediating intercultural difference—selecting people who are similar in certain ways—but from our research, the importance of this process was quite distinctive at PrimaTech, enabling the organization to more readily mitigate intercultural challenges. While most participants did not indicate that they considered the organization to be culturally American or Western, one participant explained that the company chooses people who are more culturally aligned to certain values, and that this enables easier collaboration: “At [PrimaTech], it is very different from other Indian companies. I have friends in [other companies], … and they have a lot more issues. I think this is due to [PrimaTech’s] interview process. Each person is specifically selected for how they will match...” (R3) While this selection may or may not be explicit—or true on a broader scale—the process of choosing certain qualities is quite explicit, at least internally. One participant explained how there were “unwritten rules” in the hiring process, noting the importance of the social: “Can you sustain a two-hour layover at an airport and not get bored and want to talk to this person?” (R6) He noted that this creates a basis –“the common denominator”—for having somewhat similar employees throughout PrimaTech’s global locations. Thus, while not necessarily being specific, there is a template for what is sought out socially in an employee. And since this process is held worldwide, it contributes to cultivating a workforce that reflects the values of the company, irrespective of national culture or other cultural traits.

5 Discussion and Conclusions

We conducted research within a large multinational technology company to better understand the mechanisms by which people conduct transnational work. While common sense says that presumably co-located work is easier and better for teams, the issues are more complex, especially in intercultural collaborations. From our interviews we found that social capital is, in part, impacted by whether the person is in their home context or transplanted, which in turn impacts the individual, the team, and the organization.

As our results show, location and configuration played a significant role in creating expectations, so it is also important to discuss how these expectations relate to lived experiences. Stronger and richer social capital were expected by Transplants; this led to disappointment when those expectations weren't fulfilled. This brings up issues of adaptation. In the literature on migration, it seems that findings generally indicate that the adaptations are made to accommodate the host culture. This was only supported to some extent in our interviews. Transplants adapt at some surface level, but not necessarily at a deeper level: "For myself, I have changed a bit over time. When I came to US, I still had the typical way Chinese do things. But now it is probably more mixed. To do many things, I do things closer to the Western way. But in thinking, I am still pretty Chinese." (T5) While we found only minor evidence of adaptation in terms of social bonding, it remains unclear to what extent participants' colleagues adapted to their behavior, clouding the issue further.

Previous research on immigrants has shown that surface-level adaptations may be made through the use of "relational templates" and that people may maintain multiple templates in engaging in intercultural work. [42] This concept works fairly well in explaining Remotes' interactions with distributed teammates: they maintain their capital with colleagues in their home office and colleagues in distributed offices in systematically different ways. So, you work with your colleagues in Bangalore in one way, but when you call the London office, for instance, you use a different behavioral pattern. While participants did not heavily cite cultural training provided by the organization, it seems that this may be another mediating factor—being able to clearly select and apply appropriate relational templates. Templates seem to be a useful concept in thinking about distributed work. However, with the case of Transplants, it is difficult to see how a template structure would hold, making adaptation more difficult. As indicated in the interviews, the US offices are quite multicultural, unlike offices such as Bangalore or Beijing, which tend to be quite nationally homogenous. When constantly surrounded by a local, heterogeneous team, it is impossible to have a single template that is appropriate for everyone. It is beyond the scope of this study to evaluate the degree of adaptability or co-adaptability of the team; however it seems the field of transnational research would greatly benefit from a more micro-level, longitudinal analysis of changes in behaviors based on the context of location.

Implications. Our results raise interesting questions for the configuration of teams. Are the social costs of moving around the globe more than the coordination costs of working at a distance? This has interesting implications for organizational structure, individual choices, and seemingly, technology.

A team is not merely a division of labor, so it is important to be reflective about organizational structure. Distance issues in teams are well known, but with the highly-systematic methods of interaction in distributed team, one wonders whether Remotes and these teams are reaching their full potential in such configurations. On the other hand, you have situations where workers cross borders to be close to their team. This may ameliorate many of the issues that arise with distance, but it does not mean that conditions for the exchange of intellectual capital will necessarily improve. As we have seen, the interactions that create strong social capital within a team do not necessarily change with proximity. Solid social relationships are the substrate-- the basis on which a team can thrive. When looking at issues of dispersion—temporal, spatial, configurational—it is necessary to also heavily consider how this interacts with cultural differences.

Beyond this, though, it seems more emphasis could be placed on organizational culture and human resources practices. On a practical level for the individual, it could be that, despite all its challenges, distributed collaboration provides a more optimal, structured format for intercultural work in that it requires less adaptation and concern for divergent cultural challenges. The limited contact inherent in distributed collaboration means that workers in their home country may experience more satisfying social interactions than those who move across borders to other offices. Put another way: consider, for a moment, PrimaTech's headquarters to be Rome, and how it might be easier to act like a Roman for short chunks of time than to try to become a Roman. But is this better for the team and the organization? If diversity is important, we want to embrace diverse perspectives. Maybe everyone should not have to act like a Roman. Or, maybe Romans should be expected to act differently. These factors would best be addressed through organizational culture and human resources training.

Finally, what are implications here for ICTs? It seems that, for Remotes, technology provided a solution that went "beyond being there" [43] in certain ways. The use of ICTs helped mitigate issues and supported the development of sufficient social capital. In this sense, perhaps what is needed is a return to Hollan and Stornetta's focus on creating technologies that meet needs that cannot so easily be met "in the medium of physical proximity." The realm of social capital is comprised of structural, cognitive, and relational dimensions. Ostensibly, ICTs can aid in structural social capital through the development of networks that connect people, but it seems they can aid in less obvious ways. From this work it is clear that ICTs can also play a role in cognitive and relational aspects of social capital. Moving forward, research in this area would benefit by considering the ways in which technologies can and cannot support development of social capital—distributed or co-located. In being sensitive to these possibilities and limitations, designers, individuals, teams, and organizations can seek more creative ways to build and nurture social bonds in transnational work.

Limitations and Future Work. Several limitations of this work should be noted. This study focused on Indian and Chinese participants in transnational teams; we did not include other nationalities nor the teammates with whom these participants work. Also, while we did visit the US headquarters, we did not visit international sites, and all interviews were conducted via video conferencing. Thus, our findings may be limited by interviewees not talking as openly as they might in person. By relying on interviews alone, we may also be assessing perceived, rather than actual, social capital. Including other teammates (non-transplant, non-remote) and conducting fieldwork that goes beyond a case study of single firm would certainly aid in further investigating these social capital phenomena. This could be a useful starting point for development of a survey instrument that can operationalize and validate some of these constructs, particularly in teasing apart *perceived* social capital and *actual* social capital. Future research should focus on creating a better framework for looking at behavioral dynamics in transnational teams, taking the context of location into account. This line of research should also look at whether these effects hold when the dominant culture is non-Western. On a broader level, such research could have broad implications for individuals, team structuring, and organizational mechanisms for co-adaptability.

Acknowledgements. We thank all the participants in this study for their time and contributions. We thank Google for generously supporting this research.

References

1. Ryan, A.M., Tippins, N.: Designing and Implementing Global Staffing Systems. Wiley-Blackwell (2009)
2. Bradner, E., Mark, G.: Why Distance Matters: Effects on Cooperation, Persuasion and Deception. In: Proc. of the ACM Conference on CSCW, pp. 226–235 (2002)
3. Grinter, R.E., Herbsleb, J.D., Perry, D.E.: The Geography of Coordination: Dealing with Distance in R&D Work. In: Proc. GROUP Conference of the ACM, pp. 306–315 (1999)
4. Koehne, B., Shih, P.C., Olson, J.S.: Remote and Alone: Coping With Being the Remote Member on the Team. In: Proc. of the ACM Conference on CSCW, pp. 1257–1266 (2012)
5. Olson, G.M., Olson, J.S.: Distance matters. Human-Computer Interaction 15(2), 139–178 (2000)
6. Tang, J.C., Zhao, C., Cao, X., Inkpen, K.: Your Time Zone or Mine? A study of Globally Time Zone-Shifted Collaboration. In: Proc. ACM Conference on CSCW, pp. 235–244 (2011)
7. Gelfand, M.J., Erez, M., Aycan, Z.: Cross-Cultural Organizational Behavior. Annual Review of Psychology 58, 479–514 (2007)
8. Hall, E.T.: Beyond Culture. Anchor Books (1977)
9. Hofstede, G.H.: Culture's consequences: International Differences in Work-Related Values. Sage (1980)
10. Gibson, C., McDaniel, D.: Moving beyond conventional wisdom; Advancements in Cross-Cultural Theories of Leadership, Conflict, and Teams. Perspectives on Psychological Science 5, 450–462 (2012)
11. Hinds, P., Liu, L., Lyon, J.: Putting the Global in Global Work: An Intercultural Lens on the Practice of Cross-National Collaboration. The Academy of Management Annals 5(1), 135–188 (2011)
12. Kraidy, M.M., Murphy, P.D.: Shifting Geertz: Toward a Theory of Translocalism in Global Communication Studies. Communication Theory 18, 335–355 (2008)
13. Zimmerman, A.: Interpersonal relationships in transnational, virtual teams—towards a configurational perspective. Int. J. of Management Reviews 12(1), 59–78 (2011)
14. Page, S.E.: The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton University Press (2007)
15. Stahl, G.K., Maznevski, M.L., Voigt, A., Jonsen, K.: Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. J. of International Business Studies 41, 690–709 (2010)
16. Gibson, C.B., Grubb, A.R.: Turning the tide in multinational teams. In: Shapiro, D.L., Von Glinow, M.A., Cheng, J.L. (eds.) Managing Multinational Teams: Global Perspectives, pp. 69–96. Elsevier, Oxford (2005)
17. Castells, M.: The Rise of the Network Society (1996)
18. O'Leary, M.B., Mortensen, M.: Go (Con)figure: Subgroups, Imbalance, and Isolates in Geographically Dispersed Teams. Organizational Science 21(1), 115–131 (2010)
19. Gorton, I., Motwani, S.: Issues in Co-operative Software Engineering Using Globally Distributed Teams. Information and Software Technology 38, 647–655 (1996)
20. Gupta, A., Mattarelli, E., Seshasai, S., Broschak, J.: Use of Collaborative Technologies and Knowledge Sharing in Co-located and Distributed Teams: Towards the 24-h Knowledge Factory. J. of Strategic Information Systems 18, 147–161 (2009)

21. Ramasubbu, N., Mithas, S., Kemerer, C.F.: Work Dispersion, Process-Based Learning, and Offshore Software Development Performance. MIS Quarterly 32, 437–458 (2008)
22. Olson, J., Olson, M.: Culture Surprises. Queue, 52–59 (December- January 2003)
23. Nahapiet, J., Ghoshal, S.: Social Capital, Intellectual Capital, and the Organizational Advantage. The Academy of Management Review 23(2), 242–266 (1998)
24. Irani, L., Vertesi, J., Dourish, P., Philip, K., Grinter, R.E.: Postcolonial computing: A lens on Design and Development. In: Proc. of the ACM Conference on Human Factors in Computing Systems, pp. 1311–1320 (2010)
25. Deming, W.E.: Sample Design in Business Research. John Wiley & Sons (1990)
26. Burrell, J., Anderson, K.: 'I have great desire to look beyond my world': trajectories of information and communication technology use among Ghanaians living abroad. New Media and Society 10, 203–224 (2008)
27. Olson, G.M., Olson, J.S., Venolia, G.: What Still matters about distance. In: Proc. of HCIC (2009)
28. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine de Gruyter (1967)
29. Narayan, S.: Return to India: One Family's Journey to America and Back. Special Report. Wharton, `http://knowledge.wharton.upenn.edu/india`
30. Morris, M.W., Podolny, J., Sullivan, B.N.: Culture and Coworker relations: Interpersonal Patterns in American, Chinese, German, and Spanish Divisions of a Global Retail Bank. Organization Science 19, 517–532 (2008)
31. Breton, R.: Institutional Completeness of Ethnic Communities and the Personal Relations of Immigrants. American Journal of Sociology 70(2), 193–205 (1964)
32. Goffman, E.: Forms of Talk. University of Pennsylvania Press (1981)
33. Weisfeld, G.E.: The Adaptive Value of Humor and Laughter. Ethnology & Sociobiology 14(2), 141–169 (1993)
34. Cai, D.A., Hung, C.J.: How relevant is trust anyway? A cross-cultural comparison of trust in organizational and peer relationships. In: Cheney, G., Barnett, G.A. (eds.) International and Multicultural Organizational Communication, Hapton, pp. 73–104 (2005)
35. Yuki, M., Maddux, W., Brewer, M., Takemura, K.: Cross-Cultural Differences in Relationship- & Group-Based Trust. Personality and Soc. Psych. Bulletin 31, 48–60 (2005)
36. Rocco, E., Finholt, T.A., Hofer, E.C., Herbsleb, J.D.: Designing as if trust mattered. CREW Technical Report, University of Michigan (2001)
37. Jarvenpaa, S.L., Leidner, D.E.: Communication and trust in global virtual teams. Organization Science 10(6), 791–815 (1999)
38. Meyerson, D., Weick, K.E., Kramer, R.M.: Swift Trust and Temporary Groups. In: Kramer, R.M., Tyler, T.R. (eds.) Trust in Organizations: Frontiers of Theory and Research, pp. 166–195. Sage, Thousand Oaks (1996)
39. Starbuck, W.H.: Learning by knowledge intensive firms. J. of Management Studies 30, 713–740 (1992)
40. Ulijn, J., Li, X.: Is Interrupting Impolite? Some Temporal Aspects of Turn Taking in Chinese-Western & Other Intercultural Business Encounters. Text 15(4), 589–627 (1995)
41. Geister, S., Konradt, U.: Effects of Process Feedback on Motivation, Satisfaction, and Performance in Virtual Teams. Small Group Research 37(5), 459–489 (2006)
42. Lubbers, M.J., Molina, J.L., Lerner, J., Brandes, U., Avila, J., McCarty, C.: Longitudinal Analysis of Personal Networks: The Case of Argentinean Migrants in Spain. Social Networks 32, 91–104 (2010)
43. Hollan, J., Stornetta, S.: Beyond Being There. In: Proc. of the ACM Conference on Human Factors in Computing Systems, pp. 119–125 (1992)

A Comparison of List vs. Hierarchical UIs on Mobile Phones for Non-literate Users

Indrani Medhi[1,*], Kentaro Toyama[2], Anirudha Joshi[1], Uday Athavankar[1], and Edward Cutrell[3]

[1] Industrial Design Centre, IIT Bombay, Powai, Mumbai 400076, India
[2] School of Information, University of California, Berkeley, USA
[3] Microsoft Research India, Bangalore, India
{indranim,cutrell}@microsoft.com, {anirudha,uaa}@iitb.ac.in, kentaro_toyama@hotmail.com

Abstract. Previous research has shown that low-literate users have difficulty using hierarchical information architectures and that a list design showing all items at once on a PC screen works best for search tasks. However, the limited screen space on phones makes it impossible to show more than a few items at once on a single screen. Does a hierarchical UI work better on a phone? In this study, we compared the performance of non-literate users from Bangalore, India, on a search task using a hierarchical UI (four levels deep) and a multi-page list that had forty items across seven pages of a touch-screen phone. Our results show that participants using the multi-page list perform better both in terms of time taken and percent correct even when the list UI design requires them to browse through multiple pages of items on the phone.

Keywords: Non-literate users, list design, hierarchy, mobile phone.

1 Introduction

Traditional computing software is structured in information architectures (IAs) designed in the form of hierarchies, to enable navigation of enormous information systems by concentrating on a few issues at a time. One of the principle benefits of hierarchical IAs is that space needed for navigation can be reduced by nesting. However previous research has shown that hierarchies can be difficult to use [1, 7, 9] particularly for low-literate people [3, 6]. Low-literate users have trouble understanding the concept of nesting or how a top node in a hierarchy represents a group of pages. Prior work showed that in the context of a search task of 40 items, a list UI that displayed items in a grid all at once on a PC screen worked best [6]. But displaying all selections at once is frequently not an option – for example, on a phone UI where screen real estate is constrained. What happens when it is not possible for items in a search task to be visible all at once on a list?

Do low-literate users have challenges with paging, or with hierarchies? Is their challenge cognitive organization and understanding ontologies and representations, or

[*] Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 497–504, 2013.

is it that paging itself is the challenge? And, is it ultimately easier for low-literate users to navigate multi-page lists or hierarchical IAs when screen space is limited?

To answer the above questions, we conducted a controlled experiment with 20 non-literate users in Bangalore India, on a search task on a touch-screen phone. Ten of the participants used a multiple-page list UI (where items are listed across seven pages), while the remaining ten used a hierarchical UI (4 levels deep). Both groups were asked to find five out of a total of forty household items. Our results show that participants using the multi-page phone list perform better both in terms of time taken and percent correct even when the list UI design requires users to traverse through multiple pages of items on the phone. As design recommendations in UIs for non-literate users, we suggest keeping navigation linear to the extent possible and to minimize hierarchical depth.

2 Related Work

A previous study showed that low-literate users performed poorly when using hierarchies in a search task [6]. This study compared a single-page list, a shallow hierarchy and a deep hierarchy on a PC. Participants performed best on the single-page list where all search items were visible all at once on the PC screen. Other studies have questioned the suitability of menu-based navigation for low-literate novice users in the developing world context [3, 4, 5]. However, these studies are more qualitative and anecdotal in nature. Studies with literate users have shown other problems related to the usability of hierarchical menus [1, 7, 8, 9]. Particularly relevant for this work, a study based on PDAs compared linear, hierarchical and cross-linked navigation performance in the context of a search task [2]. Results indicated that users performed best when navigating a linear structure, but preferred it mostly because of the ability to go to the ‘Home’ screen from any page since it allowed them to “start over”. However, paging between subsequent screens was actually found to be difficult. Note, that in contrast with the current work, this study [2] was conducted in a Western context with a higher literate group (with 10-14 years of formal education), half of who had previous experience playing computer games or browsing the internet occasionally.

3 Study Methodology

The methodology of our study below closely follows that of [6] for the sake of comparison. The reader is referred there for more detail.

3.1 UI Prototypes

The domain of choice for this study was commonly occurring household items. These items: a) allowed for graphical representation for the target non-literate population; b) were widely understood and did not require any domain specific knowledge; c) were gender neutral; and d) allowed for extensive categorization. Forty common household items were selected that included items of clothing, jewelry, utensils, electronics, games and sports, etc. We considered two instantiations for comparison: a list and a hierarchy of items. Each of these UI prototypes were displayed on a Samsung GT-I8350 running Windows Phone 7.5.

List. A list UI of 40 items, distributed continually over a total of 7 pages on the mobile phone; there were 6 items per page up to the 6th page (in a 3X2 matrix; 3 columns, 2 rows), and then 4 remaining items on the 7th page (in a 2X2 matrix). We took the layout of the PC list from the previous study [6] and divided up the items into groups of 6 to be placed on consecutive individual pages on the phone UI (the 7th page had the remaining 4 items). This resulted in loose categories of items per page sometimes flowing into subsequent pages. There were forward and backward arrows that had to be tapped to traverse between the pages of the list. To select any item, the test participant had to point to the item with his/her finger.

Fig. 1a. 1st page of the list UI on phone with a forward arrow

Fig. 1b. 5th page of the list UI on phone with a forward and a backward arrow

Hierarchy. A hierarchy UI of 40 items (4 levels deep and average branching factor of 3). The items in the hierarchy were organized in a top-down navigation tree based on four levels of organization: first level is how the item is used (e.g. things you wear, things you use), second is item category (e.g. Clothes, electronics, jewelry, etc.), third is item sub-category (Men's clothes, living room electronics, hands jewelry, etc.) and fourth is type of item (all 40 items such as shirt, TV set, bangles, etc.). Figure 2 shows the IA of the hierarchy. On the UI, each node of the organization was represented by a photograph that best depicted that node. To ensure that any cultural biases affecting the hierarchy were consistent with those of target participants, the previous study [6] conducted a validation of categories with members of the target community.

To select any given search item, test participants had to make four choices down the navigation tree to arrive at a given item. The interface was completely graphical with no text. Tapping on a certain graphic would take the user to the next level (sub-ordinate categories) of the hierarchy. There was the provision to go back to the previous page in the hierarchy by clicking on a "back" button at the bottom on any given page (Figure 3).

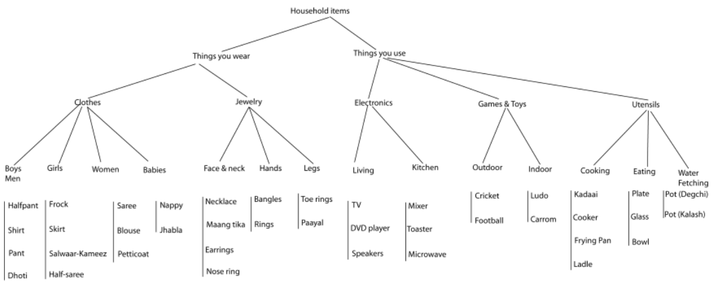


Fig.2. Hierarchy UI architecture on phone

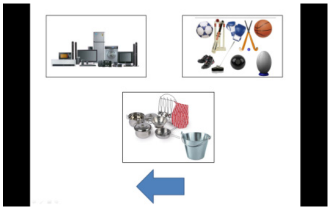


Fig.3. Screenshot of a page from the deep hierarchy UI

Process of Administering the UI Test. Every participant was randomly assigned one of the prototypes (list / hierarchy). Participants were asked to carry out five tasks: each task required them to find a given household item on the UI that was allotted, such as set of bangles, water pot, a football, a pair of shorts and a mixer-grinder. There was a time limit of 2 minutes for every task. Consecutive tasks were announced only after the previous task was over. There was a standard script of verbal instructions provided at the beginning of the test by the experimenter. We used the timing in the UI test solely as a mechanism for making progress with participants; the fact that participants were timed was not announced to them. This was done so participants did not feel that they were under time pressure – what they experienced was that occasionally, we would simply move onto another task.

Before using the UI prototype, for both the phone list UI and the phone hierarchy, like in the previous study [6] participants watched an instructional video on how to use the UIs on the touch-screen phone interface (of 1 min 58 secs for the list UI video and 2 min 8 sec for the deep hierarchy UI video). The hierarchy UI required more time to explain the concept of nesting. Both the instructional videos also had details about how to tap on the phone screen to go to the other pages by clicking on the arrows, or in the case of the deep hierarchy, on a given graphic (or back arrows). Participants could watch the video as many times as they wanted, up to a maximum of three times. The time taken to watch the video was not accounted for during performance evaluation on either of the UIs. The domain for the instructional video was animals-birds kingdom instead of household item, so there was no learning effect on the actual tests. There was no assistance provided by the experimenter during use of the UI.

3.2 Experimental Procedure

For consistency, the same researcher acted as experimenter for all participants and followed a standard script. Participants came in one by one. The researcher first gathered information about the participant such as their age, and asked if they had any formal education and technology experience, particularly with touch-screen phones. Then, each participant took a UI prototype test. The dependent variables that we measured are: number of correct tasks (maximum 5), total time taken for correct tasks, total time taken (maximum 10 minutes for incorrect and correct tasks combined). The search items that participants had to find on the multi-page list ranged between pages one to seven such that the average number of page selections was 4, so as to have an equivalent comparison with the hierarchy where the user had to make 4 choices down the navigation tree to arrive at any search item.

We conducted a between-subjects experiment design with a total of 20 participants. There were 2 kinds of UI prototypes-- list and hierarchy. First we paired the 20 participants by age, gender and mobile phone experience and then flipped a coin to decide who uses which UI prototype. This resulted in 10 participants each for the list and the hierarchy. The experimental design is illustrated in Table 1, together with mean and median ages and gender break-up.

Table 1. Between-subjects experimental design with 20 nos. of participants (m=male, f=female)

List	Hierarchy
10 nos. (5 m, 5 f)	10 nos. (5 m, 5 f)
Mean age: 38.8 yrs	Mean age: 38.2 yrs
Median age: 36.5 yrs	Median age: 38.5 yrs
4 nos. phone users	5 nos. phone users

3.3 Participants

Test participants were drawn from two urban slum communities in Bangalore, India. They were recruited through an intermediary organization doing developmental activities in the slum areas. Most people that were recruited were in informal sector jobs: construction workers, vegetable vendors, domestic workers, motor mechanics, etc. The household income of participants was less than about INR 6500 (USD 120) per month. The age range was 18-65 years. None of the participants had any formal education and could not read or write, though all of them were numerate and could recognize Indo-Arabic numerals. Their primary language of communication was Kannada. Apart from this, a few people also spoke Tamil and Hindi. None of the participants had any previous experience using touch-screen phones. Most male participants owned and used personal mobile phones for making voice calls only, dialing numbers from scratch every time. None of the participants with mobile phones used the phone book. In terms of other technology use, TVs, DVD players and CD players were common items in participant households.

4 Results

4.1 Quantitative Findings

Figure 4a illustrates the mean time taken for correct responses for each of the UIs. Results show that the multiple-page phone list UI required significantly less time to navigate than the deep hierarchy phone UI (average of 25 vs. 65.5 seconds) t(18)=4.6, p<0.001. This finding is corroborated when we take mean % correct tasks as the dependent measure of performance. Refer Figure 4b. Participants completed more correct tasks on the multiple-page phone list UI than on the deep hierarchy phone UI (average of 100 vs. 80), t(18)= -3, p=0.0077.

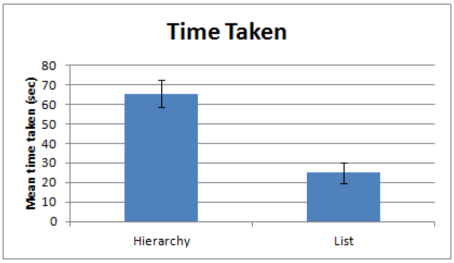


Fig. 4a. Mean time taken across UIs (±SEM)

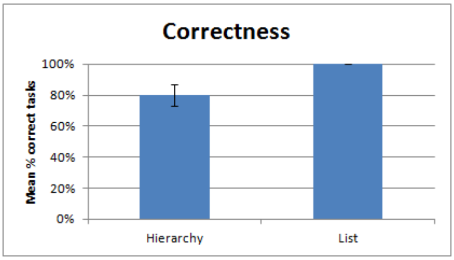


Fig. 4b. Mean % correct tasks across UIs (±SEM)

4.2 Qualitative Findings

We had a number of qualitative observations during the phone UI tests and follow-up informal conversations with the participants, some of which are consistent with the observations from the previous study [6] with respect to use of the hierarchy: First, some participants who could not complete tasks correctly or took more time on the hierarchical UI did not seem to understand the concept of nesting, or that the top graphic in a hierarchy represented a group of pages. The video instructions shown before the use of the hierarchy explained how (subordinate) items were "contained within" (super-ordinate) item categories, represented by a related graphic. It further explained how selecting that graphic would take the participant to the items contained within that category. But during the UI tasks, some participants seemed to guess randomly. They selected unrelated graphics on any given page in the hierarchy in efforts to find an assigned item.

Second, some people did not remember how to navigate back to higher levels once they had gone down the incorrect path in the hierarchy. Follow-up conversations revealed that they had forgotten the use of the "back" button from the instructional video. This could be because of problems with short-term memory, inattention during the video instructions, lack of understanding about how "back" corresponds to moving up a level in the hierarchy, or the challenge of the additional cognitive load that navigating a hierarchy imposes. However, our study does not allow us to distinguish between these possibilities.

In the case of the list UI, however, participants seemed to understand the “back” button better. When participants needed to go to the previous page, they pressed the button without much hesitation. We suspect the recall was helped by the placement of “back” right beside the “forward” button (refer Fig. 1b) that they already had to use in any given task (except in the final task where the item was on the first screen and neither the forward or back button had to be used). The difference between the two arrows was indicated by their directionality.

On the phone list UI, we observed that even though the items were spread across 7 pages and not all of the items were visible at once, participants did not hesitate to move about quickly through the pages. Once they realised that the task item was not available on the first screen, unlike in the hierarchy UI they seemed to remember instructions from the video; they would promptly press the arrow (“forward” in this case) to move to the next page to find the task item. Even though none of our participants had any previous experience using a touch-screen phone, they seemed comfortable using this device. We suspect this could be because of general familiarity with the form factor of a phone, even though the interaction was through the touch-screen.

5 Discussion

In general, we had expected the test participants on the multiple-page phone list to perform poorly since, unlike a single-page list, not all items were visible at once. Finding task items on the phone list required participants to traverse through multiple pages (total 7 pages). But we found that test participants, despite having no previous experience using touch-screen phones, seemed fairly comfortable moving about the pages using the forward and back buttons. After watching them use the phone list with such comfort, we suspect a multiple-page list design might work better than a single-page PC list UI as it can be overwhelming to scan all 40 items listed at once. Further work is needed to determine if a multi-page phone list is in fact better than a single-page PC list. If the multipage phone list performs better, it would be interesting to investigate the reasons: fewer items for scanning per page, general comfort with form factor of phone versus a PC, etc.

Since this work, unlike the previous study [6], shows that it is not the paging itself that causes the problems, there seems to be something inherent in hierarchical navigation that poses a challenge. Further work is needed to investigate why this might be. We suspect it is because list navigation really does not require a user to remember much – it is just moving back and forth. At most, the user has to remember where he/she is along a single line. Hierarchical navigation seems more complex – the user has to remember that he/she is in a hierarchy, and has to hold the hierarchy from the root in their thinking. It is possible that literate users also would perform relatively poorly on the hierarchical UI, given the circumstances of our trial. But our study does not allow us to deduce this given there was no control group of literate users; we recognize this is a potential limitation of our study.

6 Conclusions

Previous research has shown that in the context of a search task, a list UI design that displays all items at once on a PC screen works better than a UI where items are categorized under a top-down navigation tree of a hierarchy. But displaying all search items at once on a screen is frequently not an option for devices such as mobile phones where screen space is limited. We investigated the trade-off of paging through multiple pages of a list UI on a touch-screen phone, compared to a hierarchy where all of the items at a given level are all visible at once. 10 non-literate users used the multi-page list while another 10 non-literate users used the hierarchical UI. They were asked to search for five out of forty household items. Our results showed that both in terms of time taken and percent correct, non-literate users using the multi-page list UI perform better even when the list design required them to browse through multiple pages of the phone. Based on this, we suggest keeping navigation linear to the extent possible and to minimize hierarchical depth in UIs for non-literate users.

References

1. Allen, R.B.: Cognitive factors in the use of menus and trees: an experiment. IEEE Journal on Selected Areas in Communications SAC-1(2), 333–336 (1983)
2. Chaudry, B., Connelly, K., Siek, K.A., Welch, J.L.: Mobile Interface Design for Low-Literacy Populations. In: Proceedings of the 2012 ACM SIGHIT International Health Informatics Symposium (2012)
3. Jones, M., Marsden, G.: Mobile Interaction Design. John Wiley & Sons (2005)
4. Jones, M., Buchanan, G., Thimbleby, H., Marsden, G.: User interfaces for mobile web devices (www9 mobile workshop position paper). In: Proceedings of the 9th International World Wide Web Conference, Amsterdam (2000)
5. Katre, D.S.: Unorganized cognitive structures of non-literate as the key factor in rural e-learning design. I-Manager's Journal of Education Technology 2(4), 67–72 (2006)
6. Medhi, I., Lakshmanan, M., Toyama, K., Cutrell, E.: Some Evidence for the Impact of Limited Education on Hierarchical User Interface Navigation. In: ACM Conference on Human Factors in Computing Systems (2013)
7. Parush, A., Yuviler-Gavish, N.: Web Navigation Structures in Cellular Phones: the Depth/Breadth Trade-off Issue. International Journal of Human-Computer Studies 60, 753–770 (2004)
8. Tang, K.E.: Menu design with visual momentum for compact smart products. Human Factors 43(2), 267–277 (2001)
9. Tullis, T.S.: Designing a menu-based interface to an operating system. In: Proceedings of CHI 1985, pp. 79–84 (1985)

A User Study with GUIs Tailored for Smartphones

David Raneburger[1], David Alonso-Ríos[2], Roman Popp[1],
Hermann Kaindl[1], and Jürgen Falb[1]

[1] Institute of Computer Technology, Vienna University of Technology, Vienna, Austria
{raneburger,popp,kaindl,falb}@ict.tuwien.ac.at
[2] Department of Computer Science, Faculty of Informatics, University of A Coruña, Spain
dalonso@udc.es

Abstract. Web-based graphical user interfaces (GUIs) are mostly not tailored for small devices with touchscreens, such as smartphones. There is little scientific evidence on the conditions where additional taps for navigation are better or scrolling. Therefore, we conducted a user study in which we evaluated different ways of tailoring a GUI for a smartphone. Each participant performed the same task with two different layouts of the same GUI. We collected quantitative data through measuring task completion time and error rates, as well as qualitative data through subjective questionnaires. The main result is that minimizing the number of taps is important on a smartphone. Users performed significantly better when they could scroll (vertically), instead of tapping on widget elements (tabs). This preference was also reflected in their subjective opinions.

Keywords: Usability, device-tailored GUI, small touchscreen, user study.

1 Introduction

Schmiedl et al. showed in an experiment that users completed tasks on small mobile devices 30–40% faster on Websites tailored for such devices than on Websites designed for desktop PCs [1]. Smartphones and PCs have *two* major differences in this respect: size and input method (i.e., touch vs. fine-grained mouse pointing). What is then the best approach for tailoring Web-based GUIs for touch-based smartphones? Web-based usability guidelines typically agree that unnecessary scrolling [2, 3], and in particular horizontal scrolling [4], should be avoided on desktop devices. Recently published guidelines for touch-based devices, in contrast, state that scrolling is part of such a device's user experience and, therefore, not bad at all [5, 6]. The different user experience originates in the more "natural" swiping gesture, in comparison to dragging a scroll-bar with a mouse, or turning its scroll-wheel. These guidelines, however, do not contain hints on whether vertical scrolling is preferable to tabs.

Since we are not aware of any empirical studies that investigate these questions, we designed and performed a user study evaluating an intentionally simple Web-based GUI using different layout approaches for a process-oriented application (flight booking). Our focus is on vertical scrolling vs. tapping, while two-dimensional scrolling and horizontal browsing is out of scope. Based on the collected data, we present both a statistical and a subjective evaluation of the results.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 505–512, 2013.

2 Related Work

Different GUI designs and input methods for touch-based mobile phones are investigated in [7], focusing on overall interface layout (scrollable vs. tabbed GUI), input mode (keyboard vs. tapping through a modal dialog) and menu access (device menu vs. context menu). The two GUIs under study in [7] differed in all these characteristics, which were investigated through an assigned task. Nevertheless, it is hard to prove that only the investigated characteristic accounts for slower performance and that there is no influence from other characteristics. Our study, in contrast, is based on GUIs that differ only in one characteristic – the layout. In addition, our GUIs are HTML pages rendered on Web browsers, which display them similarly on different devices. This helps to minimize the influence that familiarity with specific devices may have on users' perceptions of usability and correlated time-on-task [8].

Recently, Lasch et al. investigated touchscreen scrolling techniques and their effect on car driver distraction [9]. Their results indicate that *swiping* was less distracting while driving than *traditional scroll buttons* or *kinetic scrolling*. The latter uses the same gesture as swiping, but it accelerates the menu. The menu stops automatically after deceleration or if the user taps on the screen. So, this work investigates three different ways of scrolling in a context where distraction is a safety risk. Our primary concern, in contrast, is to achieve good usability in a context where the user concentrates on a specific task. Therefore, we investigated whether scrolling was preferable to tabbed panels, or vice versa. Nevertheless, the results of both studies complement each other.

Comparable user studies have also been performed in the past, but only with desktop devices [10–12], old cell-phones [13] or PDAs [14], so their results are not necessarily valid for current smartphones. Previous user studies on touch-based devices (e.g., PDA or smartphone) assumed that the user does not like to scroll and, therefore, primarily focused on investigating strategies to avoid or minimize scrolling [15–17]. Jones et al. [18] found that users preferred vertical over two-dimensional scrolling on small screens. Our results extend these findings for current touch-based devices through the comparison of two different layouts tailored for smartphones.

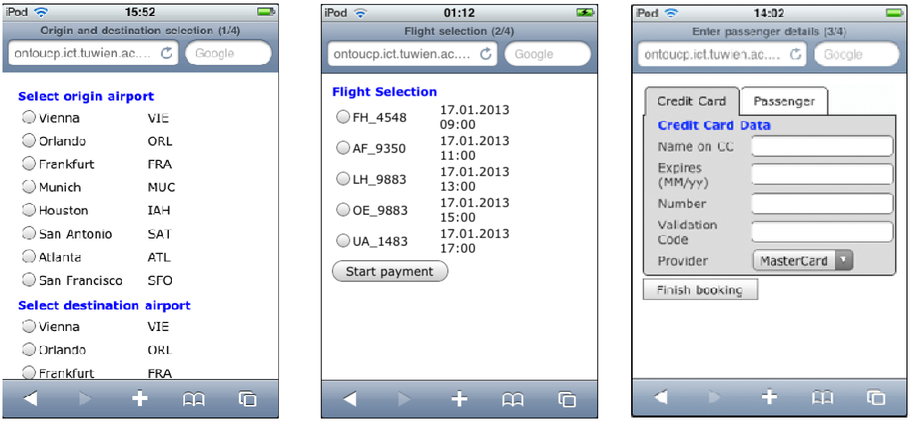
3 User Study Design

In order to evaluate the different GUIs, we combined quantitative data (task completion time and error rate) with subjective data (user questionnaires) like Jones et al. [15] and Buchanan et al. [13].

In our **study setup**, we tested two different layouts for an otherwise identical user interface. Both layouts were tailored for a small device each (a 320×480px iPod Touch / iPhone up to version 4S, which all show the GUI in the same way), more precisely in portrait mode. However, they followed different approaches (tabs vs. vertical scrolling). The application was a simplified flight-booking scenario consisting of HTML pages as follows:

- **Screen 1.** Origin and destination airports, flight date.
- **Screen 2.** List of flights available on that date.
- **Screen 3.** Billing information.

Fig. 1 shows screenshots of these screens, where Screens 1 and 3 are the ones tailored, while Screen 2 is identical for both layouts, as it is small enough to fit perfectly. Due to space limitations, Fig. 1 shows Screen 1 only in its form for vertical scrolling, and Screen 3 only in its tabbed form. During the test each layout was, of course, the same from the beginning to the end of the task – there were no changes of layout in the middle of a task.



(a) Vertical scroll-based layout (Screen 1)

(b) Both layouts (Screen 2)

(c) Tab-based layout (Screen 3)

Fig. 1. Screens tailored for iPod Touch / iPhone up to version 4S

We split the **participants** into two groups, reversing the order in which the two layouts were tested. At the end of each test, we collected their subjective opinions through a questionnaire and, sometimes, a brief and informal interview. Overall, we hired 30 participants as summarized in Table 1.

Table 1. Participant Details

Age:	18–39 years
Gender:	3 female and 27 male
Educational level:	28 students and 2 assistants with doctoral degrees
Use of mobile devices:	56.7% regular, 40% sometimes but never owned one, 3.3% never
Online flight booking:	20% regular, 40% sometimes and 40% never

At the beginning of each experiment we informed the participants about the content of the study and the **procedure**. We emphasized that the point was to test the

GUIs, not the skill of the participants. We asked them to fill in some background data and to give their consent to filming their hands operating the iPod Touch (for subsequent video annotation) and recording their voices. We fixed the iPod Touch to the table, setting it in portrait mode for the following reasons: the GUIs have been tailored for this mode; we wanted all the participants to use the same mode; filming would have been difficult if the participants were holding the iPod Touch.

We gave the following scenario description to the participants:

"Imagine it is Tuesday 14/02/2012, 11:55am and your boss Mr. Huber tells you to book a flight ticket for his wife as quickly as possible. Mrs. Huber is already waiting at the airport!

Book a flight from *Munich* to *Atlanta* on *02/14/2012* at *1pm* for Mrs. *Anna Huber* (age 47). Pay for it using her husband's (Max Huber's) *VISA Credit Card* with the number: *1258 8569 7532 1569* (validation code: *354*) and the expiration date *12/14*."

We left a printed version on the table, where it could be easily seen, so that participants did not have to remember the information. We emphasized the urgency of the task, so that the participants would concentrate on the task and finish it quickly and realistically.

After the participants completed both tasks, we collected their subjective opinions on several usability-related issues through a questionnaire and finally asked them if they had further comments, in which case we made a brief interview.

For our **analysis**, we were primarily interested in the correlations between task time / error rate and the type of layout. An independent variable was the GUI, with these possible values: tab-based (T-UI) and vertical scroll-based (V-UI). This variable is dichotomous, as we only compared two GUIs per run. A second independent variable was the order in which the two layouts were tested. The dependent variables were adjusted task completion time and error rate. For the *adjusted task completion* analysis, we measured the total time spent on a screen. We also measured the time needed for text input, screen loading, form submission, validation alerts and error messages, which are all layout-independent and may introduce bias into the results. Therefore, we subtracted all these times from the total time, resulting in *adjusted task time*.

For the *error rate* analysis we measured the number of errors each participant made on a given screen. Errors in our study occurred when the participant hit "submit" without entering some required information.

As the type of GUI is dichotomous and the calculated time is on an interval scale, we calculated the *point-biserial Pearson correlation coefficients*. Our tests also showed that the variables are nearly normally distributed and thus satisfy the prerequisites of the point-biserial Pearson correlation. Note that the difference to the more common *t-test* is that the correlation covers coherence, while the t-test deals with differences of means. The corresponding **null hypotheses** are correlation hypotheses as follows:

- $NH_{VT,T}$: *There is no statistically significant correlation (p-value = 0.05) between the adjusted task time and the type of GUI: V-UI and T-UI.*
- $NH_{VT,E}$: *There is no statistically significant correlation (p-value = 0.05) between the error rate and the type of GUI: V-UI and T-UI.*

In order to extract task completion times and error rates from our video recordings of the tests, we used the video-annotation tool Anvil[1], and converted the files to CSV in order to analyze the data with SPSS.

The participants were also given a **subjective usability questionnaire** for collecting information on how they perceived usability. With some adaptations, our definition of usability was based on [19], which aims to synthesize the best-known usability definitions in the literature. We discarded those usability characteristics that were either not relevant or identical on all layouts (e.g., security) and tried to phrase the usability criteria as short and self-explanatory questions, for which we also consulted the USE Questionnaire[2], the W3C's WAI[3], the Software Usability Measurement Inventory[4] (SUMI) and the Cognitive Dimensions framework [20].

For each question, the participant had to state the preferred layout on a Likert scale with preference being “extreme”, “strong”, “moderate”, and “equal”. The participants were explicitly asked to focus on the differences between layouts, rather than evaluate the usability of each layout independently. The reason is that for our study, which compares different layouts, their difference is relevant. Measuring usability on an absolute scale is more difficult and not necessary for our study.

4 Results

Table 2 summarizes the statistical results for the **adjusted task time analysis**. The positive correlations in Table 2 show that vertical scrolling performs better than the tab-based layout, with highly significant results. The sum of times for all screens (minus typing time, etc.; see above) is on average 25.81 seconds for the vertical scroll-based GUI (V-UI) and 36.79 seconds for the tab-based GUI (T-UI).

Table 2. Correlation between adjusted task time on a screen and its layout

	Pearson Corr.	Sig. (1-tailed)	*V-UI* av. time	*T-UI* av. time
Time Screen1 × GUI	**0.35**	0.003	12.79s	17.98s
Time Screen2 × GUI	0.12	0.189	8.17s	9.50s
Time Screen3 × GUI	**0.43**	0.000	4.85s	9.31s

The null hypothesis $NH_{VT,T}$ can be rejected for this experiment, i.e., the adjusted task time using V-UI is significantly smaller than using T-UI in Screen 1 and Screen 3. In fact, it took 54% longer to operate T-UI.

Table 3 shows the **correlation between the error rate and the two layouts** (V-UI and T-UI), together with the average number of errors for each screen. For Screens 1 and 3, the error rate is significantly smaller. On Screen 2, however, there were no errors, thus no correlation could be calculated. So, the null hypothesis $NH_{VT,E}$ can also be rejected for this experiment. That is, the error rate for V-UI is significantly smaller

[1] http://www.anvil-software.org/

[2] http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html

[3] http://www.w3.org/WAI/EO/Drafts/UCD/questions.html

[4] http://sumi.ucc.ie/en/

than that of T-UI in Screens 1 and 3. Actually, there were no errors for V-UI – all the errors can be attributed to the participants hitting “submit” before changing tabs, so this could be an argument against using tabs.

Table 3. Correlation between the number of errors on a screen and its layout

	Pearson Corr.	Sig. (1-tailed)	*V-UI* av. errors	*T-UI* av. errors
Errors Screen1 × GUI	**0.363**	0.002	0	0.23
Errors Screen2 × GUI	-	-	0	0
Errors Screen3 × GUI	**0.285**	0.014	0	0.2

Table 4 presents the **results of the subjective questionnaire**. In their overall assessment (i.e., the last questionnaire item), the majority (60%) of the participants preferred the vertical scroll-based GUI, but 30% preferred the tab-based GUI. The answers to the other questions varied much. There was only one criterion for which T-UI was widely preferred: visual attractiveness.

The error rate result was confirmed by participants in interviews, where they stated that it is easy to forget to switch tabs before hitting “submit” even when they were aware that there were two tabs. As for V-UI, the questionnaire shows that it was considered much more intuitive to navigate and to interact with, and slightly less demanding and more efficient to use in general. However, opinions were sharply divided on which GUI made information more visible and which GUI lent itself more to how the participants like to work. It is worth noting that, whenever opinion was divided, analyzing the data showed no clear relationship between the answers and the background of the participants, including whether they owned such a device or not.

5 Conclusion and Future Work

This paper reports on a user study with GUIs specifically tailored for smartphones. We used task time analysis, error rate measurement, subjective questionnaires and interviews. Participants performed significantly better when scrolling vertically, as opposed to tapping on widget elements (tabs). This preference was also reflected in their subjective opinions. The study also suggests that minimizing the number of taps is important on a smartphone.

We conjecture that our study results are not limited to such a simplified flight-booking application as the one used for two reasons. First, we defined a scenario that reduced the user task to entering pre-given data, without having to decide about actually booking a particular flight or not. Second, we used a very simple prototypical application that allowed only for exactly the interaction that was required to complete the given task, without alternatives. This also helped to keep the user focus on the interaction and allowed all users to complete the task successfully.

In future work, we plan to investigate related issues, and e.g., to study variable-sized lists, since it is not clear if vertical lists of *any* length are necessarily better than a tab-based layout. There may be a balance involved between motor performance and understanding the content.

Table 4. Subjective Questionnaire Results

Question	*V-UI* preferred			equal	*T-UI* preferred		
	extreme	strong	moderate		moderate	strong	extreme
Which interface makes information more visible?	3.33%	33.33%	10%	6.67%	30%	13.33%	3.33%
Which interface makes interaction more intuitive?	13.33%	26.67%	26.67%	30%	3.33%	0%	0%
Which interface makes it easier to figure out what to do next?	20%	36.67%	20%	10%	10%	3.33%	0%
Which interface makes it clearer how to use it?	0%	13.33%	30%	53.33%	3.3%	0%	0%
Which interface lends itself more to how you like to work?	13.33%	23.33%	13.33%	23.33%	10%	13.33%	3.33%
Which interface requires less manual interaction?	6.67%	20%	26.67%	26.67%	13.33%	6.67%	0%
Which interface demands less precision on your part?	3.33%	13.33%	16.67%	56.67%	10%	0%	0%
Which interface demands less time from you?	13.33%	10%	26.67%	33.33%	10%	6.67%	0%
Which interface makes interaction more efficient?	10%	6.67%	26.67%	33.33%	16.67%	6.67%	0%
Which interface is more visually attractive?	0%	0%	6.67%	10%	40%	23.33%	20%
Overall, which interface would you use to book a flight?	16.67%	20%	23.33%	10%	10%	13.33%	6.67%

Acknowledgments. David Alonso-Ríos's research was funded by the University of A Coruña (Spain) through the "Axudas á Investigación" programs of 2011 and 2012, and also in part by the Xunta de Galicia through projects CN2011/007 and CN2012/211, both projects also partially supported by the European Union ERDF.

References

1. Schmiedl, G., Seidl, M., Temper, K.: Mobile phone web browsing: a study on usage and usability of the mobile web. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2009, pp. 1–70. ACM, New York (2009)
2. Microsoft Corporation: User Experience and Interaction Guidelines for Windows 7 and Windows Vista (September 2010)

3. Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers Inc., San Francisco (1993)
4. Bevan, N.: Guidelines and standards for web usability. In: Proceedings of HCI International 2005. Lawrence Erlbaum (2005)
5. Apple Inc.: iOS Human Interface Guidelines (August 2012)
6. Microsoft Corporation: Windows 8 User Experience Guidelines (August 2012)
7. Balagtas-Fernandez, F., Forrai, J., Hussmann, H.: Evaluation of user interface design and input methods for applications on mobile touch screen devices. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 243–246. Springer, Heidelberg (2009)
8. Suzuki, S., Bellotti, V., Yee, N., John, B.E., Nakao, Y., Asahi, T., Fukuzumi, S.: Variation in importance of time-on-task with familiarity with mobile phone models. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 2551–2554. ACM, New York (2011)
9. Lasch, A., Kujala, T.: Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2012, pp. 41–48. ACM, New York (2012)
10. Harms, I., Schweibenz, W.: Usability engineering methods for the web: results from a usability study. In: Knorz, G. (ed.) Proceedings des 7. Internationalen Symposiums für Informationswissenschaft (ISI 2000), Saarländische Universitäts und Landesbibliothek, Univ.-Verl. Konstanz, pp. 17–30 (2000)
11. Baker, J.R.: The impact of paging vs. scrolling on reading online text passages (February 2003)
12. Peytchev, A., Couper, M.P., Mccabe, S.E., Crawford, S.D.: Web Survey Design: Paging versus Scrolling. Public Opinion Quarterly 70(4), 596–607 (2006)
13. Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., Pazzani, M.: Improving mobile internet usability. In: Proceedings of the 10th International Conference on World Wide Web, WWW 2001, pp. 673–680. ACM, New York (2001)
14. Kim, L., Albers, M.J.: Web design issues when searching for information in a small screen display. In: Proceedings of the 19th Annual International Conference on Computer Documentation, SIGDOC 2001, pp. 193–200. ACM, New York (2001)
15. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G.: Improving web interaction on small displays. Comput. Netw. 31(11-16), 1129–1137 (1999)
16. Kaikkonen, A., Roto, V.: Navigating in a mobile XHTML application. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 329–336. ACM, New York (2003)
17. Churchill, D., Hedberg, J.: Learning object design considerations for small-screen handheld devices. Computers & Education 50(3), 881–893 (2008)
18. Jones, S., Jones, M., Marsden, G., Patel, D., Cockburn, A.: An evaluation of integrated zooming and scrolling on small screens. Int. J. Human-Computer Studies 63(3), 271–303 (2005)
19. Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., Moret-Bonillo, V.: Usability: A critical analysis and a taxonomy. Int. J. Human-Computer Interaction 26(1), 53–74 (2010)
20. Blackwell, A.F., et al.: Cognitive dimensions of notations: Design tools for cognitive technology. In: Beynon, M., Nehaniv, C.L., Dautenhahn, K. (eds.) CT 2001. LNCS (LNAI), vol. 2117, pp. 325–341. Springer, Heidelberg (2001)

Evaluating Direct Manipulation Operations for Constraint-Based Layout

Clemens Zeidler[1], Christof Lutteroth[1], Wolfgang Stuerzlinger[2], and Gerald Weber[1]

[1] University of Auckland, 38 Princes Street, Auckland 1010, New Zealand
czei002@aucklanduni.ac.nz, {lutteroth,g.weber}@cs.auckland.ac.nz
[2] York University, 4700 Keele St., Toronto, Canada M3J 1P3
wolfgang@cse.yorku.ca

Abstract. Layout managers are used to control the placement of widgets in graphical user interfaces (GUIs). Constraint-based layout managers are more powerful than other ones. However, they are also more complex and their layouts are prone to problems that usually require direct editing of constraints. Today, designers commonly use GUI builders to specify GUIs. The complexities of traditional approaches to constraint-based layouts pose challenges for GUI builders.

We evaluate a novel GUI builder, the Auckland Layout Editor (ALE), which addresses these challenges by enabling GUI designers to specify constraint-based layouts via direct manipulation using simple, mouse-based operations. These operations hide the complexity of the constraint-based layout model, while giving designers access to its benefits.

In a user evaluation we compared ALE with two other mainstream layout builders, a grid-based and a constraint-based one. The time taken to create realistic sample layouts with our builder was significantly shorter, and most participants preferred ALE's approach. The evaluation demonstrates that good usability for authoring constraint-based layouts is possible.

Keywords: GUI builder, layout editing, layout manager, constraint-based layout, layout preview, evaluation.

1 Introduction

Graphical user interfaces (GUI) are widespread, even on web and mobile platforms. WYSIWYG GUI builders facilitate the creation of such GUIs by developers and designers. In GUI builders, the user can drag & drop widgets from a palette onto a GUI canvas and adjust their properties.

Today's applications are used across a wide range of screen sizes, and users often use multiple applications concurrently on the same screen. Thus users expect that GUIs can be resized and that widgets use the space allocated to an application judiciously. Early drag & drop GUI builders enabled only the creation of static layouts that are unable to adapt to window resizing or other changes to the distribution of screen space. In order to create resizable GUIs with dynamic layouts, a *layout manager* has to be used, which implements a *layout model.* The latter defines how objects in a layout, the *widgets*, can be arranged and how their resize behavior can be specified.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 513–529, 2013.

Constraint-based layout models are naturally powerful: with the notable exception of flow layouts, many other layout models including gridbag layout [15] can be reduced to constraint-based layouts, see also Figure 1. Many other layout models rely on a hierarchy of nested layouts to define more complex layouts. Within nested layouts, widgets typically cannot be aligned across different levels of the hierarchy. In contrast, constraint-based layouts can align widgets that are situated in completely different parts of a layout [10], which greatly reduces the need for nested layouts.

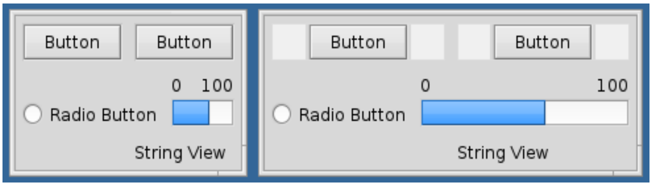


Fig. 1. Example of a constraint-based layout. When resizing, the middle row moves independently of the top and bottom rows. This layout is impossible to achieve with a single gridbag layout, e.g. as the radio button is not constrained to a column.

While constraint-based layouts are more powerful, their creation may be more complex and poses challenges. General constraints are difficult to visualize and even harder to manipulate directly. Specifying individual constraints can be tedious and error-prone, as they are situated at a low level of abstraction. Moreover, specifications may become over-constrained, i.e. have no solution, or permit widgets to overlap each other. Presumably it is for such reasons that support for constraint-based layouts in GUI builders is less developed than for other layout models. This makes it harder for GUI designers to leverage the advantages of constraint-based layouts.

In this paper we evaluate the use of simple drag & drop operations for the manipulation of constraint-based layouts. These operations are sufficient to create a wide variety of constraint-based layouts, including layouts where widgets are aligned to other widgets and no additional constraints exist between them. Our operations are designed to prevent situations where widgets can overlap. To test these operations with regard to their usability, we created a novel GUI builder – the Auckland Layout Editor (ALE) – that implements these operations. In ALE there is no need for the designer to know about the underlying constraint-based layout model.

In two controlled experiments we compared ALE with two state-of-the-art GUI builders. The experiments target the question if ALEs operations are effective for editing of constraint-based layouts. The first experiment compares ALE against a GUI builder that uses the gridbag layout model, which is currently one of the most frequently used layout models. The second experiment compares ALE against a GUI builder that supports constraint-based layouts. The evaluations show the promise of our new approach, with significantly shorter completion times for typical layout creation and editing tasks, and a strong preference from most of the participants for ALE.

Section 2 reviews related work. Section 3 presents ALE and Section 4 the edit operations used in ALE. The evaluation that compares ALE to a gridbag-based and a constraint-based GUI builder is described in Section 5. Section 6 finishes with conclusions.

2 Related Work

There are various non-constraint-based graphical GUI builders. Myers [11] summarized various techniques to create GUIs, including graphical tools to place interface objects on screens. IBuild [14] enabled the creation of complex GUIs in a WYSIWYG manner. It already supported nested layouts and spring-like "glue" layout elements. Moreover, IBuild enabled interactive testing of the resize behavior. Druid [13] predicted the intended alignment and spacing of a widget during editing, facilitating widget placement. FormsVBT [1] supported simultaneous editing of a textual and a graphical representation of a layout.

Some graphical GUI builders permit manual constraint editing. Lapidary [16] supported editing of constraint-based layouts with rich editing functions for constraints. In the constraint-based Gilt system [6], widgets could be aligned relative to user-specified tabstops or other widgets. In contrast to ALE, users had to manage tabstops themselves. In the Intui GUI builder [12] layout constraints can be edited directly by toggling between struts (fixed-size constraints) and springs (variable-size constraints) defined between and inside of widgets. However, only one constraint per widget side is supported. Finally, Rockit [9] automatically infers constraints from static drawings, using hard-coded rules and heuristics. These approaches require the user to edit or handle individual constraints. In contrast, ALE makes it possible to create constraint-based layouts without knowledge of the (somewhat complex) underlying constraint system.

Today, there are many open-source GUI builders, such as WindowBuilder Pro[1], Matisse/Swing[2], and Qt Designer[3]. There are also many commercial GUI builders, such as MS Expression Blend[4], Visual Studio[5], and the Apple Xcode Interface Builder[6]. Most of them support aligning widgets via snapping. With the exception of the Xcode Interface Builder, where Apple added support for constraints in 2012, none of them support the interactive construction or maintenance of constraint-based layouts. In Xcode widgets can be placed "freely" onto the editing canvas, and constraints can be added between the widgets. However, common GUI design guidelines such as [5], which guide designers towards appropriate layouts, generally emphasize that it is not desirable to place widgets completely "freely", i.e. at an arbitrary position in a layout. Aligned layouts are more compact and easier to understand [7]. ALE automatically keeps widgets aligned, which can lead to a faster layout creation and editing process.

Supple is an automated system that can adapt layouts to changes in display size, in particular to different devices. The system supports discrete changes of widgets, i.e. it changes the controls that are used within an input form depending on the available space [4]. However, the designer has less influence on the actual layout.

A previous usability comparison between constraint-based and grid-based layout models [19] used printed screenshots and sketching. This comparison found that the constraint-based layout model is a competitive alternative to the grid-bag layout model

[1] eclipse.org/windowbuilder
[2] netbeans.org/features/java/swing.html
[3] qt.nokia.com
[4] microsoft.com/expression/products/Blend_Overview.aspx
[5] microsoft.com/visualstudio/en-us
[6] developer.apple.com/technologies/tools

and performs better for editing existing layouts. The current paper compares constraint-based and grid-bag layout model editing directly in interactive GUI builders.

3 The Auckland Layout Editor (ALE)

Similar to other GUI builders, ALE's user interface consists of a component palette and an editing canvas (Figure 2). New widgets can be dragged from the palette into the editing canvas, where the designer can change the layout using a rich set of edit operations, as described in the next section. Here we first define the terms used to describe ALE.

In the underlying constraint-based layout model, each widget has a *minimum*, a *preferred* and a *maximum* size, also called *intrinsic* sizes. A widget should assume its preferred size if there are no other constraints for it, similar to the behavior of a sponge. By default, widgets are surrounded by a small margin, as visible in Figure 1.

There are two common types of linear constraints: *hard* constraints, which have to be satisfied, and *soft* constraints, which may be violated if necessary. When inserting a widget into a layout, some constraints are automatically derived from the intrinsic sizes of a widget: a hard inequality is created for the minimum size, and soft equalities for the preferred and the maximum size. These constraints are added in both the horizontal and vertical direction. As ALE inserts these constraints automatically, users do not need to manage the constraints for intrinsic sizes manually. The preferred size of each widget can be fine-tuned, which is comparable to changing the "weight" of a row or column in a grid-bag layout.

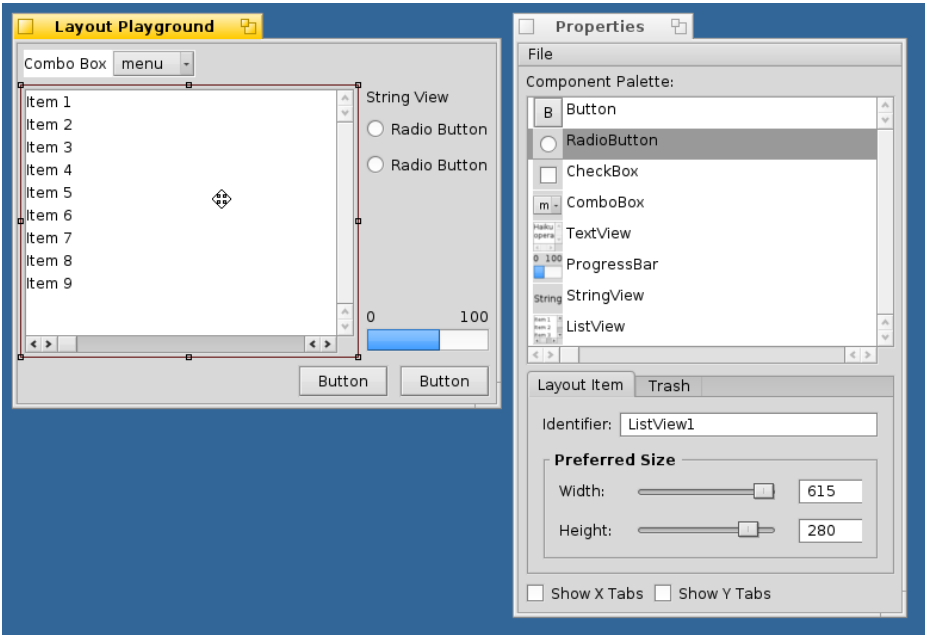


Fig. 2. Screenshot of the ALE GUI builder. New widgets can be inserted into the editing canvas (left) via drag & drop from the component palette (right).

Variables in a constraint are called *tabstops* and represent horizontal or vertical grid lines. Other frequently used names for the same concept are aligners, guides, snap lines or anchor lines. The edges of each widget are always connected to two horizontal (top and bottom) and two vertical (left and right) tabstops, which form a rectangular *area*. Connecting widgets to the same tabstop aligns them. Figure 3 shows a simple layout and its tabstops. Each GUI container, such as a panel or a window, defines tabstops for its four borders.

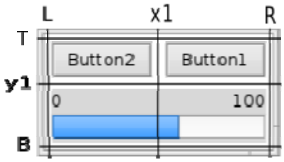


Fig. 3. Widgets are aligned to tabstops. In the picture there are four outer tabstops (L, T, R and B) and two inner tabstops (x1 and y1).

If all widgets are connected to at least one horizontal and one vertical tabstop, the whole layout can be computed directly. The reason for this is that the intrinsic sizes determine the size and the tabstops the x and y position of each widget. The chosen constraint solver controls how soft constraints are handled, which determines the final visual appearance of the layout. In our GUI builder and in the final applications, the quadratic active set method is used for solving [3]. In contrast to a linear approach, this leads to unique and aesthetically more pleasant layouts [17]. An overview of different solving techniques can be found in other works [2, 8].

The minimum size of widgets is guaranteed by a sufficiently large minimum size of the outermost container, e.g. the window. With this, a layout always stays solvable as only a soft inequality constraint is used for the maximum size. If a maximum size constraint is violated, then more space is allocated than a given widget is able to use. In this case, the widget is by default aligned in the center of its allocated space. The right side of Figure 1 illustrates this. The two buttons at the top share the full window width, and the maximum width constraints of the buttons are violated here. Thus, the buttons are centered in their allocated space, as visualized by the light gray areas.

4 ALE Layout Edit Operations

Ideally, a GUI builder should make the creation of a layout easy to learn, simple to achieve, and quick to perform. ALE's edit operations for constraint-based layouts automatically keep widgets aligned to each other as far as possible, by connecting them to existing tabstops. This makes it easier to rearrange widgets while keeping the layout consistent. Furthermore, the edit operations leave a widget connected to at least one horizontal and one vertical tabstop. This ensures that the position of a widget is always well-defined. The edit operations provided are moving, swapping, resizing, inserting, and removing of widgets. These operations enable the user to create complex layouts without manual constraint authoring.

As in most GUI builders, a GUI can be created and edited by drag & drop operations. Edit operations are started by dragging an already placed widget or the border of a widget. Then, ALE checks at each mouse position if an operation can be applied and gives corresponding visual feedback (Figure 4). For this, the operation is applied tentatively in the background. If an operation is appropriate, the user can commit the operation by "dropping" the dragged item and the result becomes visible. In general, edit operations can affect multiple widgets and a layout may differ in several places after an operation. To help the user understand such changes, a short animation visualizes how the affected widgets are changed.

Moving a Widget. While dragging a widget, its shape is visualized as a dotted rectangle. We limit the size of the dragged outline to a reasonable size to avoid problems with very large widgets. Two cases need to be considered when moving a widget from one position to another.

First, a widget can be inserted into an empty area (Figure 4). To identify the correct empty area, we attempt to snap the dragged rectangle to existing tabstops. If the widget can only be snapped on one side in a certain direction, horizontally or vertically, a new tabstop is inserted at the opposite side. If a widget cannot be snapped to any tabstop in a certain direction, then two other suitable tabstops have to be found. These are the respective tabstops of the largest empty rectangular area at the pointer position.

The combination of snapping widgets to existing tabstops and placing them into the largest empty area at the pointer makes the move operation quite versatile. For example, it is possible to place a small item accurately into the corner of an empty area by snapping it on two borders. Dropping a widget roughly in the middle of an empty area will fill the area with the widget. Furthermore, a widget can, e.g., be placed in the left half of an empty area by moving the item's left border close to the area's left border midway between top and bottom. ALE visualizes the location where a widget will be placed when "dropped" with a green rectangle (Figure 4).

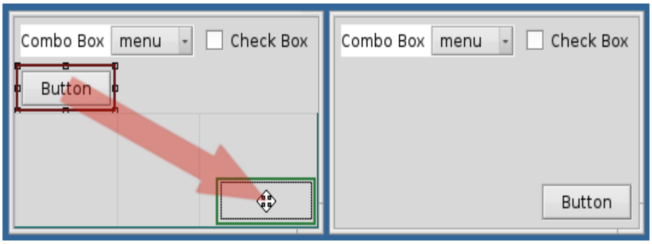


Fig. 4. Move operation: Dragging the button to the bottom-right of the empty area. The area where the button will be inserted when dropped is highlighted in green.

Second, a widget can be placed between an existing tabstop and a widget adjacent to that tabstop. This happens when a widget is dropped close to an existing tabstop and on the adjacent existing widget (Figure 5). In the following, only the insertion at a vertical tabstop is described; the horizontal case works analogously. When dropping the dragged widget, it is inserted so that its top and/or bottom are connected to those of the existing widget, using the same rules as in the first case: if the top or bottom of the dragged rectangle is close to the top or bottom of the existing widget, then they are snapped together. If only one side is snapped, either top or bottom, a new tabstop is created at the opposite side. If the dragged rectangle is not close enough to the top or bottom of the existing widget, then the dragged widget is connected to both the top and bottom of the existing widget. In Figure 5, the string view is snapped to the bottom of the right button. The left and right tabstops of the moved widget are then set so that the widget is placed between the vertical tabstop and the existing widget.

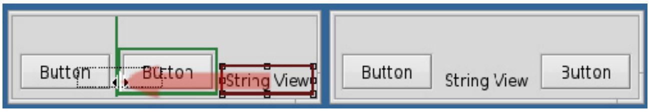


Fig. 5. Move operation: Moving a string view between two buttons

Swapping Two Widgets. This operation is triggered by dropping a widget onto another existing widget. It simply swaps the position of the two involved widgets. Here it is sufficient to connect the moved widget to the tabstops of the other widget, and vice versa.

Resizing a Widget. Dragging one of the borders or corners resizes a widget and allows the user to connect the dragged borders to different tabstops. During resizing, all relevant tabstops are visualized as light blue lines to aid alignment. A resize operation is aborted if an enlarged widget would overlap other widgets.

There are two cases to consider for resizing. First, a widget can be resized to an existing tabstop, by dragging and snapping it to said tabstop (Figure 6). Second, an item can be resized to match its preferred size when a dragged border is released without snapping it to a tabstop. In this case, a new tabstop is inserted for the dragged border and its position is calculated so that the widget gets its preferred size. Finally, the resize behavior of widgets can be controlled manually via the preferred size settings in a properties window (see right side of Figure 2).

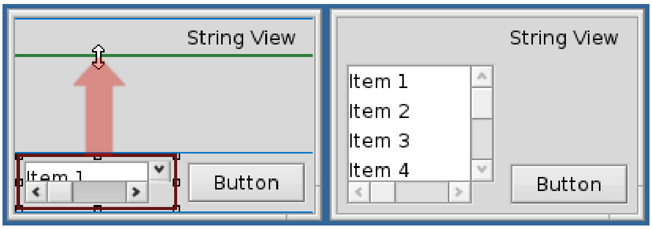


Fig. 6. Resizing the top of the list widget to the bottom tabstop of the string view

Inserting and Removing Widgets. Inserting a new widget is practically identical to the move operation. A widget can be removed from the layout by dragging it outside of the layout and dropping it.

4.1 Filling Gaps

A move, resize or remove operation can disconnect adjacent widgets that were only connected through the one widget that was edited. In this case, a gap may appear in the layout (Figure 7) and the position of some widgets may become undefined in the constraint system; the widgets are "floating".

ALE avoids such situations by checking for disconnected widgets after move, resize and remove operations. If the layout contains widgets or groups of widgets that are disconnected, then these widgets are moved one after another into the direction of the removed or resized widget. If a group was connected to the right side of the removed or resized item, the group is moved to the left, and similarly for other directions (Figure 7). All disconnected groups are moved until they are connected directly or indirectly to at least one horizontal and one vertical layout border.

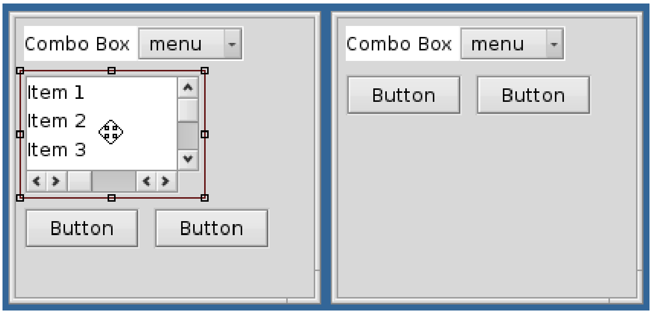


Fig. 7. Filling gaps: Upon removal of the list widget (left), the two buttons are both "floating" vertically. Moving the group of the two buttons to the top fills the gap. The top of the buttons is now connected to the bottom tabstop of the combo box (right).

4.2 Non-overlapping Layouts

ALE guarantees that all created layouts are non-overlapping. In other words, widgets do not intersect each other, regardless of the window size. This ensures that all widgets are completely visible and accessible at all GUI sizes. Non-overlap for a layout also implies that no widget intersects the boundaries of the GUI. Note that we need to differentiate between concrete layouts, as they are rendered on a screen, and *layout specifications*, which can be rendered at many different sizes and hence lead to many concrete layouts. For example, the layout in Figure 8 is non-overlapping, but as the layout size is reduced, the check boxes overlap the buttons due to a missing constraint.

ALE's operations are designed to avoid overlap. For this, the layout is tested after each operation for possible overlap. If necessary, additional constraints are added. For example, if the user resizes the layout in Figure 8 during the build process as shown, ALE would add a constraint that keeps a positive distance between the checkbox and the buttons.

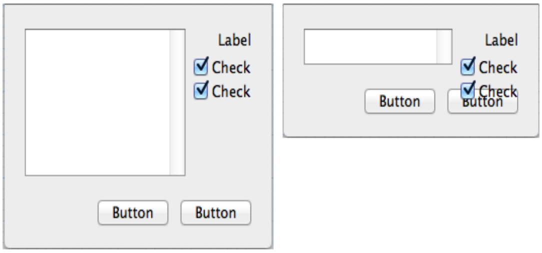


Fig. 8. The constraint-based layout on the left appears sound. Yet, when decreasing its size, as shown on the right, overlap between widgets occurs due to a missing constraint. The used constraint-based GUI builder, Apple Xcode, gave no indication of this issue.

5 Evaluation

We considered two hypotheses, addressed in the following Experiments 1 and 2:

H1 ALE makes it easier to create layouts from scratch compared to other available GUI builders.
H2 ALE makes it easier to edit existing layouts compared to other available GUI builders.

Experiment 1 compared layout creation in ALE with a popular gridbag GUI builder (**H1**). Experiment 2 compared ALE with a modern constraint-based GUI builder, considering both creation of new layouts and editing of existing layouts (**H1** and **H2**).

5.1 Experiment 1: Comparison with a Gridbag GUI Builder

In the first experiment, we targeted **H1** by investigating how our approach performs in comparison to a state-of-the art GUI builder. For this comparison we chose the GUI builder in MS Visual Studio 2010 (VS) as a representative for the state of the art, since it was popular at the time of the study. A prior study compared the usability of the gridbag and the constraint-based layout model, finding that the gridbag layout model has advantages in the creation of new layouts [19]. This supports our choice of a gridbag GUI builder such as VS for comparison with ALE when investigating **H1**.

16 participants, mostly software engineering students with experience in GUI development, were asked to perform four GUI creation tasks, each either with ALE or with VS. In each task, they were asked to rebuild a realistic GUI layout from a sample screenshot. Figures 9, 10 and 11 show the four tasks. We measured task completion time as an indicator of efficiency, and used a post-questionnaire to determine participants' preferences.

With the VS GUI builder it is not easy to modify the row- and column-span in a gridbag layout, since this cannot be done visually. It can only be achieved by opening a properties dialog. Thus, participants were instructed to nest multiple gridbag layouts, permitting users to recreate the target layouts more easily.

For both ALE and VS, a training task was given before the respective main tasks to ensure a reasonable amount of training with both tools. To counteract potential learning effects, half the participants were allocated to a group which first performed the training and tasks I and II with ALE, and then the training and tasks III and IV with VS. The second half used the tools in the opposite order.

If there were errors in the layout after the participants had finished a task, e.g. a widget was placed erroneously, the experimenter indicated the errors to them. Afterwards, timing continued and the participants had to fix these errors. In this way, all participants were able to succeed in all tasks.

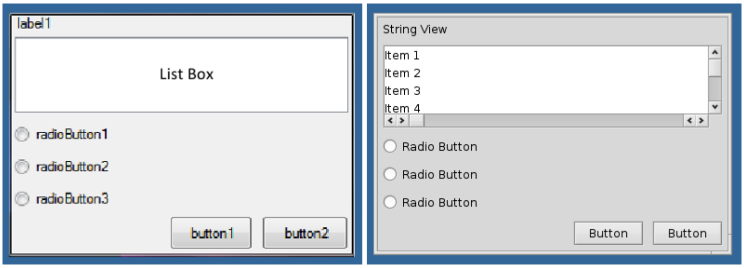


Fig. 9. Experiment 1, Task I: VS on the left and ALE on the right side

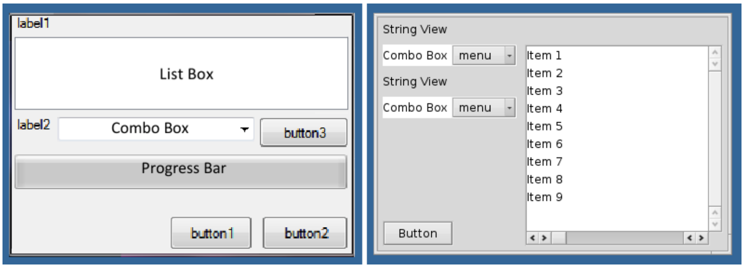


Fig. 10. Experiment 1: Task II for VS on the left and Task III for ALE on the right side

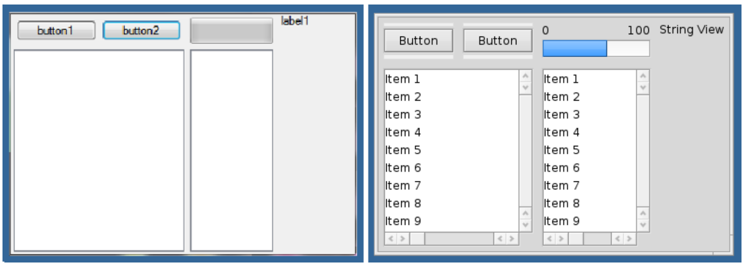


Fig. 11. Experiment 1, Task IV: VS on the left and ALE on the right side

Results and Discussion. The measurements were not normally distributed. The medians of ALE and VS were 74 and 188 seconds, respectively. A Wilcoxon signed-rank test identified a significant effect of the GUI builder (Z = -5.31, $p < 0.0001$), which supports **H1**. Pairwise Wilcoxon signed-rank tests show that ALE was significantly faster than VS for every task, with $p < 0.01$ or better. Figure 12 shows the individual times broken down by task. According to the post-questionnaire, 11 of the 16 participants preferred ALE over VS. A separate study is necessary to determine what exactly made ALE perform better than VS. It was not the aim of our current work, as we only aimed to evaluate how ALE competes with the gridbag-based VS builder at the task level.

One potential threat to validity is the fact that in VS participants did not use a single gridbag layout, but a nested gridbag layout with a column- and row-span of one. According to observations during the experiment, many participants had difficulties when nesting multiple layouts to create the desired outcome with VS, even though this was easier in the VS GUI builder compared to creating a single gridbag layout. A possible explanation is that a gridbag layout specification has to be understood more thoroughly upfront, and cannot easily be developed on the fly during the design process as with a constraint-based layout approach.

Note that for this experiment, a slightly older ALE version was used [18] than for Experiment 2. Although this version was less polished and had some small usability problems, all operations discussed here were available. It is possible that the different versions of ALE would differ slightly in their performance.

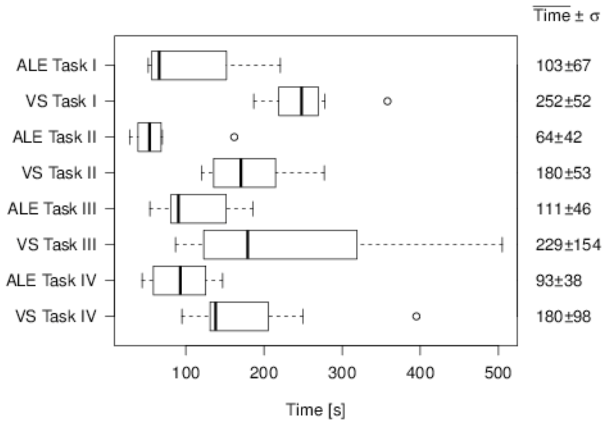


Fig. 12. Boxplot of the task completion times for all four tasks of Experiment 1

5.2 Experiment 2: Comparison with a Constraint-Based GUI Builder

In a second experiment, we tested both **H1** and **H2**. Apple's Xcode offers currently the only easily available GUI builder for constraint-based layouts. Xcode's layout model is similar to ALE's layout model, supporting simple linear hard and soft constraints. Also, Xcode permits free placement and resizing of widgets on the editing canvas. Consequently, we compared ALE with Xcode in this experiment.

The evaluation involved two main tasks (Task V and VI), preceded by a training task that was similar to the main tasks. Each participant performed all tasks once with ALE and once with Xcode. To eliminate order effects, half of the participants started with ALE, the other half with Xcode. Each task was divided into a layout creation subtask and three editing subtasks.

The first editing subtask (a) required swapping two widgets; the second editing subtask (b) required moving a widget to a position between two other widgets (see Figure 5). For these two subtasks, we expected ALE to perform better because users need only a single operation, while in Xcode multiple operations are necessary. With the free placement approach of Xcode, moving a widget between two other widgets requires the user to first move at least one of the other widgets aside to make room for said widget. Furthermore, it is necessary to manually fill the empty space that the moved widget has left behind. The last editing subtask (c) was more complex and required a combination of multiple edit operations. Figure 13 shows the layout for the creation subtask for Task V, and Figure 14 shows the editing subtask (c) for Task VI. In Xcode, participants were asked to align widgets as well as possible via the snapping functionality provided by the builder. As with the previous experiment, after each task, timing stopped, the experimenter pointed out layout errors to the participants, timing continued and participants had to fix the errors.

After finishing the tasks using Xcode and ALE, the participants were given a Likert-scale questionnaire to gather general information and to analyze their preferences. Furthermore, they were asked for comments in an open-ended question.

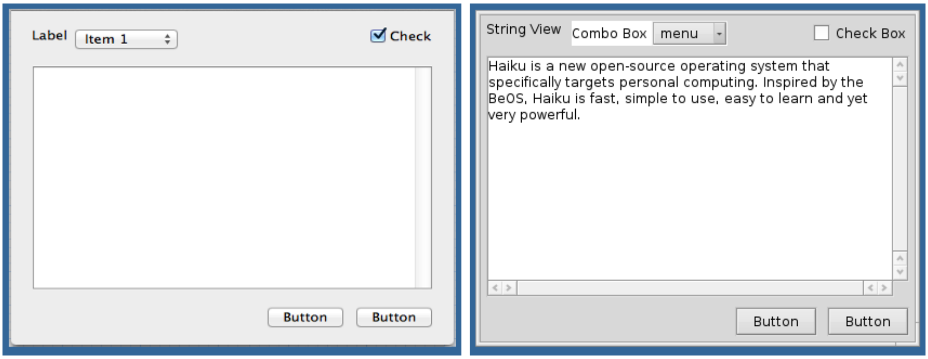


Fig. 13. Experiment 2, Task V: Xcode on the left and ALE on the right side

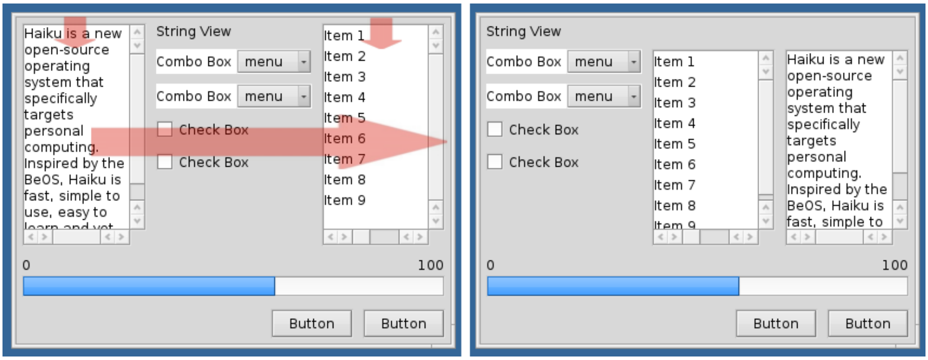


Fig. 14. Experiment 2, Task VI (c): The complex editing subtask

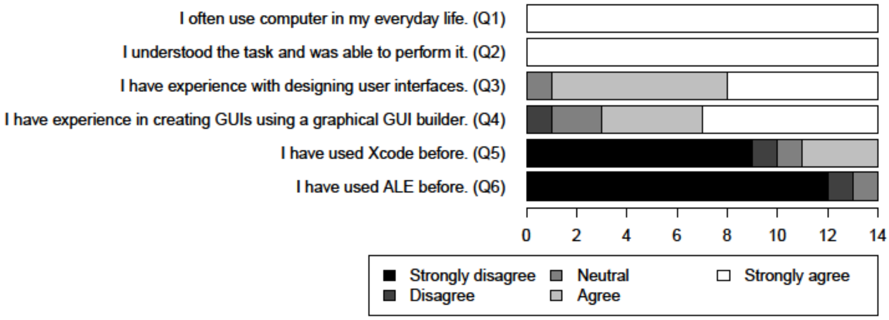


Fig. 15. Results for the general Likert-scale questions of Experiment 2

Results and Discussion. For this study, 14 participants were recruited, mostly graduate Computer Science students. All of them stated that they had understood the tasks. Most participants had experience with user interface design and were familiar with creating GUIs using a GUI builder. Four participants had used Xcode before, while only one participant had used ALE before (see Figure 15).

Overall, the measured task completion times were not normally distributed. The results for the creation tasks are shown in Figure 16.

The medians of all creation times for ALE and Xcode were 66 and 80 seconds, and for all editing times 9 and 38 seconds, respectively. Figure 17 depicts the results for the editing subtasks of Task V. The results for the editing subtasks of Task VI are shown in Figure 18. A Wilcoxon signed-rank test identifies a significant effect of the GUI builder for creation (Z = -2.05, p < 0.05), which supports **H1**. There is also a significant effect for editing tasks (Z = -9.19, p < 0.0001), which supports **H2**. For layout creation and layout editing, ALE was clearly faster for Task V and Task VI. Pairwise Wilcoxon signed-rank tests show that ALE was significantly faster than Xcode for every creation and editing subtask, with p < 0.01 or better. The swapping (a) and the moving (b) subtasks were much faster using ALE.

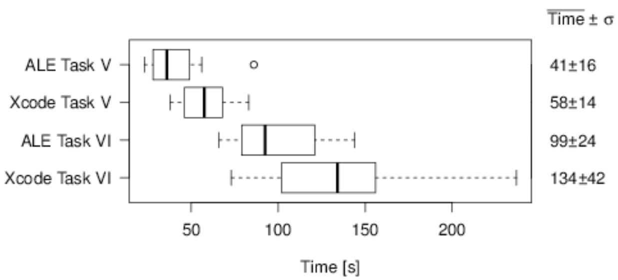


Fig. 16. Task completion times for layout creation in Experiment 2, Task V and VI

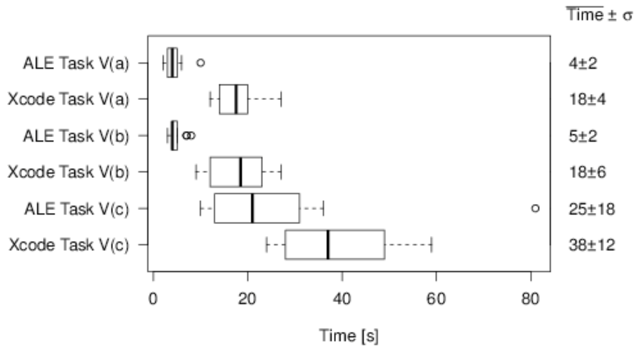


Fig. 17. Task completion times for layout editing in Experiment 2, Task V

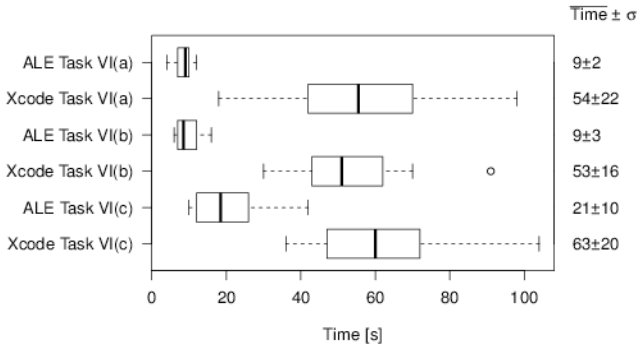


Fig. 18. Task completion times for layout editing in Experiment 2, Task VI

The results from the post-experiment questionnaire show a consistently positive response (Figure 19). Most participants preferred ALE and found it easier for creating and editing layouts. Furthermore, participants enjoyed ALE more than the Xcode builder.

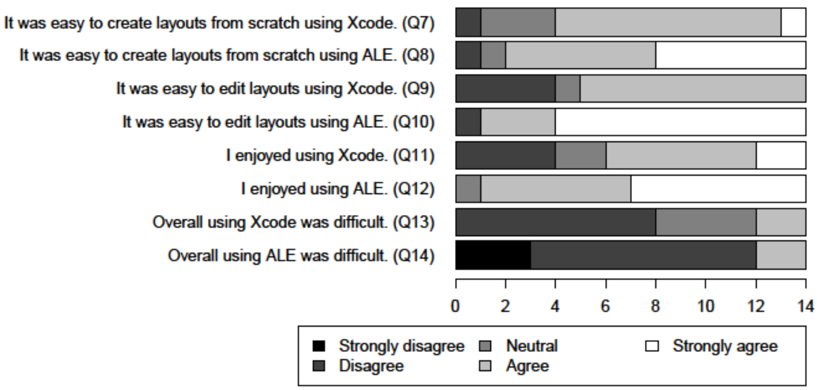


Fig. 19. Results for the Likert-scale usability questions of Experiment 2

Several participants commented in the open-ended question that they liked the swap operation and that layout editing was easier. Other comments pointed out that "*one first had to get used to the different concepts of ALE, e.g. that [a widget] cannot be placed freely.*" Another stated: "*I would imagine that ALE may perhaps be very efficient when acquainted with.*" This is consistent with our observations that participants needed more time to get used to ALE. However, after the training phase most participants were able to perform all tasks without problems.

One noteworthy observation is that with the free placement approach of Xcode it was more difficult to align items precisely. Participants made many erroneous alignments or seemed to be unaware of misalignments in the layout. Another observation is that when using Xcode, participants often first aligned a newly inserted or modified widget precisely to some other widgets, and realized only later that they had to align said widget again to achieve the target layout. ALE avoids this problem by automatically keeping widgets aligned, which is one contributor to the much shorter completion times. A threat to validity is that in Xcode, resizing a text or list widget was sometimes difficult, as users had difficulties clicking the resize handles. However, users also experienced similar minor usability problems in ALE.

Another threat is that layouts in both experiments were relatively small. Since the evaluation took already about an hour we tried to keep the layouts small. However, we believe that the layouts have a reasonable size because for larger layouts one would naturally start to use nested layouts.

6 Conclusion

We presented ALE, a GUI builder that makes it possible to quickly create and edit constraint-based layouts with simple mouse operations. ALE defines a set of layout edit operations that hide much of the complexity of constraint-based layouts from the designer. These edit operations are moving, swapping, resizing, inserting and removing of widgets. All operations maintain alignment and establish appropriate constraints automatically. ALE automatically aligns widgets to each other when placing a widget. It also moves other widgets aside if necessary. Swapping of widgets does not require manual resizing or moving of the adjacent widgets. Additionally, ALE automatically keeps layouts non-overlapping.

In two comparative evaluations, we found that ALE permitted participants to construct several realistic layouts significantly faster than with current commercial solutions, both for a gridbag and a constraint-based layout. Furthermore, we found that editing existing constraint-based layouts is also significantly faster with ALE. Participants enjoyed using ALE, and once familiar with the new edit operations, found it easier to use.

This evaluation demonstrates that it is feasible to utilize the power of constraint-based layouts in graphical GUI builders. The encouraging results from the experiments illustrate also that operations that automatically keep widgets aligned can result in a substantial boost in productivity. Overall, we see this as an indication that there is ample potential for improvements in today's GUI builders.

Future Work. Ideally, a GUI builder should be able to guarantee that all created layouts are non-overlapping, regardless of window size. ALE addresses this by automatically adding constraints when potential overlap is detected. We will describe this functionality in more detail in future work.

ALE's edit operations are already powerful enough to create many common layouts. However, a constraint-based system can be used to create far more complex layouts. For that, general constraint editing has to be integrated into ALE. We will investigate this in the future as well.

Acknowledgements. We would like to thank Rishika Mukerjee and Keerthana Puppala for helping to design and perform the first experiment.

References

1. Avrahami, G., Brooks, K.P., Brown, M.H.: A two-view approach to constructing user interfaces. SIGGRAPH Comput. Graph. 23(3), 137–146 (1989)
2. Badros, G.J., Borning, A.: The Cassowary Linear Arithmetic Constraint Solving Algorithm: Interface and Implementation. Tech. rep., ACM Transactions on Computer Human Interaction (1998)
3. Fletcher, R.: Practical methods of optimization, 2nd edn., ch. 10.3. Wiley Interscience (1987)
4. Gajos, K., Weld, D.S.: Supple: automatically generating user interfaces. In: Proc. 9th International Conference on Intelligent User Interfaces, IUI 2004, pp. 93–100. ACM, New York (2004)
5. Galitz, W.: The essential guide to user interface design: an introduction to GUI design principles and techniques. Wiley (2007)
6. Hashimoto, O., Myers, B.A.: Graphical styles for building interfaces by demonstration. In: Proc. 5th Annual ACM Symposium on User Interface Software and Technology, UIST 1992, pp. 117–124. ACM (1992)
7. Heim, S.: The resonant interface: HCI foundations for interaction design, ch. 6.6. Pearson/Addison Wesley (2007)
8. Jamil, N., Müller, J., Lutteroth, C., Weber, G.: Extending linear relaxation for user interface layout. In: IEEE 24th International Conference on Tools with Artificial Intelligence, ICTAI 2012, Athens, Greece (2012)
9. Karsenty, S., Weikart, C., Landay, J.A.: Inferring graphical constraints with Rockit. In: Proc. INTERACT 1993 and CHI 1993 Conference on Human Factors in Computing Systems, CHI 1993. ACM (1993)
10. Lutteroth, C., Strandh, R., Weber, G.: Domain Specific High-Level Constraints for User Interface Layout. Constraints 13(3) (2008)
11. Myers, B.A.: User-Interface Tools: Introduction and Survey. IEEE Software 6, 15–23 (1989)
12. Scoditti, A., Stuerzlinger, W.: A new layout method for graphical user interfaces. In: 2009 IEEE Toronto International Conference on Science and Technology for Humanity (TIC-STH), pp. 642–647. IEEE (2009)
13. Singh, G., Kok, C.H., Ngan, T.Y.: Druid: A system for demonstrational rapid user interface development. In: Proc. 3rd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology, UIST 1990, pp. 167–177. ACM (1990)

14. Vlissides, J.M., Tang, S.: A unidraw-based user interface builder. In: Proc. 4th Annual ACM Symposium on User Interface Software and Technology, UIST 1991, pp. 201–210. ACM (1991)
15. Weber, G.: A reduction of grid-bag layout to auckland layout. In: 2010 21st Australian Software Engineering Conference (ASWEC), pp. 67–74 (April 2010)
16. Zanden, B.V., Myers, B.A.: The Lapidary graphical interface design tool. In: Proc. SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology, CHI 1991, pp. 465–466. ACM (1991)
17. Zeidler, C., Lutteroth, C., Weber, G.: Constraint solving for beautiful user interfaces: how solving strategies support layout aesthetics. In: Proc. 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on HCI, CHINZ 2012, pp. 72–79. ACM (2012)
18. Zeidler, C., Lutteroth, C., Weber, G., Stürzlinger, W.: The Auckland layout editor: an improved gui layout specification process. In: Proc. 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, CHINZ 2012, pp. 103–103. ACM, New York (2012)
19. Zeidler, C., Müller, J., Lutteroth, C., Weber, G.: Comparing the usability of grid-bag and constraint-based layouts. In: Proc. 24th Australian Computer-Human Interaction Conference, OzCHI 2012, pp. 674–682. ACM, New York (2012)

KLM Form Analyzer: Automated Evaluation of Web Form Filling Tasks Using Human Performance Models

Christos Katsanos[1], Nikos Karousos[1], Nikolaos Tselios[2], Michalis Xenos[1], and Nikolaos Avouris[2]

[1] Hellenic Open University, Parodos Aristotelous 18, Patras, Achaia 26335, Greece
{ckatsanos,nkarousos,xenos}@eap.gr
[2] University of Patras, Rio, Achaia 26500, Greece
nitse@ece.upatras.gr, avouris@upatras.gr

Abstract. Filling forms is a common and frequent task in web interaction. Therefore, designing web forms that enhance users' efficiency is an important task. This paper presents a tool entitled KLM Form Analyzer (KLM-FA) that enables effortless predictions of execution times of web form filling tasks. To this end, the tool employs established models of human performance, namely the Keystroke Level Model and optionally the Fitts' law. KLM-FA can support various evaluation scenarios, both in a formative and summative context, and according to different interaction strategies or modeled users' characteristics. A study investigated the accuracy of KLM-FA predictions by comparing them to participants' execution times for six form filling tasks in popular social networking websites. The tool produced highly accurate predictions (89.1% agreement with user data) in an efficient manner.

Keywords: Web form design, task efficiency, user performance time, automated tool, human performance models.

1 Introduction

Usability of interactive web forms is a critical aspect of the overall user experience. Form filling is a data entry task, and thus user efficiency is of particular importance in the design of web forms. Current design practices are mostly empirical and rely on guidelines derived from experimental studies comparing alternative designs and usability experts' experience or observations. For instance, the type of form elements as well as their positioning in the form layout significantly affect users' performance [1].

One may argue that theoretically-based approaches have had a limited impact on web form design practices. Unlike desktop [2] or mobile interfaces [3-4], GOMS [5] and its simplified version Keystroke-Level Model (KLM) [5-6], have been rarely used to guide web form design or evaluation. In addition, if field size and position on the form layout are not taken into account in such model-based techniques, superficial results may arise. For instance, interaction with a dropdown menu theoretically takes

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 530–537, 2013.

longer than interaction with radio buttons. This is due to an additional point and click needed to open the dropdown menu. However, in one study the latter hypothesis was confirmed [7] and in another it was rejected [8].

As a result, there is a need to bridge HCI models, such as KLM, with design and evaluation practices. Previous research [9] resulted in the development of CogTool, a tool that can produce quantitative, model-based predictions of skilled performance time from tasks demonstrated on storyboard mockups of a user interface. CogTool-Explorer [10] builds upon CogTool to predict a user's goal-directed exploratory interaction with a website. Currently available model-based tools require non-trivial manual work to examine forms. In addition, if a large scale summative evaluation is needed, the evaluator has to repeat the same process without any particular assistance. Furthemore, the plethora of available functions and generic modeling nature of existing tools can overwhelm and discourage practitioners who, in most cases, need a simple tool focused on the problem at hand.

To tackle the aforementioned problems, in this paper a novel tool entitled KLM Form Analyzer (KLM-FA) is presented. KLM-FA extends the capabilities of existing modeling tools for practitioners by focusing specifically on automating the analysis of web forms. The paper is organized as follows: The tool functionality and usage is delineated in the next section, along with its internal architecture and reasoning. Finally, a validation study comparing KLM-FA results to human performance data is presented and discussed.

2 The KLM Form Analyzer Tool

The main objective of KLM-FA (available at http://klmformanalyzer.weebly.com) is to support design and evaluation of web forms in an effective and efficient manner. The tool employs web parsing algorithms, coupled with KLM and Fitts' modeling to estimate the time required to fill a web form according to different interaction strategies (e.g. using tab to move across the elements) or users' characteristics (e.g. age and typing expertise). Figure 1 presents the main interface and functionality of KLM-FA.

2.1 KLM-FA Typical Usage Scenario

First, the evaluator inputs the URL of the web form to be evaluated or selects a previously evaluated form. Next, the evaluator selects a set of analysis preferences related to the modeled user profile (typing ability, age), usage (or not) of Fitts' law in the calculations, and hypotheses about the interaction, such as initial cursor position and whether the user moves across form elements using the mouse or the keyboard. The evaluator can also assign a predefined field type to text elements (e.g. username, email) to easily specify their number of keystrokes. The tool provides an editable list of field types that covers most of the elements used. The default typical field entry lengths rely on empirical data available in the literature (e.g. mean password length [11]) and a dataset of our own with 839 registered web users' personal data.

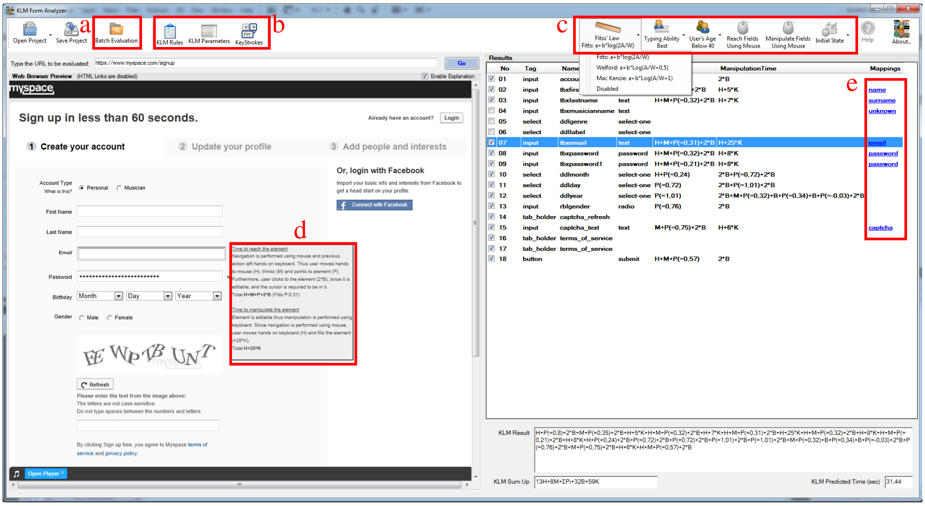


Fig. 1. Overview of the KLM-FA interface and functionality: (a) Mass scale evaluation, (b) Analysis rules and parameters, (c) Analysis preferences, (d) Tooltip explaining KLM modeling for the selected field, (e) Semantic mapping of a text-entry element to a number of keystrokes

Next, KLM-FA runs an algorithm which parses the evaluated form, produces the sequence of predicted user actions (KLM operators) based on the evaluator's selected analysis preferences and estimates task completion time based on a set of analysis parameters related to KLM and Fitts' law calculations. Based on empirical data [12-13], the tool provides a set of default values for the analysis parameters, which can be easily modified through appropriate dialogues. The internal architecture and algorithms employed by the tool are delineated in the next section.

The output of the tool is an interactive web form preview synchronized with a results list: when an element is selected in the web form preview it is highlighted in the results list and vice versa. Depending on the evaluation scenario, mental operators can be added or any element can be excluded from the analysis by simply unchecking it from the results list. In all cases, the tool updates the results in real time. Furthermore, KLM-FA provides an option that elaborates the underlying KLM modeling for each element in a tooltip. In this way, one can trace step-by-step the KLM modeling analysis by simply selecting the sequence of the form elements either in the web preview or in the list. Each evaluated form can be saved and/or subsequently modified. In addition, KLM-FA can employ mass scale summative evaluations by selecting a set of saved projects. Then, the tool runs an analysis of all the selected forms using the same settings for all projects and saves the results in an XML file.

2.2 Internal Architecture and Reasoning

The KLM-FA internal architecture comprises two conceptual layers: the user interface layer, responsible for the interaction with the tools' user, and the KLM analysis layer (named *KLMKernel*). The latter handles the elements identification through webpage parsing and the KLM modeling calculations. The fundamental data structure of the KLM-FA is the *Element* class, which is used to represent each element. Table 1 presents its main data members. The *KLMKernel* first parses a given form to produce a structured list of *Elements* (named *ElementList*) and then performs the KLM analysis and updates the *ElementList* with the results.

Table 1. Data members of the *Element* class, the fundamental data structure of KLM-FA

Data member	Brief Description
TagName	HTML form tag (e.g. <select>) or special KLM-FA tag (e.g. tabholder)
Type	HTML type of <input> tag (e.g. password, text, radio etc.)
Name	Value of the HTML name attribute
Choices	Number of choices for radio buttons and drop-down lists
MappedField	Semantic mapping of a text-entry element to a number of keystrokes
Position	X and Y coordinates of the element position in the form layout
Size	Width and height of the element
MentalExtras	Mental operators manually added or removed
Active	Flag to denote whether the element is included in the calculations
ReachTime	Predicted time required to reach the element
ManipulationTime	Predicted time required to manipulate the element
KLMexplanation	Explanatory text describing the rationale of the derived KLM operators

Parsing Module. It is responsible for parsing a provided webpage to identify existing forms and their elements. To this end, it employs two separate algorithms: a) *form identifier*, and b) *element identifier*. The *form identifier* parses the HTML DOM loaded in the internal browser and finds all forms. It filters out forms that cannot be eligible for analysis (hidden) and presents a "select form" dialogue if two or more forms are found. Then, the *element identifier* parses the selected form, identifies and stores visible fields in the internal *ElementList*. Currently, the tool cannot identify fields when either Flash or AJAX is used. However, KLM-FA provides support to manually add fields and specify their properties in a straightforward manner (e.g. clicking on unidentified field registers its position and size). The following pseudocode sketches the *element identifier* algorithm which produces an updated *ElementList*.

```
GetFormElements(FormNode, ElementList){
  foreach Element in FormNode.Elements
      if (validate_element(Element))
        if (Element.Type == "radio")
          calculate_middle_Element(formNode, Element)
        else
          if (Element.Type == "select")
            calculate_select_options(formNode, Element)
      ElementList.Add(Element)
      Element.Active = isElementInsideHiddenDiv(Element)
}
```

Analysis Module. This module performs the KLM modeling and related calculations. It takes as input the *ElementList* along with the following parameters:

- evaluator-defined preferences concerning modeled users' typing proficiency and age, mouse or keyboard usage for navigation and manipulation of the elements, Fitts' Law activation, and initial position of the user's hands and cursor,
- predefined time values for KLM operators and Fitts' constants,
- paired list of *[fieldname-keystrokes]* that is used for text entry calculations, and
- set of KLM analysis rules regarding placement of mental operators, and other specific modeling assumptions (e.g. manipulation of dropdown lists with keyboard).

For each form element the algorithm produces the sequence of required actions (KLM operators) to first reach it (ReachTime) and then manipulate it (ManipulationTime). This distinction enables flexible modeling of various user interaction strategies (e.g. tab-based navigation). In addition, the algorithm creates an explanatory text of the KLM modeling rationale which can be displayed as a tooltip in the web preview form.

Concerning Fitts' law, the analysis module calculates the pointing operator by storing the previous position of a simulated mouse cursor and updating it whenever the modeling process requires its movement to a new position. The MacKenzie-Shannon formula and constants [2] for Fitts' law are the default selection for modeling pointing device movement time. However, given the lack of consensus on the Fitts' formula [14], the tool offers additional options (e.g. Welford's formulation [2]) and it is also easy to add further formulas or modify constants values.

Finally, KLM-FA sums up the results and produces a sequence of operators and the predicted form completion time for the provided analysis preferences and parameters. The entire form analysis concludes to an updated *ElementList* that can be saved, reanalyzed with a different set of parameters or exported to an XML file. In addition, the form analysis algorithm can be executed for a set of saved forms (*ElementLists*) to rapidly produce massive KLM modeling results for the same set of analysis parameters. The following pseudocode sketches the *form analysis* algorithm.

```
Analyze(ElementList){
  TypeElement prev_el;
  foreach element in ElementList
      if (is_active(element))
        prev_el = ElementList.GetPreviousActiveElement
                   (element, nav_using_mouse());
        if (Fitts_Law is_enabled() and nav_using_mouse())
            estimate_FittsP_Reach(element);
        analyze_element_reach(element, prev_el);
        analyze_element_manipulation (element);
}
```

3 Validation Study

The aim of the study was to investigate the accuracy of the results obtained by using KLM-FA. The study compared the KLM-FA predictions with user testing data for three signup forms of popular social networking websites: facebook, twitter, and myspace. For each form, two interaction strategies were investigated: a) *mouse-based*, in which a user is assumed to interact with the form using the mouse, except for input in text entry fields, b) *keyboard-based*, in which form fields are reached using the tab key and manipulated only through the keyboard. In both interaction strategies, users were assumed to fill the fields following the form layout. All in all, times for a total of six form-filling tasks (3 forms x 2 interaction strategies) as calculated by KLM-FA and measured through user testing were compared.

Fifteen University students, 12 male, with a mean age of 27 (sd=5.8), a mean of 14 years of QWERTY keyboard usage (sd=4.2) and a mean typing speed of 42 corrected words per minute (sd=16) took part in the study. First, participants completed a short online demographics questionnaire and a typing speed test. Next, they were asked to perform 10 trials for each of the six tasks and their behavior was monitored by an in-house web-based software developed for the needs of the study. Ten trials have been used in similar studies [15] to allow users' to reach skilled performance. In this study, participants were allowed to perform additional trials if their tenth trial was not error-free (max number of trials observed = 12). Task execution times were derived from participants' last error-free trial.

In each trial, participants were first presented with an instructions webpage, followed by the actual form which appeared when they clicked on a link. In the instructions page, they were asked to familiarize themselves with the form registration data and were instructed to strictly employ a specific interaction strategy (i.e. mouse-based or keyboard-based) in order to fill the form as fast and correct as possible. In the form webpage, participants were first required to press a start button located in the top-left of the screen, which started logging of actions and ensured the same starting cursor position for all. The presentation order of both the forms and interaction strategies were counterbalanced to avoid serial order effects. Participants used an HP standard keyboard, an HP 3-button optical mouse and a TFT 17" screen with a resolution of 1280x1024. User sessions lasted about 75 minutes.

In KLM-FA, the following assumptions were used: a) the user was a poor typist (40 wpm) and aged below 40, b) system response time was negligible, c) the cursor's initial position was at the top–left corner of the page, d) tool defaults for all analysis parameters were used, apart from field entry lengths that were appropriately adjusted for each task, e) the user's hand began on the main device of each interaction strategy, and f) Fitts' law calculations were enabled in KLM-FA. KLM-FA analyses were also conducted on a TFT 17" screen with a resolution of 1280x1024. The process to evaluate all six tasks using KLM-FA required approximately 10 minutes.

Table 2 presents participants' task execution times and KLM-FA calculated times for each form and interaction strategy combination, along with the KLM-FA error rate. The error rate was calculated as the participants' mean task time minus the KLM-FA predicted time, and this difference divided by the participants' mean task time. Results show that the mean error of KLM-FA predictions was 10.9% (sd=6.4%), which is well within the 20% margin of error reported in the literature for KLM predictions in other contexts [6], [16]. The lowest and highest KLM-FA error rate values were 4.5% and 17.6% respectively. In general, KLM-FA tended to slightly overestimate (16.7% on average) and underestimate (5.1% on average) task time in the mouse-based and keyboard-based interaction strategies respectively.

Table 2. Study results showing means and, in parenetheses, standard deviations

Signup form	Interaction strategy	Participants' task time (ms)	KLM-FA predicted time (ms)	Error rates of KLM-FA predictions (%)
Facebook	Mouse-based	30739 (6742)	35320	14.9%
Facebook	Keyboard-based	27306 (7752)	25640	6.1%
Myspace	Mouse-based	33201 (6341)	39050	17.6%
MySpace	Keyboard-based	29641 (9310)	28320	4.5%
Twitter	Mouse-based	22478 (5146)	26420	17.5%
Twitter	Keyboard-based	23144 (6108)	24240	4.7%

4 Conclusions

This paper presents KLM-FA, a tool that employs predictive models of human performance to estimate execution times of web form filling tasks. In addition, a study is presented that demonstrates the accuracy of KLM-FA predictions by comparing them to human execution times for the same six form filling tasks.

KLM-FA extends the capabilities of existing general modeling tools for practitioners, such as CogTool [9], by focusing specifically on web form interaction. In this way, KLM-FA increases automation of evaluation tasks, minimizes the required effort and achieves increased simplicity and flexibility, thus increasing the chances of its adoption in actual practice. As a result, practitioners can rapidly evaluate alternative web form design approaches using a variety of scenarios. In addition, the ability of KLM-FA to evaluate keyboard-based interaction with web forms can be valuable in automated accessibility testing. KLM-FA can also be used to produce benchmark data of form completion times for specific web domains, such as social networking or e-commerce. Finally, the tools' step-by-step tracing of the KLM modeling supports learning through examples and thus can be valuable for both educators and students.

Investigating the effect of KLM-FA adoption on the learning outcome, while educating students in KLM, constitutes a future research goal. In addition, we plan to conduct additional studies that compare KLM-FA predictions with human performance data. An additional future research goal is to incorporate enriched models of KLM [17] in order to support design of web forms that enhance users' efficiency in mobile interaction contexts.

Despite the advantages of the presented automated approach, it only addresses task efficiency which is one aspect of the web user experience. Other tools that automate different aspects of web design are also available [18]. However, all such approaches should be used in conjunction with user-based methods.

Acknowledgements. This work has been partly funded by the European Territorial Cooperation Operational Programme "Greece-Italy 2007-2013" under the project Intersocial.

References

1. Bargas-Avila, J.A., Brenzikofer, O., Tuch, A.N., Roth, S.P., Opwis, K.: Working towards Usable Forms on the Worldwide Web: Optimizing Multiple Selection Interface Elements. Advances in Human-Computer Interaction 2011, Article ID 347171, 6 pages (2011)
2. Soukoreff, R.W., MacKenzie, I.S.: Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. International Journal of Human-Computer Studies 61(6), 751–789 (2004)
3. Myung, R.: Keystroke-Level Analysis of Korean Text Entry Methods on Mobile Phones. International Journal of Human-Computer Studies 60(5-6), 545–563 (2004)
4. Maragoudakis, M., Tselios, N.K., Fakotakis, N., Avouris, N.M.: Improving SMS usability using Bayesian Networks. In: Vlahavas, I.P., Spyropoulos, C.D. (eds.) SETN 2002. LNCS (LNAI), vol. 2308, pp. 179–190. Springer, Heidelberg (2002)
5. Card, S.K., Moran, T.P., Newell, A.: The Psychology of Human Computer Interaction. Lawrence Erlbaum Associates, Hillsdale (1983)
6. Card, S.K., Moran, T.P., Newell, A.: The Keystroke-Level Model for User Performance Time with Interactive Systems. Communications of the ACM 23(7), 396–410 (1980)
7. Healey, B.: Drop Downs and Scrollmice: The Effect of Response Option Format and Input Mechanism Employed on Data Quality in Web Surveys. Social Science Computer Review 25(1), 111–128 (2007)
8. Hogg, A., Masztal, J.J.: Drop-down, Radio Buttons, or Fill-in-the-blank? Effects of Attribute Rating Scale Type on Web Survey Responses. In: Proceedings ESOMAR 2001 (2001)
9. John, B.E., Prevas, K., Salvucci, D.D., Koedinger, K.: Predictive Human Performance Modeling Made Easy. In: Proceedings CHI 2004, pp. 455–462. ACM Press (2004)
10. Teo, L., John, B.E.: CogTool-Explorer: towards a Tool for Predicting User Interaction. In: Ext. Abstracts CHI 2008, pp. 2793–2798. ACM Press (2008)
11. Cazier, J., Medlin, B.: Password Security: An Empirical Investigation into E-Commerce Passwords and Their Crack Times. Information Systems Security 15(6), 45–55 (2006)
12. Kieras, D.: Using the Keystroke-Level Model to Estimate Execution Times. Unpublished report (2001)
13. Olson, J.R., Olson, G.M.: The Growth of Cognitive Modeling in Human-Computer Interaction since GOMS. Human-Computer Interaction 5, 221–265 (1990)
14. Drewes, H.: Only One Fitts' Law Formula Please! In: Ext. Abstracts CHI 2010, pp. 2813–2822. ACM Press (2010)
15. Luo, L., John, B.E.: Predicting Task Execution Time on Handheld Devices Using the Keystroke-Level Model. In: Ext. Abstracts CHI 2005, pp. 1605–1608. ACM Press (2005)
16. Ivory, M.E., Hearst, M.A.: The State of the Art in Automating Usability Evaluation of User Interfaces. ACM Computing Surveys 33(4), 470–516 (2001)
17. Holleis, P., Otto, F., Hussmann, H., Schmidt, A.: Keystroke-Level Model for Advanced Mobile Phone Interaction. In: Proceedings CHI 2007, pp. 1505–1514. ACM Press (2007)
18. Katsanos, C., Tselios, N., Avouris, N.: A Survey of Tools supporting Design and Evaluation of Websites based on Models of Human Information Interaction. International Journal on Artificial Intelligence Tools 19(6), 755–781 (2010)

The Evolution of Number Entry: A Case Study of the Telephone

Parisa Eslambolchilar, Julie Webster, and Gerrit Niezen

Computer Science Department, Swansea University, Swansea, UK
{p.eslambolchilar,j.webster,g.niezen}@swansea.ac.uk

Abstract. This paper details a user study to investigate serial digit entry on analogue and digital input platforms and the errors associated with them. We look specifically at the case of entering eleven-digit telephone numbers without a decimal point. The telephone is used as a platform for comparison, due to its clear evolution from a rotary dial to a pushbutton keypad and more recently, touch-based input. Thirty participants took part in a user study, which concluded that the touch interface was four times less accurate than the pushbutton and rotary dial interfaces. The latter two interfaces performed with similar accuracy; however, users were more than three times faster on the pushbutton keypad and recognized almost twice as many errors on the rotary dial. We have extended previous error taxonomies to include some errors relevant to sequences of numbers and built upon task-based guidelines found in the literature to suggest context-based design considerations.

Keywords: Number Entry, Interaction Design, Usability, Telephone Interfaces.

1 Introduction

The design of interactive systems has evolved over time in hopes of making them easier to use. Errors are typically blamed on a human factor, forgetting that the system should have been built to take them into account and attempt to minimise these types of problems [4]. The goal of interaction design is to aid the user in completing a task with a minimal number of errors or no error at all. An interaction that fails to accomplish this has likely failed at various levels.

Number entry is a task that is performed daily with little conscious thought. These number entry tasks seem trivial, but we still experience problems when performing them. Consider how many times we dial a phone number. It is rare to dial a wrong number, but it does happen occasionally and the results are potentially embarrassing. Similar problems occur with alarm systems where the wrong code results in a blaring sound or with personal identification numbers (PIN) for bank accounts that lock us out if entered incorrectly too many times in a row. All of these systems use a simple input method, such as a knob or keypad, laid out with either buttons or a touchscreen. In these cases, inputting a single wrong digit does not get us close to what we intended - it is instead entirely wrong. A wrong digit in a telephone number will result in speaking to a different person all together. This specific category of numbers,

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 538–545, 2013.

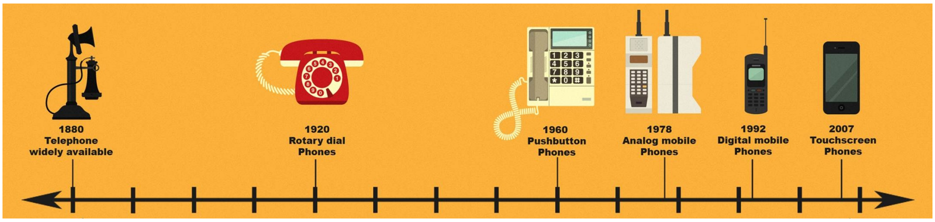


Fig. 1. The evolution of the telephone [5]

which includes telephone numbers, PIN numbers, and credit card numbers do not have a decimal point. The telephone gives us a good vehicle for exploration because it demonstrates a clear evolution from one input device to another, as shown in Figure 1.

2 Related Literature

This section details literature that has influenced this case study of the telephone by looking at errors and their taxonomies as well as number entry interfaces.

2.1 What is Error?

James Reason [3] defines two major categories of error: slips (or lapses) and mistakes. *Slips* result from the incorrect execution or incorrect planning of a correct sequence of actions, so an error is made even though a person has the knowledge needed to perform the number entry task. In contrast, *mistakes* occur when a person has incorrect or absent knowledge of the task they are aiming to complete. In other words, mistakes result from the correct execution of an incorrect sequence of actions. For example, consider the misinterpretation of feedback: Reading "121" as the expected value "12.1" could be a slip, whereas thinking a symbol signifies "On" when it signifies "Start" is probably a mistake.

2.2 Taxonomies of Error

Zhang et al. [7] proposed a cognitive taxonomy to categorise medical errors at the level of individuals, based upon Reason's taxonomy. Wiseman et al. [6] built on Zhang et al.'s taxonomy by focusing on number entry errors - errors that occur when a series of digits is being entered. In the medical domain, an example could be specifying values for an infusion pump, such as rate of infusion or the total volume to be infused. The study in [6] consisted of 20 participants each entering 30 numbers, where the length of numbers ranged from two to five digits. Half of the numbers entered were integers and the other half contained decimal points. A total of 350 errors were gathered and categorised into 21 classes of error, based on Zhang et al.'s taxonomy.

2.3 Number Entry Interfaces

Oladimeji [2] identified three categories of number entry interfaces:

- Serial digit entry – number is entered sequentially from left to right, e.g. telephones, ATMs or calculators,
- Independent digit entry – user controls the digits that make up the number separately, e.g. using up/down buttons to change the digit (0-9) and left/right buttons to change the position of the digit being entered, and
- Incremental number entry – user increases or decreases a number analogous to scrolling through values on a number-line, e.g. using up/down buttons, sliders or knobs.

There are certain trade-offs to consider when selecting a number entry interface, including speed of entry, severity of errors and space required for the interface. Oladimeji et al. [1] compared keypad interfaces to up/down button interfaces (press and hold to increase or decrease values) and showed a factor of two fewer unnoticed errors with up/down buttons. Fixation experiments showed that users are less likely to notice errors on the display when using keypads, as they fixate twice as much on the keypad itself (compared to the up/down buttons).

We focus our research on serial digit entry and the errors associated with these types of interfaces, when not using a decimal point. We focus on three types of serial digit number entry interfaces: the rotary dial, the pushbutton keypad and the touch display. The telephone was chosen as a platform for comparison due to its clear evolution from analogue interfaces to digital interfaces.

3 Experimental Setup

Existing studies on number entry research focus on a specific domain, for example entries on infusion pumps in the medical domain [2, 6]. The focus of this study was on more everyday interaction - serial digit number entry without a decimal point. This is the type of number entry we perform almost on a daily basis, for example entering PIN numbers or telephone numbers, where one wrong digit invalidates the whole number being entered.

A user study was performed to compare a rotary dial interface, touch interface and a keypad interface (Figure 2). A total of thirty participants completed the user study task. All were recruited through emails sent to the students and staff at the university and were affiliated with the university in some way. Most were postgraduate students and staff, 14 were male and 16 were female. Three participants were between 17 and 20 years old, 19 participants where in the 21-30 age group, four participants were between 31 and 40 years old, and four participants were between 41 and 50 years old. Participants were fluent in English for questionnaire purposes and were compensated for their time with a £5 shopping voucher. Before the sessions, the order of interfaces was randomly assigned and 50 11-digit numbers were randomly generated to allow all 30 users to enter the same 50 numbers on all three interfaces.

Participants were allowed five practice runs to familiarise themselves with the interface before the 50 11-digit numbers that were then recorded. The time it took to enter each number was recorded, along with the number entered compared to the number

Fig. 2. The three telephone-style interfaces used by participants: the rotary dial (left), pushbutton keypad (middle) and touchscreen keypad (right)

given to the participant. Participants were not provided with a display to see what digits they had entered, so they had to rely on the inherent tactile and visual feedback provided by the input devices. This was done to maintain consistency between the three interfaces, as a visual display would not have been used on older interfaces.

Participants were not allowed to correct their entries. If they perceived that they made an error, they were asked to notify the researcher and continue entering the rest of the number. This made it possible to determine how many errors would have been recognised, for a given number being entered. To alleviate fatigue, participants received a five-minute break before using the next interface. After thirty user sessions, data was collected from a total of 4,293 numbers, amounting to 46,807 digits in total, including digit additions and omissions due to errors. These results were analyzed in a variety of ways, detailed in Section 4.

3.1 Hardware Setup

A rotary dial interface from an early 1960s telephone was used as the analogue input device. When the dial spins, a cog inside moves the same number of clicks as the number selected, opening and closing a switch. For example, moving the number “6” around and letting go will cause the cog to engage a switch six times. An Arduino Uno microcontroller was used to count the number of pulses and each digit was outputted to the computer along with a timestamp.

A button interface for a FEZ Spider Kit[1] was used as the digital input device. The interface consisted of membrane buttons laid out as a traditional 3x4 keypad and was connected to the kit. An SD card module was used to store the numbers entered on the device, along with a timestamp for each digit.

The touchscreen interface was created on a resistive touchscreen with the FEZ Spider Kit. A 3x4 keypad was created on the touchscreen to look like the physical keypad. The LCD, touch-capable screen was attached to the mainboard and USB-powered modules of the kit. An SD card reader was utilized in the same way as on the pushbutton interface.

---

[1] The FEZ Spider Kit is a .NET Gadgeteer electronic set produced by GHI Electronics, consisting of a mainboard and a number of attachable modules.

Table 1. Total incorrect digits and correlation of error rate and order of interfaces

Interface	Total Incorrect Digits	Correlation
Rotary	190 (1.19%)	0.344
Button	180 (1.12%)	0.615
Touch	896 (5.81%)	**0.046**

A small button labelled "Next" was placed next to all three interfaces, which the user pressed to indicate that he/she was ready to enter the next number. The interfaces were mounted on platforms (Figure 2) to allow users to interact with the devices in the manner that they were the most comfortable.

4 Results

A digit-by-digit analysis was conducted of each number entered to determine what types of errors were made. Table 1 summarizes the total number of digits that were incorrect in each interface.

We used a subset of the error taxonomy of Wiseman et al. [10] to classify the errors made in our user study. Wiseman's taxonomy included numbers with decimal points, so some categories did not apply to our study. The following definitions were used:

- Digit(s) wrong – a different digit was entered from what was intended (e.g. '1' instead of '4')
- Digit missing – a single digit was omitted (e.g. '14' instead of '124')
- Digit added – a single digit was added (e.g. '124' instead of '14')
- Anagram – digits were reversed (e.g. '14' instead of '41')

Based on observations during in our study, we have added the following error classes to more fully understand user errors when entering whole numbers:

- Repeat digit missing – a single digit the same as the previous one was omitted (e.g. '14' instead of '144')
- Repeat digit added – a single digit the same as the previous one was added (e.g. '144' instead of '14')
- Repeat *n* digit pattern – a pattern *n* digits long was repeated (e.g. '1414' instead of '14')

Table 2 shows the total number of each type of error on each interface, as well as the frequency of errors on the interfaces. In the tables, n represents the number of errors made and r is the frequency, calculated to be $100n/N$ percent to represent the frequency that particular error occurred on that interface.

The most common error on the rotary dial interface was entering an incorrect digit. Missing digits were the next most common error, but they were almost eight times less likely than entering the wrong digit. A total of 180 digits were entered incorrectly on the button interface, slightly less than the rotary dial and more than five times less than the touch interface. The most common error on the button interface was adding one repeat digit. This may have occurred because users held down a button for too long hoping to make sure that it registered, but instead resulted in multiple instances

Table 2. Rotary Errors (R), Button Errors (B), Touch Errors (T)

Error Type	nR	rR (N=15928)	nB	rB (N=15906)	nT	rT (N=15389)
Wrong digit	120	0.75	33	0.21	150	0.97
Digit missing	28	0.16	34	0.21	509	3.31
Added 1 repeat digit	19	0.12	94	0.59	30	0.19
Added 1 digit	13	0.08	8	0.05	87	0.57
Added 2 repeat digits	4	0.03	-	-	-	-
Anagram	3	0.02	3	0.02	19	0.12
Repeat digit missing	-	-	6	0.04	96	0.62
2 digit pattern missing	2	0.01	-	-	4	0.03
Repeated 2 digit pattern	-	-	1	0.01	-	-
Repeated 3 digit pattern	1	0.01	1	0.01	1	0.01

of the same digit. A missing digit was the next most common error. The most common error on the touchscreen was a missing digit, likely a result of the user not tapping the screen hard enough to register their selection. A wrong digit and missing a repeat digit were the next most common errors but were roughly four and five times less common, respectively, than the omission of a single digit.

We conducted a mixed model analysis for each interface to determine if learning or fatigue affected the user as they completed the number entry task (Table 1). The analysis was performed by comparing the order that the interface was given to that participant and the error rate. Performance when the participant used the touchscreen first had a significance of 0.046 compared to a value of 0.022 when used second. No significant difference was observed for performance on two other interfaces.

Figure 4 shows a binary representation of all of the errors made on each interface, where each row represents a participant and each column represents a number being entered. Effects of learning would suggest that errors were concentrated at the beginning, while effects of fatigue would suggest errors concentrated at the end. Based on visual analysis of the figure, it appears that the rotary interface may have caused fatigue because there was a slightly higher concentration of errors at the end. In contrast, the touchscreen interface looks like it may have benefitted from learning as the user progressed through the task because the errors occurred slightly more often at the beginning. The button interface does not appear to have been affected by either because the errors look evenly spread across the use of the device.

5 Discussion

There are trade-offs to consider when selecting a number entry input device, as demonstrated by the data collected in our experiment. Our results showed that the button interface is only 0.07% more accurate than the rotary dial. Because the error rate is very close between these two interfaces, other factors such as the recognized error rate or the speed of entry may become important factors in selecting the most appropriate input device. Based on our results, the button interface is slightly more accurate may not always make it the right choice, because the recognized error rate was 13.80% higher on the rotary dial.

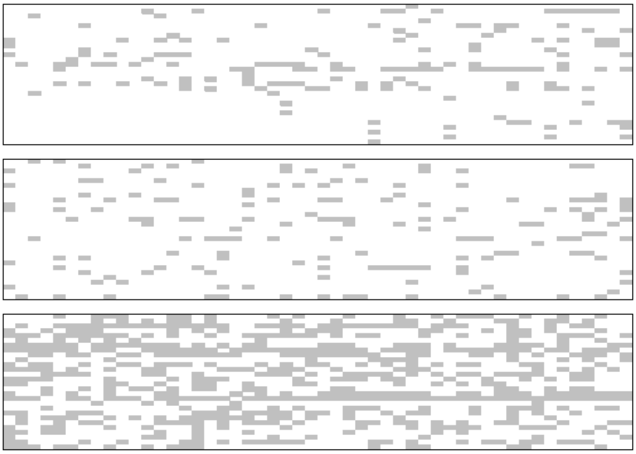


Fig. 3. A binary representation of all errors made on each interface: rotary (top), button (middle) and touch (bottom). The rows represent the thirty participants and columns are 50 numbers.

The most common errors on each interface vary and show that the errors must be dependent on the interface and therefore, independent of the user. This kind of analysis is important because we can note that some errors only occur on certain interfaces, or some occur on all interfaces at the same rate, such as ‘Anagram’, the reversal of two digits. This means that although it may not be possible to select an input device to reduce ‘Anagram’ errors, selecting the rotary dial will likely reduce ‘Repeat digit missing’ errors since they did not occur on the interface when tested in our study. By knowing what kinds of errors are common on each interface, it is possible to design devices that should work best for the objectives of that task.

One important observation from our study came from users’ interaction with the resistive touchscreen. Most touchscreen mobile phones and tablets on the market today use capacitive screens, which only require a tap of the fingertip to register the selection. Our resistive screen is more similar to the types of touchscreens used on other types of devices, such as ATMs, PIN pads and GPSs. The error rate in our study may have been so high on the touchscreen because users simply tapped the screen expecting the digit to register and did not watch for the visual feedback provided by the button on the screen to confirm the correct number. This would explain why the most common error was ‘Digit missing’, because users did not tap the screen hard enough, even if they thought they had entered the number correctly.

6 Conclusions and Future Work

In this paper we focused on analysing the number of errors that occurred during interaction with three number entry interfaces on a telephone platform. Telephone inputs are not outdated; on the contrary we use numeric keypads everyday at cashpoints. Although physical rotary dialers are nearly gone, they are available as smartphone apps.

From our study, it is evident that multiple trade-offs need to be considered when selecting an input device, with the type of task not being the only design criterion. There are a number of contexts where the most appropriate design for the user may be different than what is best for the task, so a compromise must be made. Instead of compromising to sacrifice usability for one group, considering the context could instead create a successful design. Many users have memorized a certain position of their fingers on a PIN pad to quickly enter their code, so if the layout was changed from the usual '1' at the top to having the '1' at the bottom, many users would enter their PIN numbers incorrectly in the context of paying at the till in a busy shop filled with distractions.

By considering learning and fatigue as factors in a device's use, an informed decision can be made about the most appropriate input device for the task. Since one device may have shown learning, one fatigue and one neither, we assume that the interfaces themselves caused the errors. This suggests that the interface choice is critical to the usability and performance of a device.

As future work we would like to repeat this study using other platforms of number entry, such as capacitive touchscreens, to find out if the touchscreen error rates are still so much higher than the other two interfaces. We would also like to compare these results with the errors caused by continuous input method of entering numbers such as knobs. This would lead to providing generalised recommendations for number entry interfaces. Devices could also be fine-tuned to offer more intuitive interaction for users. For example, incorporating real-world physics into an input device's response could assist users, such as a rotating wheel on a touchscreen that mimics the effects of friction when slowing down.

References

1. Oladimeji, P., Thimbleby, H., Cox, A.: Number Entry Interfaces and their Effects on Error Detection. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 178–185. Springer, Heidelberg (2011)
2. Oladimeji, P.: Towards Safer Number Entry in Interactive Medical Devices. In: Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2012), pp. 329–332 (2012)
3. Reason, J.: Human error. Cambridge University Press, Cambridge (1992)
4. Thimbleby, H.: Press On: Principles of Interaction Programming. The MIT Press, Cambridge (2007)
5. Webster, J., Eslambolchilar, P., Thimbleby, H.: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 596–597. ACM Press, New York (2012)
6. Wiseman, S., Cairns, P., Cox, A.: A Taxonomy of Number Entry Error. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction, pp. 287-196 (2011)
7. Zhang, J., Patel, V.L., Johnson, T.R., Shortliffe, E.H.: A Cognitive Taxonomy of Medical Errors. Journal of Biomedical Informatics 37, 193–204 (2004)

Informing the Design of an Authoring Tool for Developing Social Stories

Aurora Constantin[1,*], Helen Pain[1], and Annalu Waller[2]

[1] University of Edinburgh, School of Informatics, Crichton St 10, Edinburgh, EH8 9AB UK
A.Constantin-2@sms.ed.ac.uk, helen@inf.ed.ac.uk
[2] University of Dundee, School of Computing, Perth Road, Dundee, DD1 4HD UK
awaller@computing.dundee.uk

Abstract. This paper describes the design of an authoring tool, ISISS (Improving Social Interaction through Social Stories), for supporting practitioners using social stories to enhance social interaction in children with Autism Spectrum Conditions (ASC). The goals of the research are: 1) to determine practitioners' current procedures when working with social stories; 2) to discover how technology can better support such practitioners in the development and use of social stories that focus on improving children's social communication skills. An exploratory study was conducted with experienced practitioners, resulting in a number of design principles. Two low-functioning prototypes were developed, and explored in a second study. Further work is discussed.

Keywords: Paper-based Interfaces, Design, Autism, ASC, Educational Tool, Assistive Technology for Children with ASC, Authoring Tool.

1 Introduction

Children with Autistic Spectrum Conditions (ASC) frequently have difficulties with social interaction [19]. These can be addressed through the use of social stories. The concept of social story was devised by Carol Gray as "a short story—defined by specific characteristics—that describes a situation, concept, or social skill using a format that is meaningful for people with ASD" [7]. It is written from the student's point of view and is a guide to follow when they have difficulties with a social situation (see Fig. 1). They are used to help the child acquire appropriate behaviour, reduce inappropriate behaviour, teach routines, teach skills, or cope with transitions and novel situations. Meta-reviews of social story interventions [11, 16] focus on their effectiveness and on the impact of variables such as goal addressed, length of intervention, or participant characteristics. They conclude that social stories are promising, though it is not clear what variables are crucial to their efficacy.

[*] Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 546–553, 2013.

The aims of this research are: 1) to determine practitioners' procedures when working with social stories, and 2) to discover how technology can better support such practitioners. The longer term goal is to evaluate if computer-based technology can support and enhance the writing, presenting and assessment of social stories.

Taking my temperature
Sometimes if I am sick I get very hot.
The doctor needs to know how hot I am.
He will take my temperature with a thermometer.
He will put the thermometer in my ear. This will not hurt.
I might hear a beep or click noise, and this will not hurt.
The doctor can see if I am sick. Then he can help me get better.
The doctor will be happy and mum will be happy.

Fig. 1. Example of social story

Section 2 of this paper presents a short review of social stories and their importance for children with ASC. Section 3 describes a study which aimed to uncover practitioners' procedures and practices in order to see what support they need. The implications for design are considered in section 4. Section 5 assesses current tools for social stories in relation to the design principles and requirements elicited from the study. The design of an authoring tool to support practitioners in developing social stories is proposed. Section 6 presents the formative evaluation of the initial prototypes. Section 7 includes conclusions and future work.

2 Background

Psychological research has identified three dominant cognitive theories of autism: theory of mind, weak central coherence, and executive dysfunction [15]. Individuals with ASC frequently have impaired Theory of Mind: they have difficulty understanding what other people think, feel, or intend to do [2, 6, 13]. A social story can provide information about how other people think and behave in a specific social situation, using perspective sentences that refer to other people's thoughts, feelings, actions and motivation. In this way a social story might reduce or remove the confusion and ameliorate the deficit in Theory of Mind.

The weak central coherence theory describes the inability to build higher level meaning. An individual with ASC may be too focused on detail and miss the 'whole picture' [9, 15]. Social stories are helpful in guiding individuals with ASC to identify the relevant details for a specific event and to correct mistaken assumptions; providing logical connections can help the individual to grasp the bigger picture.

Executive Function is a generic term for functions such as: initiating, sustaining, shifting and inhibition actions [24]. It is thought to be responsible for handling novel situations where the routine is insufficient for optimal performance [15]. Social stories seem to be useful in reducing or removing the deficit in executive functions, by providing an individual with ASC with planning and organizing strategies in specific social situations, with initiation and impulse control.

Gray [8] suggests that social stories (SS) should be customized to meet the distinct needs and skills of the child, such as reading and comprehension skills, learning style, interests and attention. She provides criteria and guidelines to support the development of SS, including use of the first person; sharing social information in a simple, literal way; answering "wh" questions (who is doing, what, where, when and why?) and how questions, and use of positive language, with reference to positive rather than negative behaviour. Illustrations should be used when appropriate. She defines different sentence types to be used, including descriptive sentences (factual statements); perspective (describing a person's internal state, thoughts, feelings, beliefs); cooperative (identifying what others will do to help); directive (identifying a suggested response or a choice to a situation); affirmative (enhancing the meaning of previous statements), and control sentences (identifying strategies for recalling or applying information in SS). Descriptive, directive, perspective and control sentences may be either complete or partial (e.g. *"Mom and Dad will feel _____ if I finish all my dinner"*). She recommends a ratio between sentences of 0-1 directive and control to 2 or more descriptive, perspective, affirmative and cooperative.

3 Exploring How Practitioners Develop Social Stories

The approach taken in this research is primarily a Participatory Design approach, with practitioners as the main stakeholders. A study was conducted to better understand the following: 1) the process of developing SS; 2) the challenges encountered by practitioners in doing so; 3) to collect examples of SS developed, and typical content; 4) the tools currently used to support social story development. This study involved 4 practitioners who worked in special schools and had considerable experience in developing social story interventions for children with ASC. The practitioners were invited to participate in the study, and informed of its purpose. They were asked to think about their procedures and practices in SS, in advance of participation. The 2 hour study session took place outside of the school setting.

Each practitioner was asked to write a social story. Practitioners were invited to express their thoughts aloud ('think aloud' protocol) while building the social story. Following this, semi-structured interviews were used to better understand the practitioners' experience and challenges when using SS with children with ASC. Both activities were recorded and video transcripts analysed qualitatively, using open, axial and theoretical coding [17] to build a conceptual framework of social story intervention. Further literature was reviewed and incorporated into the framework. The results suggested that practitioners would value support for developing SS using a specialised computer tool, as an alternative to pen-and-paper or word processing methods currently used. The outcomes led to design principles and initial requirements for a 'social stories authoring tool', described below.

4 Supporting Practitioners in Developing Social Stories

The study revealed that a tool for supporting practitioners in developing SS might help them organise their work, and address the main challenges they face, including re-using resources developed based on individuals' profiles. Design should also consider Gray's guidelines and best practice. The steps that practitioners use are identified and elaborated, along with other core concepts, in Fig. 2: the exterior arrows suggest that challenges and content are determined by steps and goals, while the interior arrows suggest that the four concepts are part of social story intervention. These led to functional requirements including being able to create new stories; extend and edit existing ones; browse and search a library of stories and related images; annotate story sentences; present them in various forms; create and edit child profiles; manage social story development through the use of status display, reminders, etc. and to record progress through and manage assessment of SS.

Fig. 2. The conceptual framework of social story intervention

A number of challenges were identified, and these are incorporated in design principles to guide the design of tools to support social story development.

Ease the Practitioners' Workload. A major challenge that practitioners encounter is the time spent in preparing educational materials. Although SS seem to be less demanding than other educational strategies, the whole process of preparing, writing, presenting and assessing a social story is labour intensive. To address these problems, any support tool needs to be simple, intuitive and must help practitioners organise their work, and support the steps in the development process. The practitioners identified that the tool should allow them to reuse SS, symbols and photographs and to monitor the impact of SS on children. Data about the child's progress should be accessed and presented in various ways, enabling practitioners to get new insights into the child's behaviour and assess the success of the social story.

Design for Customizability. A common desire is to quickly customize newly created or re-used SS. Users should be allowed to create resources for each child, to add their own SS, symbols, photographs, rewards, as well as interests and strengths (which should be accessible from the child's profile). The stories should also be customizable to the story topic content, be re-usable and sharable with other practitioners and children. Text to speech capabilities should be added to accommodate children with reading difficulties. Options to choose between various layouts should be provided.

Design for Engagement. This could be met by customization to the child's interests and familiar context (e.g. images of familiar people). Practitioners could add rewards at the end of SS (e.g. animated characters, songs) adapted to each particular child. Social stories with partial sentences check comprehension, but may also make the system more interactive, potentially improving the child's engagement with the tool.

5 Current Tools for Social Stories

Applications for Creating Social Stories. e.g. Story Builder [21], Story2Learn [22], Sandbox Learning [18], Pictello [14], Social Stories [20], and StoryMaker [23], were considered in relation to the need to support the steps proposed for story development, and the design principles and requirements, as above. These applications focus primarily on building, editing and presenting SS, but do not generally support checking the comprehension, monitoring the story, or assessing SS. None of the applications provides an option to annotate the type of sentences, as Gray defines them. The applications are presented as allowing customisation, but this consists largely of changing font sizes and colours, background and the type of voice (when using text to speech technology, e.g. Pictello). None permits change of the story layout, nor provide the option to create resources for a particular child (e.g. favourite pictures, symbols or rewards) or for their reuse when creating new stories for that child. None permit the creation of a profile of the child, or storing information about the SS created, or their progress.

The limitations identified means these applications are not viable as the basis for a computer-based application to support practitioners in developing SS. Anecdotal evidence shows that practitioners sometimes use generic tools to create SS: Communicate: In Print [5], Boardmaker [3], or Comic Life [4], though these do not satisfy the requirements. This research suggests that, to better support practitioners, research is needed to investigate whether a computer based tool could be developed that satisfies more of the requirements and principles identified through studies with practitioners.

Authoring Tools. enable users to author material using an intuitive interface. In his analysis of the state of art of authoring intelligent tutoring systems (ITS), Murray [12] argues that authoring tools: a) reduce the effort used in creating ITS; b) reduce the necessary skill threshold for developing ITS; c) support the author in articulating or organizing her work; d) scaffold good design principles; e) facilitate rapid prototyping. Many of these arguments agree with the principles and requirements defined above, and support the use of an authoring tool to help practitioners in their work, enabling rapid customization, flexibility and requiring no programming skills.

6 Formative Evaluation of Initial Prototypes

A second study was conducted with practitioners. The goals of this study were: 1) to explore design alternatives; 2) to discover usability problems and solutions to overcome them; 3) to generate new ideas to include into the systems' features and interaction, and 4) to refine the system specification. Two versions of low-functioning prototypes were created using Balsamiq Mockups [1], according to the requirements and design principles (see screenshots examples in Fig 3: a & b). They were explored with ten practitioners having experience in using SS with children with ASC. In each two-hour session constructive interaction was applied, a version of the thinking aloud protocol which requires two practitioners at a time [10]. The practitioners were presented with four scenarios and invited to follow each scenario using each prototype, by pressing on buttons and menus, and simulating typing. A researcher playing the "computer" role changed the screens according to the users' actions. The main researcher, acting as a mediator, encouraged the practitioners to express their opinions and preferences for various features of the prototypes. Brainstorming was then used to find solutions for the problems encountered and to generate new ideas. Qualitative analysis of session recordings was used to determine practitioners' preferences, usability problems and solutions, and to extract new ideas for design.

The problems found were categorized into five groups: 1) page layout; 2) navigation/workflow; 3) concepts and terminology; 4) content, and 5) functionality. Based on these, the system interfaces and specification were refined.

Figure 3 shows the revised working prototype. The tab navigator containing pictures, symbols, layouts and resources (for a child) was preferred on the left side (Fig.3:d). The navigation buttons were placed in left and right bottom corners of the story panel. The option to annotate the sentences was considered of high importance for practitioners and included in the working prototype (Fig.3:e).

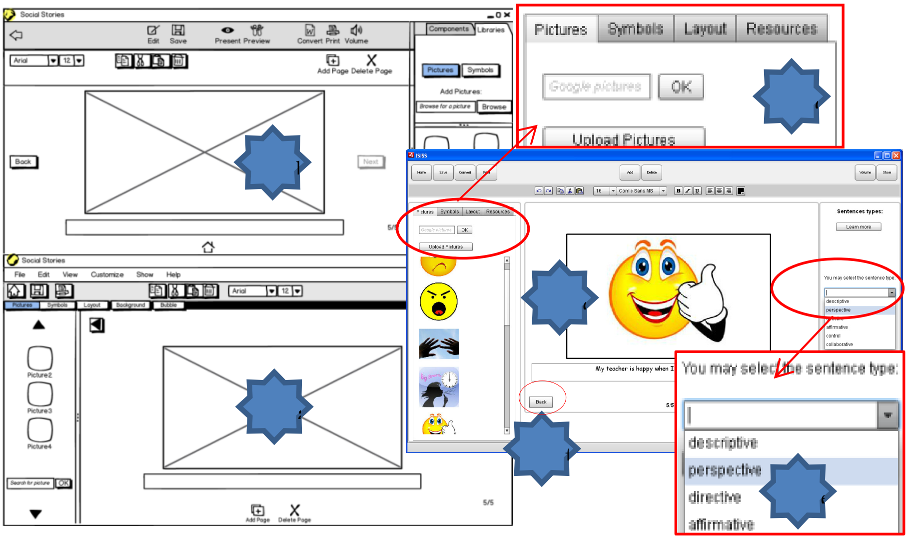


Fig. 3. The Prototypes: a) initial prototype 1; b) initial prototype 2; c) working prototype; d) tab navigator for customization; e) annotate sentences; f) navigation button

7 Discussion and Future Work

This paper reports on research that informs the design of an authoring tool for writing, presenting and assessing SS for children with ASC. This research has been guided by the methods of Participatory Design. The first study uncovered practitioners' procedures when developing SS. The focus was on the core concepts and themes that could be translated into design principles and system requirements. Although practitioners were aware that tools for social story interventions exist, most of them preferred paper and pencil to write SS, existing tools being either too complex to use or not flexible enough. They sometimes use tools not specific for SS, in order to access the library of symbols (in [5] and [3]) and because of the simplicity of the interface (in [4]). Through the second study the requirements were clarified, different design strategies explored, and interfaces and specifications refined. In both studies the practitioners contributed in creating knowledge and showed interest and enthusiasm for the project.

These are the first steps in a project aimed at better scaffolding for practitioners in developing SS. In turn this may increase the positive impact of SS on social communication skills of children with ASC. The results have been incorporated in an evolutionary prototype authoring tool, which will be further evaluated with practitioners.

References

1. Balsamiq Mockups, `http://www.balsamiq.com/`
2. Baron-Cohen, S., Leslie, A.M., Frith, U.: Does the autistic-child have a theory of mind? Cognition 21(1), 37–46 (1985)

3. Boardmaker, http://www.mayer-johnson.co.uk/what-is-boardmaker/
4. Comic Life, http://www.comiclife.com/
5. Communicate: In Print, http://www.widgit.com/products/inprint/index.htm
6. Frith, U.: Autism: Explaining the enigma. Blackwell, Oxford (1989)
7. Gray, C.: Social Stories (2003), http://www.thegraycenter.org/
8. Gray, C.: Social Stories 10.0. Jenison Autism Journal 15(4), 2–21 (2004)
9. Happé, F., Frith, U.: The weak coherence account: detail-focused cognitive style in autism spectrum disorders. Journal of Autism and Developmental Disorders 36(1), 5–25 (2006)
10. Holzinger, A.: Usability engineering for software developers. Communications of the ACM 48(1), 71–74 (2005)
11. Kokina, A., Kern, L.: Social Story interventions for students with autism spectrum disorders: A meta-analysis. Journal of Autism and Developmental Disorders 40(7), 812–826 (2010)
12. Murray, T.: Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. International Journal of Artificial Intelligence in Education 10, 98–129 (1999)
13. Perner, J., Lang, B., Kloo, D.: Theory of mind and self-control: more than a common problem of inhibition. Child Development 73(3), 752–767 (2002)
14. Pictello, http://www.assistiveware.com/product/pictello
15. Rajendran, G., Mitchell, P.: Cognitive theories of autism. Developmental Review 27(2), 224–260 (2007)
16. Reynhout, G., Carter, M.: Social Stories™ for children with disabilities. Journal of Autism and Developmental Disorders 36(4), 445–469 (2006)
17. Saldana, J.: The Coding Manual for Qualitative Researchers. Sage Publications, London (2009)
18. Sandbox Learning, http://www.sandbox-learning.com
19. Shattuck, P.T., Seltzer, M.M., Greenberg, J.S., Orsmond, G.I., Bolt, D., et al.: Change in autism symptoms and maladaptive behaviours in adolescents and adults with an autism spectrum disorder. J. Autism Dev. Disord. 37, 1735–1747 (2007)
20. Social Stories, http://a4cwsn.com/2012/09/social-stories/
21. Story Builder, http://www.leedsmet.ac.uk/inn/usabilityservices/download2.htm
22. Story2Learn, https://itunes.apple.com/gb/app/stories2learn/id348576875?mt=8
23. StoryMaker, http://udm4.com/iPhone/StoryMaker_for_Socia-3170588
24. Zelazo, P.D., Müller, U.: Executive function in typical and atypical development. In: Goswami, U. (ed.) Handbook of Childhood Cognitive Development. Blackwell, Oxford (2002)

Introducing New Perspectives in the Use of Social Technologies in Learning: Social Constructionism

Antigoni Parmaxi[1], Panayiotis Zaphiris[1], Eleni Michailidou[1], Salomi Papadima-Sophocleous[2], and Andri Ioannou[1]

[1] Department of Multimedia and Graphic Arts, Cyprus University of Technology, Lemesos, Cyprus
[2] Language Centre, Cyprus University of Technology, Lemesos, Cyprus
{antigoni.parmaxi,panayiotis.zaphiris,eleni.michailidou, salomi.sophocleous,andri.i.ioannou}@cut.ac.cy

Abstract. This paper reports on a qualitative study of the use of social technologies, explored in the context of an intensive 650-hour Greek language course. Qualitative content analysis of instructors' field notes, students' and instructors' reflections, interviews and a focus group was employed aiming at identifying the use of social technologies as a platform for constructing an online artifact. To triangulate the findings, the study also collected data by observing students' activity with social technologies. A code scheme was developed which manifests the use of social technologies as a social constructionism platform identifying its major dimensions: exploration of ideas, construction of online artifact and evaluation of the constructed artifact. Actions within each dimension that indicate the manifestation of social constructionism are identified and discussed. This study revealed results in favor of the use of social technologies as social constructing platforms suggesting a new framework for their use.

Keywords: social technologies, web 2.0 technologies, social constructionism, online artifact.

1 Introduction

The emergence of social technologies transformed the way we communicate, learn and interact with others. Among other tools, social or Web 2.0 technologies received substantial consideration from instructional designers, researchers and practitioners. Each stakeholder explores these technologies from different angles aiming at describing and explicating how these technologies are used, by whom and for what purpose.

The potentials of these technologies have also expanded the possibilities of teaching and learning and several projects have evolved which exploit blogs and wikis as social writing platforms. As [2] points out, "social writing platforms appear to be logistically useful tools for a variety of campus needs, from student group learning to faculty department work to staff collaborations". However, the potentials of these technologies are not limited to this framework. This study aspires to widen the applicability of these technologies drawing on the theoretical framework of

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 554–570, 2013.

constructionism [21-23]. The essence of constructionism can be summarized in its conviction that individual learning occurs more effectively when learners understand the world around them by making tangible objects. Thus, the main aim of this study is to explore the potential that social technologies offer in facilitating teams of learners in order to socially construct an online artifact.

2 Objectives and Related Work

2.1 Objectives

Social technologies have been widely researched, however little work has been done on their use as social constructing platforms. In this study, we exploit this possibility by providing insight to the use of social technologies in a longitudinal Greek as a second language (L2) course. The aim of this study is broken down in the following objectives:

1. Explore the potential that social technologies offer in facilitating teams of learners in order to socially construct an online artifact
2. Develop a code scheme that captures the core dimensions of social technologies as social constructing platforms
3. Explore the role(s) adopted by students and instructor within a social constructing environment

In the sections that follow an overview of the relevant literature is provided related to Computer Assisted Language Learning, Social Technologies in CALL and Papert's theoretical framework.

2.2 Computer Assisted Language Learning (CALL)

Gamper and Knapp [7] define Computer Assisted Language Learning as "a research field which explores the use of computational methods and techniques as well as new media for language learning and teaching". The popularity of the field of CALL has increased incrementally, especially with the advent of recent technological developments.

The question of CALL effectiveness is brought forward repeatedly in CALL research [1], [9], [12], [34-35], [37], [39]. However, this is not a simple yes or no question as often other parameters need to be considered. Zhao [39] in his literature review and meta-analysis for assessing the potential of technology for improving language education, showed that the use of technology in language learning is at least as effective as teacher-delivered instruction. However, he points out that these findings should be interpreted with caution due to the limited number of studies that provided satisfactory data for his meta-analysis; journals' tendency to publish studies with positive gains; the fairly small samples and rare employment of random sampling in the studies and the fact that all of them were carried out with college students and adult learners. Moreover, in most cases teachers were implementing the technology use, thus adventuring the possibility of the "Pygmalion effect". Zhao [39] also exploits four issues that need to be addressed in order for information and communication technologies to significantly improve language learning. First, curricula and course content needs to be developed with an eye to include a wide range of available technologies that drive pedagogical solutions. Second, further research needs to be conducted in order to exploit effective ways of technology use. Third, since technology use and research has not expanded to K-12 classrooms, technology needs to be promoted and research needs to be encouraged in K-12 language classrooms. Lastly, there is a need for large scale systematic empirical assessment of technology uses to support language learning. Taking these into consideration, the question of CALL effectiveness, needs to be refined to examine how technology is being used, by whom, in what context and for what purpose [12]. As [39] acknowledges it is not the technology per se that can be classified as effective or ineffective, but the specific way in which the technology is used.

2.3 Social Technologies in CALL

Unlike the traditional Web 1.0, Web 2.0 changed the direction in communication on the internet [11]. With the reception of texts, sounds and images being the dominant activity, the readable Web 1.0 was followed by Web 2.0, the 'writable web', where the dominant activity is the creation of new content. Several studies indicate that these technologies provide fertile ground for collaboration, active participation, creative thinking, connectivity, and sharing of information and ideas among users, engaging learners in authentic, real-world situations and interactive communication [8], [10], [17], [24], [30]. Social collaboration in tagging and annotating documentation for navigation support is also highlighted by several studies, emphasizing the importance of social collaboration and its added value in learning environments [40-41].

Social software came into view as a major element of the Web 2.0 movement [2]. Prevalent software of this movement are blogs, wikis, podcasting, videoblogs, Myspace, Twitter and Facebook. These types of software differ significantly from static web-pages in the sense that they are open to the world and editable by everyone. Research conducted exploring the use of this software in the field of CALL focused on its effectiveness in supporting language learning, often in comparison with traditional instruction. For example, [3] studied the effects of blog-centered writing instruction vis-à-vis in class instruction on students' writing performance. The results indicated that learners who used blog software in their writing courses performed better than those who received only in class instruction in specific areas of writing, such as content and organization. In another study, [28] explored the development of meaningful interactions on a blog used by English as a Foreign Language (EFL) Master's students in France. The aim of this study was to measure the potential added value of a blog for the purpose of language learning, and more specifically for the development of learners' written expression. Findings demonstrated positive results in respect of participation and development of meaningful interactions, although the blog is not perceived as a "real life" one but a pedagogical one. Lee [14] provided a new insight into the practicality of implementing Web 2.0 tools in L2 teaching and learning. More specifically, this study explored the effectiveness of combined social networking applications, such as blogs and podcasts, for an intercultural communication and awareness between Spanish and American university students. The results of this study yielded that the use of blogs and podcasts offered promising benefits to both American and Spanish students who afforded unique opportunities to explore the target language and culture using digital technologies. On the same line, [15] explored the effectiveness of using a wiki for collaborative writing. The results indicated that wikis had a positive impact on students' writing through collaborative engagement. In another study, [16] reported on a study using blogs as out of class assignments for developing learners' language competence. The results showed that regular blogging impacts positively on learners' writing fluency and increases their motivation for writing for a broad audience. Moreover, peer feedback on the content incites further discussion, whereas feedback from the instructor on linguistic elements promotes focus on form for language accuracy. The study concluded by indicating two essential elements for the implementation of blog projects in L2 instruction, namely learners' critical thinking and technological skills. In another study, [6] also evaluated the usefulness of blogs for peer feedback on L2 writing. This study's findings demonstrated that blogs can be a valuable tool for peer feedback, however issues of students' and tutors' training as well as the aptness of tool against other learning technologies needs to be taken into consideration.

Furthermore, researchers explored how the features of Web 2.0 tools could be exploited so as to improve many aspects of traditional teaching. For example, a study that reports the experience of blog integration into an advanced Italian course by [18] showed that blogs not only offer a useful tool for practicing reading and writing skills, but can also promote learners' interaction and raise a sense of class community. For this to happen, however, careful attention to two key aspects is required: the way in which the use of the blog is integrated into the course content and structure, and the teachers' role in moderating and facilitating blog interaction.

Learners' interaction in a Web 2.0 environment is also under the microscope of research. Bradley, Lindström and Rystedt [5], for example, explored what interaction is developed in the wiki and how written interaction promotes language learning. Results revealed that there are different types of posted interaction among group members on the wiki. Students co-operate, namely they post individually on a common theme, but they also collaborate, they produce joint texts and then make alterations and additions. On the same line, [13] focused on collaborative construction of meaning among 40 Non Native Speakers (NNS) pre-service EFL teachers in a long-term wiki-based activity. Their results showed that students benefit from opportunities to practice autonomy in flexible learning environments. Their research also shed light on students' collaborative autonomous language learning abilities, and on the nature of students' individual and group behavior when attending to meaning. In another study, [33] turned to the use of voice blogs and explored them as a platform for an extensive study of language learners' speaking skills. The study demonstrated that a series of blogging stages were adopted by learners, including conceptualizing, brainstorming, articulation, monitoring and evaluation, as well as a series of strategies to cope with difficulties related to blogging. Additionally, learners did not only perceive blogs as a learning platform, but also as a means of self-presentation, information exchange and social networking. Finally, this study suggested blogging as a dynamic forum that enhances extensive practice, learning motivation, authorship and development of learning strategies [33].

Turning to the challenges faced in the use of wiki in the language classroom, [20] pointed out that the integration of wiki in the learning and teaching process does not necessarily guarantee learning outcomes, but the key to success are the well-designed activities blended into the curriculum. On the same line, [19] suggest further research to be conducted to identify the impact of ICT towards teaching and learning outcomes.

2.4 Constructionism

The present study draws on the theoretical framework of constructionism as it is developed by Seymour Papert [21-23]. Constructionism can be summarized in its evoking idea of learning-by-making [22]. Papert [22] unwraps his theory by highlighting its difference and similarity from Piaget's constructivism:

Constructionism--the N word as opposed to the V word--shares constructivism's connotation of learning as "building knowledge structures" irrespective of the circumstances of the learning. It then adds the idea that this happens especially felicitously in a context where the learner is consciously engaged in constructing a public entity, whether it's a sand castle on the beach or a theory of the universe [22]

Papert's theory can be summarized in his vision of a new educational environment in which learners build meaningful knowledge artifacts by taking advantage of the ubiquity of new technologies around them. The constructed artifact can be exposed, discussed, explored and admired, it is "'in the world' –a sand castle or a cake, a Lego house or a corporation, a computer program, a poem, or a theory of the universe" [23]. Papert's [23] constructionism is based on the assumption that knowledge is gained when students find this knowledge for themselves and formal or informal education can promote knowledge attainment by providing moral, psychological, material and intellectual support. In addition to having students finding knowledge by themselves, Papert supports that computers are needed, as an environment through which rich activities can be developed, so called "microworlds".

Papert [23] sets as a main principle of constructionism the necessity for "'objects to think with', objects in which there is an intersection of cultural presence, embedded knowledge and the possibility for personal identification" [23]. For Papert, a constructed computational object-to-think is a computer-controlled cybernetic animal (Turtle) within the LOGO computer language. Within the LOGO environment, the Turtle is an abstract object that can be made to move by giving commands. However, the role of Turtle is not to replace thinking but to enable learners to think with. In the environment of LOGO, programming is introduced when children experience programming by having the Turtle programmed to act in response to new commands [23].

The linkage between Distributed Constructionism and CALL has been established by [38] who explored the implementation of Distributed Constructionism through a Participatory Design methodology for an Online Learning Community. Throughout this study, the learners collaborated in developing the content of an online Modern Greek language course, peer reviewed and published content contributions, and participated in participatory design teams. In this study the Participatory Design was implemented as a four step process, namely: (a) build bridges with the intended users; (b) define user needs and recommendations to the system; (c) develop a prototype and (d) incorporate feedback and carry on the iteration. Additionally, Distributed Constructionism was employed to enhance the learning experience and community development. The findings revealed that Distributed Constructionism enhanced the learning experience of both the passive users and the Participatory Design team, whose contributions included replying to other students' language enquiries, helping out students to cope with technical problems and helping them explore resources to enhance their learning of the Greek language.

In another study, [29] provoke the need for a fundamental change in approaching teaching and learning. Thus, they attempt to contribute to the ongoing debate by presenting some key principles of constructivism as a new learner-centered paradigm for learning. Moreover, they consider Papert's constructionism as a basis for putting theory into practice for language learning, as a formula that could serve in the future "as the guiding principles for curriculum design, materials development, and classroom practice" [29]. Taking this as a stepping stone, the current study explores the applicability of Papert's theoretical framework in the use of social technologies in CALL.

3 Setting

This paper presents the results of a longitudinal inquiry of social technologies as social constructing platforms, in learning and teaching Greek as an L2. Social technologies were integrated in a one-year foundation course at a Greek speaking public university in the Republic of Cyprus. The course lasted for the academic year throughout September 2011 till May 2012. The class met face-to-face every day for five hours, in a total of 650 hours. In-class activities were held face-to-face and online, as well as out-door activities in order for students to practice the language in authentic, real-world situations. The course was particularly designed to meet the needs of university students who planned to study nursing. In the first semester, the language and content were drawn from students' experiences and other key learning areas such as nursing. In the second semester, the language and content were drawn exclusively from nursing. The course and the materials were tailored to meet the academic and professional needs of the nursing students.

All material related to the study was collected during an intensive 650-hour Greek language course at a newly established public university in Cyprus. The university accommodated approximately 2500 undergraduate and postgraduate students. The official language of the university is Greek. At the time of the research, 94% of the students attending the university were of Greek Cypriot ethnicity, 4.45% Greek, 0.3% Kenyan, 0.25% Ugandan, and the remaining 1% were made up of numerous other ethnicities, including British, Russian, Albanian, Bulgarian, Serbian, Iranian and Rumanian, according to university records.

3.1 Students

The participants in the intensive course were four male students from Kenya and Uganda, who came to Cyprus, for five years, on full scholarships. Students enrolled in the Greek course upon their arrival in Cyprus, had sessions every day for five hours. Despite the small sample, this study aimed to explore students' use of social technologies in depth. This study's horizon is to go in detail and in depth, despite the small sample, having participants work with social technologies in a long-term course, and collect data rich in detail about the use of social technologies.

The students' age ranged from 19-23 years. All students were fluent English speakers; none of them had any knowledge of Greek upon arrival in Cyprus. Their computer skills were in general at basic to intermediate level. Three of them were able to turn the computer on and off; all of them had difficulties in advanced functions such as sending emails and attachments; document processing and use of keyboard. Additionally, they had minimal knowledge of social technologies, two of them created a Facebook account upon arrival to Cyprus, and none of them had any previous knowledge of blogs, wikis, Google documents or Dropbox.

3.2 Instructor

The instructor was a female, with four years experience in teaching Greek as an L2. The instructor was both participant and observer and her role provided access to as a wide-range of data as possible.

3.3 Social Technologies

Participants utilized five social technologies throughout the course: wikis, blogs, Facebook, Google documents, and Dropbox. The use of all technologies is explained in the following sections. The instructor set up two class wikis, Greek4Practice wiki and Lexicon wiki. Wikispaces was employed for creating the course wiki because of its simple, user-friendly interface that allows page layout to be easily changed. It is free and password-protected, easy to create and update. It uses open editing functionality and lets users create unlimited internal wiki pages and links. Users can also add other multimedia features including images, audio and video files to support the content. Wikispaces is currently available in many languages, including Greek, which enabled students to develop their site in the target language. Basic functions within the wiki include file or picture uploading, editing, creation of links and view of the history of pages. Wikispaces also allows its users to monitor the activity of the wiki and compare the differences between any two versions of the page.

The instructor created a Facebook group in which all participants were invited to join. Only members of the group were able to see the group information and content. Students were allowed to freely post anything of their interest on the Facebook group and make comments using the target language.

Following [4], the instructor set up one blog for the course, as it is more likely for classmates to interact with each other in one space. The blog allowed students to post and comment, upload material ant track the history of blog entries. For the instructor interface, the class blog tracked all posts and comments history.

Google documents were developed for sharing material related to the course. The instructor created and shared a folder of Google documents with students, who were allowed to view and edit. Google documents allows users to share, open and edit the document simultaneously. The Google service also enables users to view the revision history, additions made to a document, with each author distinguished by color.

Finally, all participants shared a Dropbox folder which included photos taken throughout the outdoor activities held. Dropbox enables all member of a shard folder to edit and re-post files. The version history is kept for 30 days.

4 Methodology

The linkage between constructionism, social technologies and CALL and thereafter the generation of theory will emerge from the data collected throughout the intensive course of Greek described in the section above. This study's horizon is to go in detail and in depth, despite the small sample. Throughout the intensive Greek course, the students are involved with social construction of artifacts within social technologies, including wiki, Facebook, Blogger, Google Documents and Dropbox.

4.1 Data Collection

The data was collected using a variety of methods: a questionnaire, in class observations and daily field notes kept throughout the course by the researcher-instructor, instructors' and learners' weekly reflective diary kept on the wiki. Interviews were also conducted which allowed us to elicit qualitative data about the process that participants followed within social technologies. 16 interviews (4 per student) were conducted aiming at capturing students' overall impression and challenges of their learning process. A protocol was followed to explore students' opinions on overall experiences throughout the course. Each interview lasted approximately 45 minutes to one hour. The interviews were tape recorded and transcribed verbatim. Finally, students participated in a focus group which lasted approximately 30 minutes, and written notes were taken during the focus group. Table 1 briefly describes the types of data collected. To triangulate the findings, the study also collected data by observing students' activity within social technologies.

Table 1. Overview of Data Collected

Data	*Purpose*
Questionnaire	Insight into students language and computer literacy
Students' Reflections	Self evaluation of their activities outcomes and process adopted
Instructors' reflections	Reflection of activities outcomes
Instructors' field notes	Overview of the process adopted and activities held
Interviews	Reflection on activity process and outcomes
Focus group minutes	Overview of process adopted by the group

4.2 Development of Code Scheme

In order to become acquainted with the data, we first read all the data set thoroughly. This enabled us to acquaint a holistic view of the course development during analysis and take its context into account. Also, reading the course outline and profiles of the participants helped us to gather peripheral information about the course. Throughout this process, insights and ideas emerging from the data have been reported as memos within the Qualitative Research Software Nvivo. The purpose of this stage of analysis is “to ensure that the theoretical ideas which have emerged in the first round of coding can be systematically evidenced in the data, thus addressing the validity of the research results” [36].

In our code scheme, consecutive sentences that construct the same meaning are taken as one text unit and coded into a single code. This ensures that each coded segment captures the essence of described events in detail and it is still seen within its context [26]. The aim of this process is to classify and elucidate telling the story of the data [25]. A shortcoming of this approach is that the decision of what constitutes a meaning can be very subjective. To address this issue, we followed [26] approach, which developed a procedure as a guide for determining the unit of analysis. An inter-coder reliability test with a sample of the data set revealed that two independent coders agreed on the segmentation in 81% of the cases.

In the second step, we analyzed the data set within the Qualitative Research Software Nvivo, extracting key words and themes observed. When we had a collection of themes and patterns that described the data, we sorted and grouped the codes and used them to develop the code scheme. Data was coded based on the target of an activity, for example when participants mentioned that they collected material from real situations in order to build their artifact within social technologies, we coded the segment under social technologies.

In the third step, we examined the code scheme by sensitizing concepts from Papert’s theoretical framework [25]. This procedure was repeated iteratively, until a final code scheme was developed. Saturation was reached, when no new codes could be found and the data set could be sorted into the existing codes without any discrepancies. To make the code scheme as objective as possible, a codebook was developed, which clarified the description of the codes further. This codebook includes characteristics that distinguish the codes from each other and facilitates analysis process. To measure the inter-coder reliability, the codebook was given to another independent researcher who coded 10% of the data set. Cohen’s KAPPA was calculated to be 0.72 which according to [32] is considered to be substantial.

5 Results/Discussion

Over the two semester course (26 weeks), the four participants and the instructor made a total of 1096 edits on the first wiki and 2086 edits on the second wiki. On average, each participant made 219 edits on the first wiki and 417 on the second wiki. On the Facebook Group the four participants and the instructor made 301 posts and 495 comments. On the blog a total of 26 posts and 40 comments were made and 1158 files were uploaded by the four participants and the instructor in the shared Dropbox folder.

5.1 Social Technologies as Social Construction Platforms

In this section the code scheme that manifests the use of social technologies as social construction platforms or as “objects-to-think with” is presented. Overall, three categories emerged: exploration of ideas, construction of artifact and evaluation of constructed artifact.

Exploration of ideas
Orientation: Text units which refer to setting up the goals of an activity, providing objectives for a specific task (often the instructor challenges the students to identify why a specific activity takes place and how it should be formed).

Brainstorming: Text units which refer to making a list of ideas or content that could be used in the constructed artifact. Text units also refer to sharing notes and ideas within social network channels.

Material exploration: Text units which refer to exploration and collection of material by taking photos from real situations that learners experienced and by searching the web. The issue of cultural information exchange is prominent here since students often conducted activities out of class in order to collect material.

Construction of artifact
Outlining: Text units which refer to translating material from English to Greek, mapping the main and supporting ideas (before moving to putting the ideas down).
Editing material: Text units that refer to editing material, during the construction of the artifact. Editing material includes adding links and other multimedia material. Editing the material is rather a social than an individual task.

Evaluation of artifact
Revising: Text units which refer to the process in which the participant corrects production-errors. Spell check and automatic correctors are used. Revising is rather an iterative than an instantaneous process.

Peer reviewing: Text units which refer to peer reviewing the artifact in terms of organization, content and language usage. Comments were also employed for providing feedback within social technologies as a method for monitoring and evaluating a certain activity.

Instructor reviewing: Text units which refer to the instructor reviewing the constructed artifact in terms of organization, content and language usage.

Presenting/Publishing: Text units which refer to students presenting the constructed artifact to their classmates. Publication of the constructed artifact was done also via social communication channels (Facebook).

In the following sections we report the three dimensions that manifest social constructionism. We structure our discussion around the aforementioned dimensions, along with the actions occurring within each dimension that indicate the manifestation of constructionism within social technologies.

5.2 Exploration of Ideas

The first stage involved orientation towards the tool and the idea. In this stage, goals and objectives are set and the instructor often challenges the students to take preliminary decisions for exploring their and other's ideas. In the case of Papert's constructionism within the LOGO environment learners contact free with the learning environment; however in the case of applying constructionism within social technologies the exploration of ideas facilitated by the instructor is a vital step in the process. During this phase, the instructor introduces the tool to the learners through tutorials and step by step workshops. Students that participated in the course had difficulties in coping with the tools; however, the use of computer enabled them to enhance their language literacy:

> *Fred 4th interview: When we started to work on wikispace writing, logging in and all those stuff I realized that it needs more practice because it is not easy as such. We need to go after links inside the wikispace and the good thing that I like in wikispace is that I realize that it helps a lot mostly when you write something in Greek.*

Students need to explore a great deal before gaining mastery of how the technology works. However, the task is engaging and carries students through the learning process:

> *Nelson reflections: I really enjoyed this week, I learned how to put the voki in the lexicon, and I also learned a lot of things by adding new verbs in the link related to Nursing.*

A major theme in constructionism is that the computer is seen as a "carrier of cultural 'germs' or 'seeds' whose intellectual products will not need technological support once they take root in an actively growing mind" [23]. Taking this a step forward, social constructionism assumes that learners can socially exchange "germs" or "seeds", throughout the brainstorming phase. In the framework of social constructionism, learners interact and exchange material throughout social communication channels. Facebook has been used as a tool for listing resources related to a specific topic. As it is shown in figure 1, an orientation task has been set by the instructor on Facebook requesting from students to search for material related to a specific topic. Participants have been listing related material by posting comments with material related to the specific topic.

The last action of this phase includes exploration of the material gathered, as well as out of class activities for photographing and collecting material in tasks that participants decided to include material from real situations.

Fig. 1. Screenshot of social brainstorming in Facebook group

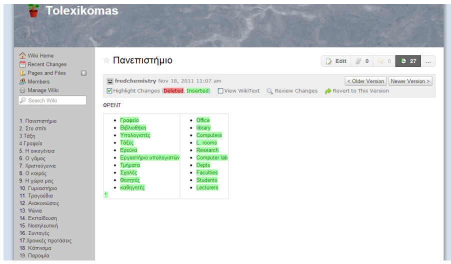
5.3 Construction of Artifact

Learners begin their construction experience by translating material from English to Greek and mapping the main and supporting ideas:

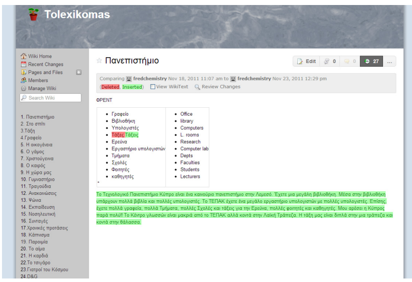
> *Instructor's reflections: Students tried to read the material they found sentence by sentence and with the use of the translator to understand the basic information before start working on their own material.*

Having the material understood, students move to more advanced commands. Students worked together in building an online dictionary which "can be shown, discussed examined, probed and admired" [23]. To this aim, during the exploration phase students focused in finding the topics that would be included and moved on in the construction phase, first by simple text editing and then by executing more complex actions such as picture uploading, adding plug-ins, videos, or other multimedia material (see fig. 2). Throughout this phase, learners are challenged to go through the artifact and enrich their computational and linguistic competences. The examples vary, but in each case learners practice in the use of language in a rich "object-to-think".

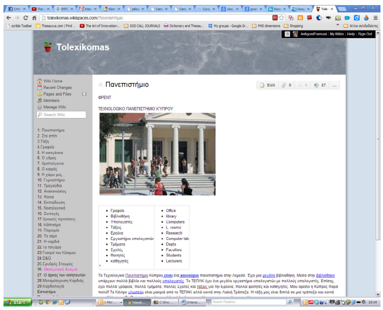
Screenshot 1: Lexicon page November 17, 2012



Screenshot 2: Lexicon page November 18, 2012



Screenshot 3: Lexicon page November 23, 2012



Screenshot 4: Lexicon page November 23, 2012

Fig. 2. Screenshots of the process of constructing the online dictionary

Participants viewed this process of constructing the artifact as highly iterative and powerful since learners had to involve systematically in problem solving and explaining the constructed artifact to their potential audience:

Saul 3rd interview: This procedure helps me because when you stick on doing something maybe you learn more. I have not been knowing how heart transplantation is called in Greek but I think I will not forget it, because when you look a certain word maybe once you can forget it easily but this one I will not forget it. I have just got it right now and many other words I have been working on.

5.4 Evaluation of Artifact

A central theme in constructionism is that "people seldom get anything exactly right on the first try" [23]. Within this framework, the construction of an artifact is seen as an iterative process that includes several modifications and revisions. The actions that participants followed in this stage include presentation of the constructed artifact to their peers either face to face or by publishing an artifact into social network channels. Participants during the presentation of their artifact receive feedback from their peers:

> *Saul reflections: In class I tried to do my best in presenting my work on the wiki and from the mistakes that I made I learned the correct.*

Additionally, peers' and instructor's comments enable them to identify and correct their mistakes:

> *Siraj 4th interview: In the blog we were discussing and exchanging views, so through those discussions we could see the mistakes of one another, written mistakes by reading through one another's posts.*

Moreover, participants were monitoring the constructed artifact regularly, thus the evaluation of the artifact is seen as an iterative process. In the stage of revision, participants moved back and forth within the constructed artifact, making iterations in terms of organization of material, content and language usage:

> *Siraj interview: When I go to the Wikispaces I may write something wrong but after two or three weeks I go back and read through and I realize that I made a mistake. Maybe I did not know about that thing before and then I get to know. As I am passing through that text I see that I made a mistake and I correct it. And if I correct is not very easy to forget it.*

Additionally, the instructor often challenged participants by highlighting their mistakes within the wiki or by providing comments in Facebook. As participants were reviewing what they have written they were challenged to think till they find the correct answer:

> *Instructor field notes: I tried to point out their mistakes on the wiki by highlighting incorrect sentences and also pointing out orally their mistakes.*

In social constructionism environment, students are not criticized for errors but are rather encouraged to proceed on, and build on their mistakes.

5.5 Role(s) Adopted by Students and Instructor

Instructor's role in the social constructionism framework can be marked as facilitator, supporter and reviewer. The instructor facilitates the orientation phase and reviews the constructed artifact. However, the instructor acts more as a member of the construction team rather than a judge. To this aim, the instructor supports students emotionally by giving advice and encouragement related to their progress:

> *Instructor's reflections: Remember that in every fight the first step is to believe in yourselves. What you have done so far proves your potentials.*

Students act primarily as active constructors and reviewers of the artifact. Typically, in a language class the aim is to memorize as much information as possible, however in the social constructionism platform learners are encouraged to focus and understand their errors and involve in the process of correcting them.

6 Conclusion and Future Work

The current study explored social technologies from the perspective of constructionism. The three dimensions that emerged along with the respective actions that accompany each dimension reveal further dynamics of social technologies as social constructing platforms or in Papert terms "object-to-think". A social constructionism action model that takes into consideration the dynamics of social technologies could be represented in the triptych: *exploration of ideas*, *construction* and *evaluation of artifact*. This triptych captures the actions that take place throughout the social construction of an online artifact process. The learner is an energetic part of the whole process starting from exploration throughout the evaluation of the artifact. Peers and instructor are also involved in the process in multiple actions. The instructor acts as facilitator in orienting the ideas in the exploration phase; supporter for participants in the construction phase and reviewer in the evaluation phase. Peers are involved in co-forming decisions in the whole process. Social technologies are an integral part of the process; however the essence of social constructionism lies in the artifact itself that produces understanding through construction of an explicit representation.

In our view these actions provide a view of different aspects of how the construction of an online artifact manifests in practice. From the perspective of knowledge creation, the construction of an online artifact within social technologies allows learners to think and understand abstract scenarios by linking them with a tangible artifact. From the perspective of design, this paper views constructionism as a fertile ground for learners to experience the design of an online artifact not as learners but as designers and researchers. Although we frame social constructionism within the limits of CALL, we believe that the emergent dimensions can serve future efforts to support learning, collaboration and problem solving.

The results of this study reveal that social constructionism demands a series of actions, including orientation, brainstorming, material exploration, outlining, editing material, revising, instructor reviewing, peer reviewing and presenting. These actions are expected to supply designers, instructors, researchers and practitioners with a better understanding of the affordances of social technologies, leading to a new perspective of their use.

We encourage future research to explore the validity and applicability of our code scheme in areas other than language. Moreover, an overall framework needs to be developed that will reveal theoretical and methodological aspects of social constructionism. Taking into consideration the intertwined relationship between language and culture, further research will be conducted exploring whether cultural scenarios can provide an exemplary framework through which social constructionism can be implemented. Finally, we seek for further studies that explore the components of constructionism in other environments and learning subjects, online or offline, which could result in its wider applicability as a means for enhancing knowledge.

Future research could be conducted in applying social constructionism in social and 3D environments as well as mobile applications in order for learners to construct their "objects-to-think". The framework and the actions that are described in this study may also inform several stages of research in HCI, enabling the analysis, design, development and evaluation process within a social environment following the aforementioned actions of social constructionism. Moreover, the actions that take place throughout social constructionism could inform designers to refine the development of social platforms so as to facilitate the construction of online artifacts.

From the perspective of practitioners, social constructionism can inform curriculum design, materials development and classroom praxis. Research on the possible contributions of social technologies as social construction platforms is still in its infancy. More research on different artifact types and different subjects could clarify aspects of constructionism and help create better pedagogical approaches for various instructional purposes.

Acknowledgements. We would like to thank Fred Kipkorir, Nelson Kiprotich, Mutyaba Saul and Sirajdin Ssemugenyi for their support and invaluable contribution for conducting this study.

References

1. Abraham, L.B.: Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. Computer Assisted Language Learning 21(3), 199–226 (2008)
2. Alexander, B.: Web 2.0: A New Wave of Innovation for Teaching. Educause Review 41(2), 32–44 (2006)
3. Arslan, R.Ş., Şahin–Kızıl, A.: How can the use of blog software facilitate the writing process of English language learners? Computer Assisted Language Learning 23(3), 183–197 (2010)
4. Bloch, J.: Abdullah's blogging: A generation 1.5 student enters the blogosphere. Language Learning & Technology 11(2), 128–141 (2007)
5. Bradley, L., Lindström, B., Rystedt, H.: Rationalities of collaboration for language learning in a wiki. ReCALL 22(2), 247–265 (2010)
6. Dippold, D.: Peer feedback through blogs: student and teacher perceptions in an advanced German class. ReCALL 21(1), 18–36 (2009)
7. Gamper, J., Knapp, J.: A review of Intelligent CALL systems. Computer Assisted Language Learning 15(4), 329–342 (2002)

8. Glassman, M., Kang, M.J.: The logic of wikis: The possibilities of the Web 2.0 classroom. The International Journal of Computer-Supported Collaborative Learning 6(1), 93–112 (2011)
9. Goldberg, A., Russell, M., Cook, A.: The effect of computers on student writing: A meta-analysis of studies from 1992-2002. Journal of Technology, Learning and Assessment 2(1) (2003)
10. Ioannou, A.: Collaborative Learning: The Promise of Wikis. Word Journal of the International Linguistic Association 38(3), 213–224 (2011)
11. Kárpáti, A.: Web 2 technologies for Net Native language learners: A "social CALL". ReCALL 21(2), 139–156 (2009)
12. Kern, R.: Perspectives on technology in learning and teaching languages. TESOL Quarterly 40, 183–210 (2006)
13. Kessler, G., Bikowski, D.: Developing collaborative autonomous learning abilities in computer mediated language learning: attention to meaning among students in wiki space. Computer Assisted Language Learning 23(1), 41–58 (2010)
14. Lee, L.: Promoting intercultural exchanges with blogs and podcasting: a study of Spanish–American telecollaboration. Computer Assisted Language Learning 22(5), 425–443 (2009)
15. Lee, L.: Exploring wiki–mediated collaborative writing: a case study in an elementary Spanish course. CALICO Journal 27(2), 260–276 (2010)
16. Lee, L.: Fostering reflective writing and interactive exchange through blogging in an advanced language course. ReCALL 22(2), 212–227 (2010)
17. McLoughlin, C., Lee, M.J.W.: Social software and participatory learning: Pedagogical choices with technology affordances in the Web 2.0 era. In: Atkinson, R.J., McBeath, C., Soong, S.K.A., Cheers, C. (eds.) ICT Providing Choices for Learners and Learning Proceedings Ascilite Singapore 2007, vol. 2007, pp. 664–675. Ascilite Singapore (2007), `http://www.ascilite.org.au/conferences/singapore07/procs/mcloughlin.pdf` (retrieved)
18. Miceli, T., Murray, S.V., Kennedy, C.: Using an L2 blog to enhance learners' participation and sense of community. Computer Assisted Language Learning 23(4), 321–341 (2010)
19. Nisiforou, E., Laghos, A.: An Overview: The Development of ICT in the Educational System of Cyprus. In: Global Learn Asia Pacific, vol. 2011(1), pp. 534–539 (2011)
20. Papadima-Sophocleous, P., Parmaxi, A.: The use of wiki in teaching and learning Greek as a second language for specific academic purposes: challenges and future perspectives. In: Buston, J., Tsagari, D., Doa, F. (eds.) Foreign Language Instructional Technology: Theory & Practice, pp. 128–139. University of Nicosia Press, Nicosia (2012)
21. Papert, S.: Mindstorms: Children, Computers and Powerful Ideas. Basic Books, New York (1980)
22. Papert, S., Harel, I.: Situating Constructionism. In: Papert, S., Harel, I. (eds.) Constructionism, vol. 36, pp. 1–11. Ablex Publishing (1991)
23. Papert, S.: The Children's Machine: Rethinking School in the Age of the Computer. Basic Books, New York (1993)
24. Parker, K.R., Chao, J.T.: Wiki as a Teaching Tool. Learning 3(3), 57–72 (2007)
25. Patton, M.Q.: Qualitative research & evaluation methods. Sage Publications, Incorporated (2002)
26. Pfeil, U., Zaphiris, P.: Patterns of empathy in online communication. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 919–928. ACM Press (2007)
27. QSR International. Nvivo 10 qualitative research software

28. Rivens Mompean, A.: The development of meaningful interactions on a blog used for the learning of English as a Foreign Language. ReCALL 22(3), 376–395 (2010)
29. Ruschoff, B., Ritter, M.: Technology-Enhanced Language Learning: Construction of Knowledge and Template-Based Learning in the Foreign Language Classroom. Computer Assisted Language Learning 14(3-4), 219–232 (2001)
30. Seitzinger, J.: Be Constructive: Blogs, Podcasts, and Wikis as Constructivist Learning Tools. Learning Solutions eMagazine, 1–16 (July 31, 2006), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.4071&rep=rep1&type=pdf (retrieved)
31. Smith, P.L., Ragan, T.J.: Instructional design, 3rd edn. Wiley, New York (2005)
32. Stemler, S.: An overview of content analysis. Practical Assessment Research Evaluation 7(17), 479–498 (2001)
33. Sun, Y.C.: Voice blog: an exploratory study of language learning. Language Learning and Technology 13(2), 88–103 (2009)
34. Taylor, A.M.: The effects of CALL versus traditional L1 glosses on L2 reading comprehension. CALICO Journal 23(2), 309–318 (2006)
35. Taylor, A.M.: CALL-based versus Paper-based glosses: Is there a difference in reading comprehension? CALICO Journal 27(1), 147–160 (2009)
36. Welsh, E.: Dealing with Data: Using NVivo in the Qualitative Data Analysis Process 1. Approaches to Qualitative Data Analysis 3(2), 1–7 (2002)
37. Yun, J.: The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. Computer Assisted Language Learning 24(1), 39–58 (2011)
38. Zaphiris, P., Zacharia, G., Rajasekaran, M.: Distributed Constructionism through Participatory Design. Group, pp. 164–179 (2003), http://books.google.com/books?hl=en&lr=&id=GaUGjcpjo7AC&oi=fnd&pg=PA164&dq=Distributed+Constructionism+through+Participatory+Design&ots=nr3a7UZes7&sig=Mwum2xiy41oDbi2x4zrRUe-0Z8M (retrieved)
39. Zhao, Y.: Recent development in technology and language learning: A literature review and meta-analysis. CALICO Journal 21(1), 7–27 (2003)
40. Bateman, S., Farzan, R., Brusilovsky, P., Mccalla, G.: OATS: The Open Annotation and Tagging System. Information Sciences (2006), http://scholar.google.com/scholar?hl=en&q=OATS+the+open+annotation+and+tagging+system&btnG=Search#0 (retrieved)
41. Bateman, S., Brooks, C., McCalla, G., Brusilovsky, P.: Applying Collaborative Tagging to E-Learning. Human Factors 3(4), 1–7 (2007), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.8892&rep=rep1&type=pdf (retrieved)

Usability Specialists as Boundary Spanners – An Appraisal of Usability Specialists' Work in Multiparty Distributed Open Source Software Development Effort

Netta Iivari

Department of Information Processing Science, University of Oulu,
P.O. Box 3000, 90014 Oulu, Finland
netta.iivari@oulu.fi

Abstract. This study examines the work of usability specialists in a multiparty research project producing an open source learning application for children, with children. Children as a user group has been acknowledged decades ago and methods for involving them have been devised, but there is a lack of research examining what happens to children's input in practice, when integrated with actual development. The paper contrasts the empirical findings with the existing research on the usability specialists' roles and with the knowledge management literature on boundary spanning, which argues that for successful knowledge sharing and arriving at shared understandings there needs to emerge boundary spanners and boundary objects and a new joint field of practice within which the experts involved can collaborate. This paper argues for the boundary spanner position to be acquired by usability specialists. Instances of successful boundary spanning are described and conditions for successful boundary spanning are discussed.

Keywords: Usability specialist, boundary spanning, boundary object, children.

1 Introduction

This study examines the work of human computer interaction specialists (HCI), whose work practice has been studied under the varying labels such as usability professional [7, 13], usability specialist [4, 6], usability practitioner [25] and HCI practitioner [12, 15]. Here, these experts will be called usability specialists [in line with 3, 13], as this is a widely known and accepted job role in industry and there already is an association for usability professionals. This choice naturally makes usability as the main goal of their work, even though they may place their emphasis also or actually even more on usefulness or high quality user experience.

However, even though the importance of usability specialists' work has generally been accepted in the industry, there also are pertinent problems in their work. One of those has been their work of not having impact on the solution under development [e.g. 4, 7, 15]. Usability specialists, if not totally ignored, may be in informative or consultative roles, meaning that they may only be allowed to provide information or

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 571–588, 2013.

to comment on predefined design solutions, without having decision-making power or ability to directly impact the design solution [15]. Also studies in open source software development have shown that usability specialists, when trying to contribute to the development, tend to remain isolated and alienated and their work not necessarily having any impact on the actual solution [e.g., 1, 26, 27]. However, it is recommended that usability specialists should be in participative role, actively taking part in the design process, having decision-making power regarding the solution [15, 27] if not acting as the designers having authority to make the important design decisions concerning the product quality [e.g. 8].

This paper will examine what kinds of roles usability specialists adopt in multiparty, distributed IT development settings. The case inquired in this paper is a multiparty, distributed research project developing an open source application for children, with children. This study thus truly represents a multicultural setting for HCI: the project participants are located in different countries, working in different organizations (companies or universities) and represent different disciplines (information technology, human computer interaction, educational science) and generations. Children, i.e. the future users of the application, are important participant group who possess valuable knowledge on what being a kid entails [11] that should be utilizable during the development, among other expertise. Although there has been extensive research interest in the means enabling children to take part in IT development [e.g. 10, 24], there is a lack of research on the participation of children in large-scale multiparty development efforts. The literature on global, distributed, open source and multiparty development efforts reveals that nowadays there might be multiple organizations, professions, areas of expertise, disciplines, and nationalities involved in IT development, making collaboration challenging [1, 5, 19, 21, 22, 26]. To make sense of the work of the usability specialists in such a complex setting, indeed involving multiple organizations, professions, areas of expertise, disciplines, nationalities as well as generations, we will rely on theoretical framework on boundary spanning [22]. This framework maintains that to be able to successfully span the boundaries of different organizational and professional settings, there needs to be people acting as boundary spanners as well as common, shared objects acting as boundary objects [22]. The conditions for such to emerge are discussed in this paper.

The paper is structured as follows. The next section discusses the existing body of knowledge related to the work of usability specialists, pinpointing areas in need of future study. The existing categorizations of usability specialist roles are discussed and the theoretical framework of boundary spanning is introduced and combined with the role repertoire presented. The third section introduces the empirical case of this interpretive case study as well as the research methods used in relation to studying the case. The fourth section presents the empirical findings, the fifth section discussing their implications and limitations as well as paths for future work.

2 Theoretical Framework

Even though user interface design has aroused researchers' interest already during 1980s, empirical studies on the work of usability specialists are more recent. During 2000s, there has been increasing interest in this topic. Gulliksen and his research group have reported numerous studies, mostly concerning public authorities in Sweden [e.g. 3, 12, 13], while Iivari has analyzed usability specialists work practices in product development in Finland [15, 16]. These are qualitative inquiries on the matter, but there is also some survey research carried out [6, 12, 37]. Moreover, one can identify recent HCI journal issues empirically addressing usability work in organizations (see e.g. Interacting with Computers 18(4), International Journal of Human-Computer Interaction 21(2), International Journal of Technology and Human Interaction 5(3)). Furthermore, there also are some studies addressing usability specialists' work in recent development settings such as in distributed open source software development [e.g. 26, 27], reporting on usability specialists' ways of working when face-to-face contact is impossible. However, the literature remains silent of multiparty development efforts, where there are numerous stakeholder groups with which usability specialists have to cooperate. On one hand, the HCI literature has touched upon the issue in the sense that it maintains that usability specialists have to act as usability champions and evangelists inside their organizations and try to seduce and convince numerous stakeholder groups, i.e. developers, managers, sales, marketing and documentation, to buy into usability [2, 8, 13, 15, 16, 25, 37]. On the other hand, the literature on multiparty IT development efforts reveals that there nowadays are parties from numerous organizations and even countries involved, representing different kinds of areas of expertise: not only technology, marketing and business, but also strategy, manufacturing, education, curation, meteorology etc. [5, 19, 20, 21, 22]. In this kind of a design team collaboration is a true challenge [e.g. 20].

2.1 Usability Specialist Roles

There already are studies that have defined a set of roles for usability specialists to adopt during development. Iivari [15] has identified four of those: informative, consultative, participative and designer roles. In the informative role, usability specialists provide information to developers about users based on their empirical studies concerning the actual or potential users or on their general state-of-the-art HCI knowledge [15, 16]. However, this role is problematic as it does not necessarily have any effect on the actual design solution [15]. In the consultative role, usability specialists provide feedback to the already made design decisions, again either relying on the user feedback gathered through empirical evaluations or on the expert evaluations utilizing the general HCI knowledge [15, 16]. This role may not have any impact on the actual design solutions either and another problem is that this role may position usability specialists as police, only pointing out negative issues that may hinder their possibilities to have any impact even more [15]. In the participative role, furthermore, usability specialists are accepted as active participants in the design process, having some decision-making power regarding the solution among the other team members

[15, 16]. In this situation usability specialists indeed can contribute, but it is a challenge to involve a usability specialist in every relevant design situation [15]. Moreover, in the designer role, usability specialists are given the authority to make the important design decisions concerning the product quality based on their expertise [3, 8, 15, 16]. However, this may actually remove the job role of the usability specialist altogether that has also proven out to be problematic as in the situation in which every designer should be also an expert in usability; usability may end up in being taken care of by nobody [15].

Finally, also a facilitator role has been identified in the HCI literature [4, 7, 13, 16, 17, 30]. Either it is emphasized that usability specialists should orchestrate design session and facilitate collaboration among developers and users [4, 7, 13, 16, 17, 30] or usability specialists are positioned as evangelists that should advocate usability inside their organization even more broadly, targeting the management, sales, marketing and documentation [2, 8, 13, 15, 16, 17, 30, 25, 37]. The developers, in particular, should perceive usability specialists as team members and allies [25, 37], but also the other stakeholder groups should be addressed [2, 8, 25, 37].

2.2 Usability Specialists as Boundary Spanners

In multiparty design teams, in which there are experts representing different disciplines, professions, organizations and nationalities cooperating, it has been reported that communicating, collaborating, and arriving at shared understandings are very challenging [5, 19, 20, 21]. However, knowledge sharing and the creation of shared understandings are also postulated as vital for successful multiparty design teams. For these reasons, in addition to the existing HCI literature on the work practice of usability specialists, this paper relies on the theoretical framework on boundary spanning [22], which focuses on a successful knowledge sharing.

Within this framework, important is the emergence of boundary spanners, boundary objects and a new joint field of practice, within which the agents involved can share knowledge [22]. The emergence involves that the boundary spanners achieve at least a peripheral understanding of each involved practice as well as legitimacy as negotiators on behalf of the parties whose interests they are attempting to represent. In addition, they need to transform their own practices so that they accommodate the interests of the other parties. [22] They also need to create and use boundary objects, which in turn need to attain local usefulness and common identity from the viewpoint of each involved party [22, 29]. The boundary objects may be existing ones or they may be created or modified by the boundary spanners specifically for the purposes of boundary spanning [22].

Boundary spanners and particularly boundary objects have been extensively discussed in the literature and shown to contribute to knowledge sharing and to the development of shared understandings among various experts [20, 22, 23, 29] as well as to mobilizing for design action and to legitimizing design knowledge [23]. While boundary objects are not the main focus of this paper, they will be touched upon in the empirical analysis. The notion of boundary spanners, on the other hand, will specifically be used, to give more depth to the analysis of the roles assigned for usability specialists. It will enable making sense of how usability specialists can act as facilitators in a distributed multiparty design team: how they can enable knowledge sharing and the creation of shared understandings among the various experts. According to Levina and Vaast [22], this involves 1) gaining at least a peripheral understanding of each involved practice; 2) gaining legitimacy as negotiators on behalf of the fields whose interests they are trying to represent 3) transforming their own practices so that they accommodate the interests of the other fields; 4) producing boundary objects, which need to gain local usefulness and common identity from the viewpoint of each involved practice. With this analytic lens, the empirical case will be examined.

3 Research Design

The case examined in this paper is a research project that develops a learning application for children, with children. The overall idea for the application came from researchers within the field of educational science, and the justification for the specific features of the application are derived from educational science research. These researchers had earlier been involved in creating an initial version of this application and this project was to produce an enhanced version of it. The other parties—IT experts, usability experts, and additional educational science experts, representing different research institutions and two IT companies—became involved in the project ideation later on. The project participants were located in four countries, two IT companies, and five research institutions; the children involved were from numerous schools and kindergartens from different countries.

The project was to last for three years. The first year was about specifying the requirements for the forthcoming application and designing the application, including both educational, usability and software design. The second year was to be mainly about implementing the application, while the third year was to be about evaluating the application from the viewpoint of project goals. The contribution of the usability specialist was significant especially during the first year of operation on which this analysis will concentrate. During the first year, work related to following aspects was to be carried out: requirements specification, usability and design. In addition, development work was to begin during the last two months. The work was distributed and involved multiple project partners. The project partners extensively relied on email and a shared data repository for knowledge sharing and communication. In addition, video conferencing and voice chat solutions were used and two face-to-face meetings organized during the year. The responsibility of the work to be carried out was a shared responsibility of multiple partners of different disciplinary backgrounds, working in different institutions and countries. For all major project deliverables, however one specific institution was nominated as the responsible leader.

The usability specialists had as their task to take care of usability and child-centeredness. They worked extensively with children during the year. Numerous children were involved in two different countries. The work involved different kinds of experiments connected with evaluating certain design ideas, empirical user testing, paper prototyping with children, interviews, observation, and different kinds of design sessions [see e.g. 10, 11, 24]. The children acted as testers, informants, and design partners [see 10], drawing, talking to researchers, using the earlier version of the application or the new version presented as a paper prototype, and playing games or creating prototypes of the application of art craft materials, among other activities.

This paper reports an interpretive case study on the work of usability specialists in this multiparty, distributed open source development effort [see e.g. 18, 31]. In this study, in line with Klein and Myers [18: 69], it is assumed that "our knowledge of reality is gained only through social constructions such a language, consciousness, shared meanings, documents, tools, and other artifacts" and the study does not try to identity independent and dependent variables, but instead to understand the complex case and human meaning making in it in more depth. My role can be characterized as "involved researcher" – I had a direct personal stake in the outcomes and interpretations, but on the other hand I was able to get a direct sense of the field from the inside [31]. I was involved in the project already during the funding application preparation and I acted as a manager and supervisor of more junior researchers working on the project. Therefore, I acted both as a participant and as an observer in the project. I represent one of the usability specialists in the case.

The research material consists of documentation produced during the project. The material consists of official project documents, project deliverables (of particular interest are the requirements specification and design documents), different kinds of memos and unofficial documentation (e.g., different kinds of sketches, scenarios and drawings) and email correspondence among the project participants. The documentation was created independently of my research interest for the purposes of the project, but they were collected to form the research material to be examined.

During data analysis, I first reviewed the project activities from the viewpoint of usability specialists: I identified the instances in the empirical material that could be connected with their role in the development. After collecting this huge amount of evidence together, I started to make sense of it within the analytic framework developed. I identified the boundaries that needed spanning in this case. Levina and Vaast [22] consider organizational and professional boundaries which apply quite well to the usability specialists' interaction with the educationists and developers in this case. Afterwards, I identified successful cases of boundary spanning: I looked for situations in which the usability specialists had succeeded in facilitating collaboration and knowledge transfer across the identified boundaries – the identification was based on actual evidence of the impact of the usability specialists' work or on the other parties acknowledgement of the usability specialists success in creating a new joint field of practice where the experts successfully shared knowledge and created shared understandings. After identifying such cases, I also gathered data on whether the usability specialists had aimed at 1) gaining at least a peripheral understanding of each involved practice; 2) gaining legitimacy as negotiators on behalf of the parties whose interests they are trying to represent 3) transforming their own practices so that they accommodate the interests of the other parties; 4) producing boundary objects that have local usefulness and common identity from the viewpoint of each involved practice.

4 Empirical Insights

In this section two differing types of successful boundary spanning will be presented. In the first type the usability specialists succeed in influencing the designs and prototypes created by the educationists and the developers, making them to integrate into their work objects issues that the usability specialists considered important after working with children. In the second case, the usability specialists succeed in facilitating shared understandings among the educationists and the developers related to specifying the software requirements for the project, the other parties thanking the usability specialists as a useful 'link' between the developers and the educationists.

4.1 Usability Specialists Impacting the Educationists' and the Developers' Work Objects

As mentioned, the usability specialists used different kinds of methods for working with children, i.e. with the future users of the application. The children took part as informants, testers and design partners [10], Through all the usability activities, one can conclude that the usability specialists become relatively well informed of their users. The usability specialists also seriously tried to take into account that they were working with children, carefully planning all their sessions with children, including play and singing, for example: "*We implemented the program as planned. As a start and addition, [a usability specialist] sang with the children a [song] to remind the children who we were.*" (Usability specialist, memo)

The educationists started the project work by identifying requirements for the application. First they sent their ideas through email, but soon they captured them into a lengthy PowerPoint presentation on the matter. Afterwards, they started to capture their ideas into hand drawn scenarios of use. Later on, they delivered the requirements in a table format. All these requirements documents were sent to the other project partners through email. Especially the educationists wished for the developers' feedback to the ideas presented, while also the usability specialists were free to comment.

The usability specialists, on the other hand, started planning their empirical work with children. Before carrying it out, they asked for input from other parties. The educationists hoped for children's feedback to some icons and design ideas as well as children's ideas relating to some features planned for the application. Based on their empirical work with children, the usability specialists informed the other parties of their results. They also created two formal project deliverables: Usability Requirements and Usability Design. The first one described the evaluation results concerning the earlier version of the application, user feedback to some initial design ideas of the educationists as well as children's own ideas and designs. The usability specialists had also carried out expert evaluations on the earlier version of the application and on the scenarios, the results of which they presented as well as some general state-of-the-art HCI knowledge on interaction design and children. The Usability Design document, on the other hand, described the screen contents, the functions available, possible user actions and system responses. All in all, the usability specialists were positioned in informative, consultative and designer roles, as they were providing information and feedback to the other project parties as well as making important design decisions themselves [see 15].

The educationists, then again, created their own Educational Requirements and Educational Design deliverables. The usability specialists had taken the educationists' scenarios as a basis and evaluated and refined them together with children, based on which they had created their usability design. The educationists, however, had continued their work with the scenarios; hence, those that the usability specialists had evaluated were not the most current ones anymore. The educationists had then based their designs on the refined scenarios that neglected some of the results of the usability specialists' work. While producing their documents, neither party carefully examined the other party's documentation to prevent conflicts and overlapping work. Instead, both parties, when delivering their documents, mentioned that there might be some overlap between their and the other party's documents, and asked others to check that. Unfortunately, the educationists' and the usability specialists' design documents were scheduled to be delivered at the same time, even though the usability specialists document was expected to create '*usability on top*', which naturally was impossible as they did not have the educational design at hand when creating their own design. These two designs ended up in being in conflict with each other and the educationists and the usability specialists had to negotiate the designs (adding the participative role to the usability specialists' role repertoire [see 15]).

The usability specialists critically reviewed the educationists' design from the viewpoint of children and pointed out, based on their empirical data, many issues that should be modified to better suit the target user group, e.g.: *"We are wondering here together with [another usability specialist] why the user interface for the younger age group has been done anew and our findings from last spring neglected? In the project, feedback was gathered and ideas generated based on the scenarios produced [by the educationists]. Now it seems that this feedback has been ignored but instead the work has been continued based on the own scenarios (for instance the door, house, and the pictures of (…) and (…) children have been left out …). I would say that one should prefer already evaluated designs."* (Usability specialist, email) *Briefly related to some central functional elements (…) 1) House and home in general was a central and important element for the 5 to 6 years old and it should be kept in the main menu. (…)The [tutor] should also be kept in the same place, for instance in the upper right corner. (…) the same place in every screen was found to be good for 5 to 6 years old in the prototyping; (…)"* (Usability specialist, email)

Related to the many of the suggestions, the educationists made changes to their design documents. However, this did not happen related to all suggestions, instead the educationists referred to certain project goals or to their authority to settle the educational aspects when keeping certain issues as they were. The educationists and the usability specialists sent their design documents to each other and exchanged numerous emails on the matter. In addition, the usability specialists created a document called a List of concerns, in which they in a table format listed all the problems they identified from the educationists' design. In addition to the issues that according to the usability specialists were to be changed in the educationists' design due to their user data, the usability specialists also identified unclear or inconsistent issues from the document. Afterwards, the educationists made some changes to their designs, but ignored other issues. They listed their responses in the List of concerns document, pointing out what they had changed, what they hadn't changed and reasons for the decisions, marking their response in a different font color.

The developers also received the usability specialists' requirement and design documents. The developers seemed to value those, especially some documents that were produced before creating the formal project deliverables on the matter. An informal Usability Requirement document, which the usability specialists started by graphically outlining the possible use cases of the application, was thanked: "*UI group requirements will be sent later (remark: [a usability specialist] sent within [software requirements] writing process …), extremely helpful for getting use cases).*" (Developer, memo) In addition, a developer reminded other project parties related to the initial Usability Design document: *"[A usability specialist's] students have sketched quite detailed design for [parts of the application]. It is based on your scenarios and have been usability tested (paper prototyping) with 5-6 years old children. The student group has made magnificent work! (…) I hope you can continue from that. (Just to remind to make sure that you are not doing overlapping work)*" (Developer, email) The developers were not very happy with the adopted waterfall development model but instead relied on the development of prototypes, into which they integrated the usability specialists design as soon as it was sent: "*The project is following waterfall model where only one cycle from requirements analysis to design, to implementation, and finally to testing and experimenting is done. (…) Thus, our process is far from ideal in a research project where results are unclear beforehand. (…) We have tried to overcome this limitation by using different process and schedule internally (…). For example, we have already built prototypes and framework for (…) user interface, and produced an initial user interface design. We will then modify those according to the requirements and design when the corresponding documents are ready.*" (Developer, email) Thus, the usability specialists' initial usability design became implemented very fast into the developers' prototypes.

4.2 Usability Specialists Facilitating Shared Understandings in the Design Team

The usability specialists also facilitated shared understandings among the design team during some occasions. For instance, in a situation in which a developer needed to know the maximum amount of files a child should be able to save during a use session, first an educationist replied, giving the answer from the perspective of educational science research, but also indicating that the question actually belongs to the field of the usability specialists. A usability specialist replied, indicating that studies with children should be carried out to be able to answer the question, but offering still some initial guesses. On the other hand, she also indicated that there is not only the question related to the amount of files to be saved, but also a question of versions as the idea was that children should be able share files as well as to continuously update them. The question of how to represent these versions to children should be considered. Additionally, she pointed out that handling this could be a nightmare for developers but anyway this was a logical problem that had to be solved. Here, the usability specialist reflects on her knowledge gained through working with children as well as indicates how these kinds of questions are dealt with by usability specialists, i.e. by empirically inquiring them. Interesting is that the usability specialist also started to consider the question from the viewpoint of versions and even from the viewpoint of coding – clearly indicating she had some understanding also of the developers' practice, not only of that of HCI.

Another case during which the usability specialists were acting as facilitators between the worlds of developers and educationists was when producing software requirements. There was a formal project deliverable related to which extensive collaboration between the project parties and different disciplines was expected. One educationist was, though, positioned as responsible for this deliverable. However, it was very difficult for her to gain input for this document from the project partners. The educationist sent numerous emails requesting other parties to contribute. Finally, other educationists sent some user characteristics descriptions, asking: "*I am very unsure what else you need. Could you please specify?*" (Educationist, email) A developer informed her: "*There are no requirements at the moment. So, we know who our users are, but not what to do with them or for them. (…) The document should be such that it could be given to a person who has not participated in the project and he could start to design or implement the (…) functionality based on the detailed requirement descriptions.*" (Developer, email) The educationists were expected to produce the main part of the document, i.e. the functional requirements. They had created their Educational Requirements deliverable and also an early version of Usability Requirements document sent by a usability specialist was available. In addition, the educationists had a template provided by the developers, even though the developers criticized the template as being too formal for the purposes of this kind of a research project and warned of not producing useless content just for the sake of it: *"I was not sure if that level of formality was needed in a research project (…) I have suggested that the sections with non-relevant information and empty content (mainly those non-functional requirements) can be omitted. (…) I see no practical reason to generate dummy content if it does not serve research purposes."* (Developer, email)

The educationists, based on the information presented above, specified the requirements and sent the document to the other project partners to comment. At this point of time a usability specialist critically reviewed the document. She pointed out that the document confused design solutions with software requirements as well as missed some things that are usually included in software requirements documents. There emerged a lengthy discussion between the project partners concerning this matter, the educationists pointing out that they were not educated for creating this kind of documents. The developers criticized the division of work that had led to this situation; all project parties agreeing that the division of work should have had been different: *"So, the deliverables were originally meant to be written by non-software professionals/researchers. We (SW persons) have participated in many video conferences and physical meetings where the requirements were discussed. However, it was not exactly clear to us what non-software people really wanted software to do, nor we had enough time to decipher that."* (Developer, email). A usability specialist offered to go through the document with her team and to improve it. The project participants happily accepted the offer: "*The project plan (…) was as argued by [a developer], mostly constructed by non-software focused people, who created the entire research idea. We [the educationists] only had some tiny little background experience on developing the (…) application. We were not familiar enough what a project like this could bring in front of our eyes. (…) [The usability specialists] are working on [the software requirements document] and trying to find a consensus with [the educationists] in the software requirements*" (Educationist, email). At this time, the educationists thanked the usability specialists as a highly useful "*link*" between them and the developers. The usability specialists later delivered the Software Requirements Specification document for other parties to review. The educationists went through the revised document and modified it further. Also a developer commented on some requirements and priorities. Afterwards, the educationists informed that they had checked the priority ratings when finalizing the document. All in all, this incidence again shows that the usability specialists possessed not only understanding of HCI, but also of software engineering and through their work, they clearly helped the design team to move towards a shared understanding of the software requirements.

5 Discussing the Conditions for Successful Boundary Spanning

The framework utilized in this paper maintains that to be able to successfully span the boundaries between diverse organizations and areas of expertise, there needs to be people acting as boundary spanners as well as common, shared objects acting as boundary objects [22]. Two forms of successful boundary spanning were identified from the case: 1) the usability specialists informing the educationists and the developers about their user data and succeeding in impacting the actual application design, there emerging a new joint field of practice between the usability specialists and the educationists or between the usability specialists and the developers, within which these experts could share knowledge and arrive at shared understandings about the appropriate application design; 2) the usability specialists facilitating shared understandings among the design team, all parties taking part within the new joint field of practice within which they could arrive at a shared understanding of the software requirements. In both cases a new joint field of practice emerged for the usability specialists, educationists and developers, but not involving the children. All these instances involved also the use of certain kinds of objects that succeeded at least partially to act as boundary objects that gained local usefulness from the viewpoint of each involved practice and that ultimately succeeded in transferring knowledge across the boundaries. Next, the findings of this study are discussed in relation to the conditions of successful boundary spanning as described by Levina and Vaast [22].

1) Gaining at least a peripheral understanding of each involved practice; This refers to the usual usability specialists' activity involving field studies and empirical evaluations together with the actual or potential users. In this case, the usability specialists interviewed and observed children, carried out different kinds of empirical evaluation sessions as well as organized several design sessions. Through all this work, one can say that the usability specialists likely gained at least a peripheral understanding of children: their needs, skills and preferences relevant from the viewpoint of the forthcoming application. On the other hand, the usability specialists also succeeded in showing that they possessed some IT skills and education, as they indicated of knowing how software development ought to proceed with associated documents (use cases. software requirements) as well as what is involved in designing software (cf. the versioning problem). This likely has happened already though their education, as usability specialists are many times educated in some sort of an IT department that includes also other courses than HCI. This enables the usability specialists to have at least a peripheral understanding of the practice of the developers, too.

The educationists were likely the most exotic group for the usability specialists in this project but no specific effort was placed on understanding their practice. Of course, some insights were gained almost naturally during the whole year of collaboration. The educationists sent the educational requirements material and wishes for the usability specialists' empirical work with children soon after the project started. Later on, more formal Educational Requirements and Design documents were sent and through those the usability specialists could again educate themselves. However, if aiming at adopting the boundary spanner position in multiparty IT development efforts, this is one place for the usability specialists to improve their practice: to consciously try to gain at least a peripheral understanding of each involved practice, not only of that of the users, straightaway when the collaboration is to begin.

2) Gaining legitimacy as negotiators on behalf of the fields whose interests they are trying to represent: The main responsibility of the usability specialists is to 'represent the users', the ignored group in systems development and computer science [9]. However, related to 'representing the users' there is a wide-spread problem of usability specialists not having any actual legitimacy to act as these 'representatives', as users usually are not even aware that there is this kind of a specialists group in the development speaking on behalf of them [16]. This applies especially in the product development context, in which users are not working inside the same organization or inside a specific customer organization for which the solution is developed; in these cases users might even be aware of usability specialists and their representation work. In the case examined in this paper the children did not see the usability specialists as their representatives in the development, as this would have involved an explicit effort of informing the children of that matter that did not happen. Moreover, the developers and the educationists were not informed of the position of the usability specialists as boundary spanners either, due to which the developers or the educationists unlikely viewed the usability specialists as negotiators trying to represent their interests. On the other hand, the educationists independently pointed out that the usability specialists were equipped to act as a link between them and the developers that indicates that this kind of role was still given to the usability specialists, therefore granting some legitimacy to their work, nevertheless. Despite that, more explicit positioning into this role is needed. Usability specialists should also become better informed of those whose needs they are to represent and on behalf of whom they are to speak.

3) Transforming their own practices so that they accommodate the interests of the other fields; The usability specialists have a repertoire of usability methods to use when carrying out their work. Related to children, there also is a body of work published on the matter [see e.g. 10, 24]. Related to working with children, furthermore, the usability specialists placed extra effort on finding suitable ways of engaging the children and for creating a nice atmosphere for the joint sessions. Likely this contributed to the success of the design and evaluation sessions in which valuable feedback, insights and design ideas were gained concerning children and the application. On the other hand, the usability specialists seemingly also tried to please the developers, e.g. through including some notations widely used by the developers (i.e. the use cases). No specific evidence related to the usability specialists trying to transform their own practice to better suit the interests of the educationists was encountered, however, if not counting the educationists being allowed to express wishes for the empirical work to be carried out by the usability specialists with the children.

4) Producing boundary objects, which need to gain local usefulness and identity from the viewpoint of each involved practice; In this case there were some objects used that can be argued of having gained local usefulness and common identity from the viewpoint of several parties. These objects were the Usability Requirements and Usability Design documents, the List of concerns created related to the conflicting designs produced, some emails negotiating the conflicting designs and the Software Requirements Specification document. The Usability Requirements and Usability Design documents were thanked by the developers as providing useful information. It might be that especially the use cases presented in the Usability Requirements document helped the developers in their work and made them to thank the usability specialists' documentation. On the other hand, the Usability Design document, describing the screen contents, available functions, possible user actions and system responses, enabled the developers to finalize their early prototypes and for that reason proved out to be highly useful as well. At that point the developers, due to their preference on early prototyping and iteration, needed as exact specifications as possible and the usability specialists succeed in providing such during a convenient time. The design documents provided by the educationists and the usability specialists later on contained the same information, while they were delivered too late from the viewpoint of the prototype development. In addition, the Educational Design document; even though very extensive description was not very exact on all details that might have made it less useful for the developers. All in all, one can conclude that the quite traditional usability specialists' documents were appreciated in this case by the developers. The results are also in line with studies that argue that the developers value redesign proposals and elaborate problem descriptions [14], even though in this study the documents did not only present usability problems and their redesign proposals, but also totally new requirements and designs. Nevertheless, the design provided by the usability specialists was very concrete and directly utilizable by the developers, which has been reported of being valued by developers [14]. In more theoretical terms one can argue that these boundary objects succeeded in transforming design knowledge between two different worlds and mobilized for design action, i.e. they enabled the developers to progress along the design path [see 23].

The educationists and the usability specialists did not utilize each other's documentation as should have been the case. When finalizing their design documents, both parties mentioned that there might be some overlap between these design documents. It seems that those documents were mainly used for legitimizing the design knowledge of the each party [see 23], not for creating shared understandings. These design documents therefore cannot be conceptualized as boundary objects transferring knowledge between these two practices, while one can say that the detailed emails on the matter and the List of Concerns document that both parties modified acted as such and enabled negotiating the design on a detailed level. In both cases, the descriptions were sent back and forth, each party adding comments and modifying text. These very mundane and rather dry and abstract textual descriptions served as boundary objects in this case. They enabled negotiating the designs at detailed level and exchanging opinions before reaching a consensus. They evidently allowed finding a common language across these different social worlds as well as acceptable common solutions [see 23]. These kinds of mundane tools, e.g. email messages and word processing software documents with tables and multiple font colors, were successful in this multiparty design effort. The main requirements for such boundary objects may be easy modifiability and support for negotiations involving multiple voices. The lack of context in these descriptions necessitated, however, more lengthy descriptions that would have been the case if the comments had been added to the design documents itself. Additionally, numerous advanced groupware systems would have been available to support this task, but were not considered in this project.

The Software Requirements Specification document, furthermore, enables an interesting analysis of the collaboration between these three parties. One of the usability specialists critically reviewed the document, pointing out many unclear issues, controversies as well as design solutions that should have been avoided at this point of time. The usability specialist also revised the document. At this point the position of the usability specialists as a link between the developers and the educationists was pointed out. Afterwards, the educationists critically reviewed the document and made their own modifications, not accepting all suggested by the usability specialists. In addition, the developers entered the discussion by requesting prioritization of the requirements and some progress was also achieved. While the template for software requirements specification as such was not perfect and a software requirements specification document, altogether, is not recommended to be used as a boundary object in multiparty projects, it, and maybe its limitations as well, enabled the usability specialists to acquire the facilitator position in the project, translating the educationists' ideas into more formal language and at the same enabling the developers to relate the descriptions more easily to their own specifications. A joint effort relating to defining the requirements is suggested also for other multiparty efforts, while there clearly is a need for boundary spanners, be they usability specialists or other kind of experts, to create a joint field of practice within which different parties can express themselves in their own language while at the same time somehow make their language, or at least allow others to transform their language, understandable to the other parties. In this kind of a situation a facilitator is needed to make sense of the requirements and to enable the different parties to comment on and to negotiate further the requirements. The boundary objects supporting this work would thus need to promote shared representational means among the participants and to transform diverse design knowledge towards a common solution [see 23].

6 Conclusions

This paper empirically examined what kinds of roles usability specialists adopt in multiparty, distributed IT development setting sand contrasted the empirical findings with the existing research on the suitable roles for usability specialists to take in the development and with the theoretical framework on boundary spanning [22] that outlines the conditions for successful knowledge sharing and arriving at shared understandings in collaborative settings involving numerous organizations, disciplines, areas of expertise and nationalities [5, 19, 21, 22], and maybe even generations [10]. The paper argues for HCI research to acknowledge that it would be useful for usability specialists to view themselves as boundary spanners, facilitating knowledge sharing and arriving at mutual understanding among multiparty design teams, involving at least users and developers in addition to usability specialists, as well as possibly sales, marketing, management and documentation [2, 4, 7, 8, 13, 15, 16, 25, 28] and perhaps even numerous other areas of expertise, disciplines, nationalities and generations [5, 10, 19, 20, 21, 22, 24, 26].

Theoretically, this paper shed light on the role repertoire usability specialists rely on in IT development. Findings from knowledge management research were utilized for scrutinizing and enriching the usability specialists' role repertoire. The data showed that the informative, consultative, participative and designer roles were all adopted by the usability specialists in the case project [in line with e.g. 3, 13, 15, 16]. The usability specialists were providing information about users based on their empirical work with children and on their general state-of-the-art HCI knowledge, as well as offering both empirical user feedback and feedback based on their expert evaluations [see 15]. They were also allowed to produce their Usability Design, offering them the designer position, while the later negotiations with the educationists settled them more into the participative role, acting as active participants, having some decision making power among other design team members [see 15]. Regarding the facilitator role, furthermore, even though the usability specialists were not orchestrating joint design sessions in which the developers and the educationists collaborated with the children [see 4, 7, 13], they nevertheless facilitated collaboration inside the design team including themselves, the developers and the educationists.

In this case the novel aspect is related particularly to this: to the ways the usability specialists contributed inside the design team. The analysis offers insights on how to develop the facilitator role of usability specialists further to enable usability specialists to act as usability evangelists and advocates inside their organization [e.g. 1, 2, 8, 13, 15, 16, 25, 37]. Although the whole HCI practice of 'representing the user' can be considered as a mediating practice between design and use [17, 30] and along these line the HCI methods as mediating information between divergent worlds, the framework of boundary spanning contributes by targeting the focus on the design team members as well as by requiring the objects and methods used to be modified to suit the needs and interests of the particular groups with whom the usability specialists are to collaborate. Moreover, the importance of gaining legitimacy as a representative of the other parties as well as truly advocating the interests of all these other parties were brought up. The usability specialists in this case did not try to represent the interests of the educationists. They were thus acting as *user advocates* [cf. 13], speaking on behalf of the users to the developers and the educationists, as well as *developer advocates*, speaking on behalf of the developers to the educationists, but it seems that they did not try to gain a peripheral understanding of the practice or to change their methods or tools to accommodate the interests of the educationists [cf. 22]. If the usability specialists were to acquire this position, they should, nevertheless, try take into account and negotiate on behalf of all the involved parties. Furthermore, gaining legitimacy as a representative of all the involved parties should be sought for.

The boundary objects identified in this case were of various kinds. The Usability Requirements and Usability Design documents thanked by the developers likely provided something concrete enough at a convenient time [cf. 14], transforming design knowledge between different social worlds and mobilizing design action [cf. 23]. The same documents, however, did not work for coordinating the efforts of the educationists and the usability specialists, in which case they seemed mainly to be used for legitimizing the design knowledge of each party instead [cf. 23]. On the other hand, plain emails and word documents sent as email attachments and stored in the shared data repository served a successful boundary objects through the use of which the parties negotiated a shared understanding of the appropriate application design. The Software Requirements Specification and the work involved in its creation also succeeded in creating a joint field of practice within which a mutual understanding of the functional requirements for the application could be reached. These very dry and abstract descriptions and the distributed way of working relying very heavily on email succeeded in serving the purpose. However, HCI research could figure out more appropriate tools for boundary spanning work, this including likely not only documentation support but some procedures for their production as well.

The boundary spanner position should be valuable for practitioners working as usability specialists in industry. This role could be utilized in multiparty, distributed and global development efforts where people representing numerous fields of expertise, organizations, nationalities or generations should together contribute to a common goal. In such cases the usability experts could consider how they could gain at least a peripheral understanding of each involved practice, not only of that of users. In addition, they could try to gain legitimacy as negotiators of the other parties' interests. Moreover, they could consider how they could transform their own practice so that it better accommodates the interests of the other involved practices. Finally, they could consider their work objects as potential boundary objects that should gain local usefulness and identity from the viewpoint of the other parties involved.

There are some limitations concerning the results. Those have been gained only from one case that is very specific in many ways. More cases should be included in the analysis. This case was a research project, while industrial cases would enrich the analysis. This analysis also considered distributed design. Face-to-face setting would offer additional opportunities for the usability specialists to accomplish the job of a boundary spanner. Another area of future work would be to devise novel or enhanced methods and tools for usability specialists acquiring this position.

Acknowledgements. I would like to thank all the project participants, including the children. This research has been funded by the Academy of Finland and EU.

References

1. Barcellini, F., Detienne, F., Burkhardt, J.: Participation in online interaction spaces: Design-use mediation in an Open Source Software community. International Journal of Industrial Ergonomics 39, 533–540 (2009)
2. Bloomer, S., Croft, R.: Pitching Usability to Your Organization. Interactions 4(6), 18–26 (1997)
3. Boivie, I., Gulliksen, J., Göransson, B.: The lonesome cowboy: A study of the usability designer role in systems development. Interacting with Computers 18(4), 601–634 (2006)
4. Borgholm, T., Madsen, K.: Cooperative Usability Practices. Communications of the ACM 42(5), 91–97 (1999)
5. Boujut, J., Blanco, E.: Intermediary Objects as a Means to Foster Co-operation in Engineering Design. Computer Supported Cooperative Work 2, 205–219 (2003)
6. Bygstad, B., Ghinea, G., Brevik, E.: Software development methods and usability: Perspectives from a survey in the software industry in Norway. Interacting with Computers 20(3), 375–385 (2008)
7. Bødker, S., Buur, J.: The Design Collaboratorium – a Place for Usability Design. ACM Transactions on Computer-Human Interaction 9(2), 152–169 (2002)
8. Cooper, A.: The inmates are running the asylum: Why high-tech products drive us crazy and how to restore the sanity. Sams, Indianapolis (1999)
9. Cooper, C., Bowers, J.: Representing the users: Notes on the disciplinary rhetoric of human-computer interaction. In: Thomas, P. (ed.) The Social and Interactional Dimensions of Human-Computer Interfaces, pp. 48–66. Cambridge University Press, Cambridge (1995)
10. Druin, A.: The Role of Children in the Design of New Technology. Behaviour and Information Technology 21(1), 1–25 (2002)
11. Druin, A., Stewart, J., Proft, D., Bederson, B., Hollan, J.: KidPad: a design collaboration between children, technologists, and educators. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 463–470. ACM, New York (1997)
12. Følstad, A., Bark, I., Gulliksen, J.: How HCI-practitioners want to evaluate their own practice. In: Morch, A., Morgan, K., Bratteig, T., Ghosh, G., Svanaes, D. (eds.) Proceedings of the Fourth Nordic Conference on Human Computer Interaction, pp. 417–420. ACM, New York (2006)
13. Gulliksen, J., Boivie, I., Göransson, B.: Usability professionals – current practices and future development. Interacting with Computers 18(4), 568–600 (2006)
14. Hornbæk, K., Frøkjær, E.: Comparing usability problems and redesign proposals as input to practical systems development. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 391–400. ACM, New York (2005)
15. Iivari, N.: Understanding the Work of an HCI Practitioner. In: Morch, A., Morgan, K., Bratteig, T., Ghosh, G., Svanaes, D. (eds.) Proceedings of the Fourth Nordic Conference on Human Computer Interaction, pp. 185–194. ACM, New York (2006)
16. Iivari, N.: Culturally Compatible Usability Work - An Interpretive Case Study on the Relationship between Usability Work and Its Cultural Context in Software Product Development Organizations. Journal of Organizational and End User Computing 22(3), 40–65 (2010)

17. Iivari, N., Karasti, H., Molin-Juustila, T., Salmela, S., Syrjänen, A.-L., Halkola, E.: Mediation between design and use – revisiting five empirical studies. Human IT – Journal for Information Technology Studies as a Human Science 10(2), 81–126 (2009)
18. Klein, H., Myers, M.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly 23(1), 67–94 (1999)
19. Lawrence, K.: Walking the Tightrope: The Balancing Acts of a Large e-Research Project. Computer Supported Cooperative Work 15(4), 385–411 (2006)
20. Lee, C.: Boundary Negotiating Artifacts: Unbinding the Routine of Boundary Objects and Embracing Chaos in Collaborative Work. Computer Supported Cooperative Work 16(3), 307–339 (2007)
21. Levina, N.: Collaborating on Multiparty Information Systems Development Projects: A collective Reflection-in-Action View. Information Systems Research 16(2), 109–130 (2006)
22. Levina, N., Vaast, E.: The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. MIS Quarterly 29(2), 335–363 (2005)
23. Mark, G., Lyytinen, K., Bergman, M.: Boundary Objects in Design: An Ecological View of Design Artifacts. Journal of Association for Information Systems 8(11), 546–568 (2007)
24. Markopoulos, P., Read, J., MacFarlane, S., Höysniemi, J.: Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers. Morgan Kaufmann Publishers Inc., San Francisco (2008)
25. Mayhew, D.: The usability engineering lifecycle: A practitioner's handbook for user interface design. Morgan Kaufmann Publishers, Inc., San Francisco (1999)
26. Nichols, D., Twidale, M.: Usability Processes in Open Source Projects. Software Process Improvement and Practice 11, 149–162 (2006)
27. Rajanen, M., Iivari, N., Anttila, K.: Introducing Usability Activities into Open Source Software Development Projects – Searching for a Suitable Approach. Journal of Information Technology Theory and Application 12(4), 5–26 (2011)
28. Rosenbaum, S., Rohn, J., Humburg, J.: A Toolkit for Strategic Usability: Results from Workshops, Panels, and Surveys. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 337–344. ACM, New York (2000)
29. Star, S., Griesemer, J.: Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science 19, 387–420 (1989)
30. Tuovila, S., Iivari, N.: Bridge Builders in IT Artifact Development. In: Österle, H., Schelp, J., Winter, R. (eds.) Proceedings of the 15th European Conference on Information Systems, pp. 819–830 (2007)
31. Walsham, G.: Interpretive case studies in IS research: nature and method. European Journal of Information Systems 4(2), 74–81 (1995)

Vimprint: Exploring Alternative Learning through Low-End Mobiles

Sheetal K. Agarwal[1], Jyoti Grover[2], Anupam Jain[2], and Arun Kumar[2]

[1] IBM Research – India, 8th floor, G2, Manyata Embassy Business Park, Bengaluru, India
[2] IBM Research-India, Plot 4, Block C, Vasant Kunj, Phase II, New Delhi, India
{sheetaga,jyogrove,anupamjain,kkarun}@in.ibm.com

Abstract. The Internet today provides a plethora of applications to assist anyone wanting to learn a new subject, language or a concept. Resources available include dictionaries, translation engines, downloadable e-books, tutorials, online courses etc. The rapid proliferation of smart phones has further provided richer visual applications that assist the user in learning on the go. However, all these applications are dependent upon the availability of Internet and/or an expensive computing device such as a smartphone or a computer. This puts them out of reach for a large section of society that consists of underprivileged people (economically or literacy-wise) and who probably need such tools the most. Also, many learning applications are pull-based and depend on the user's motivation to keep coming back for more. We present Vimprint - a system that offers an alternative mode of learning through telephony voice applications over any touchtone phone with a push-based interaction module. We present its design in the context of vocabulary building. Application of Vimprint system in the field is a work-in-progress and we present the results of a preliminary study conducted to assess its effectiveness.

Keywords: Social Computing, Developing regions, Voice Applications, Education, Diversity, Inclusiveness.

1 Introduction

The Internet has revolutionized the way learning can happen today. Online resources form a critical part of the learning process for any concept. Interactive web and smartphone applications engage the learner making it a fun experience. However, barely 35% of the world's population has access to Internet [1]. This makes all the rich online resources out of bounds for a vast majority of its potential users. Even for those who do have access to the Internet, illiteracy or English illiteracy acts as another major deterrent as most learning tools are offered in English.

In India, ability to converse in English not only improves chances of employment, it also elevates one's social image [4]. To address this demand from

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 589–596, 2013.

users in addition to classroom coaching and school based courses, books are published in local languages to teach English, cable TV operators provide simple English lessons [19] daily among other popular tools for language learning. While books are not interactive, cable TV based learning programs are broadcast in nature and do not offer an interactive experience. Classroom coaching is still the best option available but is expensive for many and skilled teachers are hard to find.

Recent research has attempted to introduce ICT in education in developing countries with the aim of providing a good supplementary channel to classroom teaching. Educational games [8, 5] to encourage math skills, improving vocabulary, learning shapes and colors etc., have been created and tested in primary schools in developing countries and have been found to be beneficial. While these games have been shown to be effective, they require a mobile, either a smart phone or a feature phone, or a computer where the game can be downloaded. Even though the mobile penetration has improved phenomenally, mobiles are often a shared resource in a family and its primary use is to make and receive phone calls. Users are barely able to use basic mobile phone features such as address books and the Short Messaging Service (SMS).

Among various technology solutions proposed to empower the underprivileged [3], telephony voice based systems have been particularly successful. Such applications break the barrier of illiteracy and make technology accessible in local language even to those who cannot read or write. Examples of scenarios where such telephony voice applications have been deployed include educational services in rural India [12], game based learning [18], voice based community knowledgebase [6], information dissemination and online discussion for farmers [9], healthcare [11] and job matchmaking [16]. In the context of learning, BBCs Janala[1] is a popular service in Bangladesh where users can access 3 minute English tutorials for 5 Taka (0.01 USD). The pervasive access to phones along with low call costs has driven the success of this service.

However, most of the work done for this user segment, in learning space, focuses primarily on delivering the content. A pull model is adopted relying on the learner's own interest and enthusiasm to revisit the content for retention. In the absence of a push mechanism, the less motivated learners tend to drop off thus defeating the purpose. In a classroom setup, the push based interaction is somehow enforced by the teacher through prompt repetitions in class or through assignments and tests. In this paper, we present a telephony voice based learning tool - *Vimprint*, focusing on the design of its vocabulary building module. Vimprint adopts a push based interaction method to increase retention of words accessed. It employs an adapted version of PimSleur's Graduated-Interval Recall method [10]. We believe that an active push based approach would simulate the aspect of a teacher that ensures retention of concepts taught, and hence result into more effective learning applications.

[1] `http://www.bbcjanala.com/`

2 Related Work

A few SMS/MMS based systems have been employed to teach a language [15, 13]. However, SMS based learning is an impractical option for many who are not literate enough to read and understand SMS content. For others, lack of spoken content leaves out the important elements of language teaching such as pronunciation, accent, etc. MMS on the other hand, does not work on most low-end phones which dominate in the developing world.. Telephony applications have been tried in the context of teaching English through multi-lingual story playback [2, 14] but they are pull based and depend upon learner's own motivation to pick up the concepts taught. Kumar et al [8] use application installed on the phone and mention about the possibility of indicated teaching over telephony voice as an explorable alternative.

3 VIMPRINT: Vocabulary Building

Vimprint is a learning system accessible through voice interaction over any touch tone mobile phone (including a landline). In this paper, we specifically focus on its vocabulary building module that aims to promote word literacy by enabling users to lookup words they have heard from others or read somewhere. On placing an ordinary phone call to the application, user is prompted to speak the word he wants to lookup. Vimprint retrieves and plays a list of words sounding similar to the user's utterance.

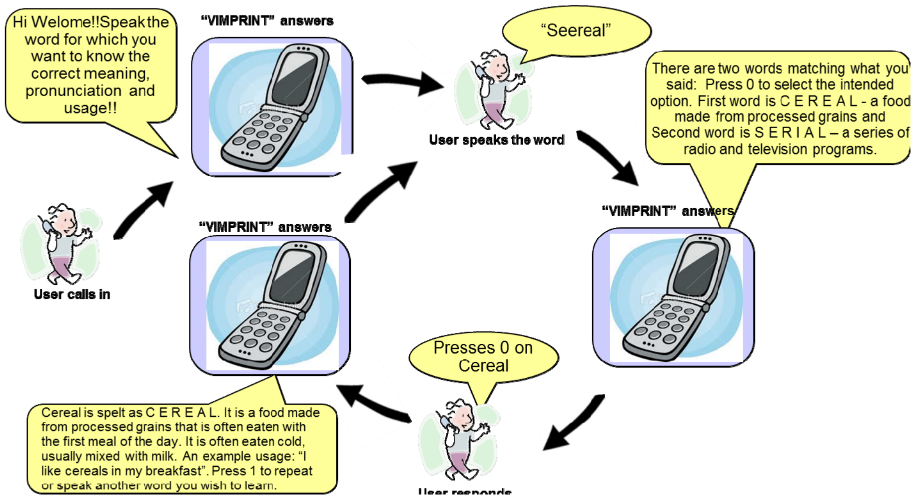


Fig. 1. User Interaction with Vimprint System for learning new words

For example, if the user says ``cereal" the API will return ``serial" and ``cereal". Along with every word it also plays its meaning to enable the user to identify the intended word. Figure 1 shows the interaction of a user with the Vimprint system for learning a new word.

Once the user identifies and selects the intended one, it's pronunciation, it's spelling and it's usage in a sample sentence are played. Vimprint focuses on enabling users to retain the words by following Pim Sleur's Graduated-Interval Recall method [5] for learning.

Graduated-Interval Recall

Typically, a number of techniques are used to improve the stickiness of a word in memory. These include using sample sentences from learners' day-to-day scenarios, synonyms or antonyms, repeating the word several times etc. Pim Sleur method is a popular method for learning a foreign language. The method follows the principle of recalling words at progressively increasing intervals to increase the possibility of moving the word from user's short term memory to long term memory. This is also known as the spacing effect. Pim Sleur's 1967 memory schedule is as follows: 5s 25s 2min 10 min 1hr 5hrs 1 day and so on. It follows the equation:

$$T1, T2 = 5 \times T1, T3 = 5 \times T2, \ldots Tn = 5 \times T(n-1)$$

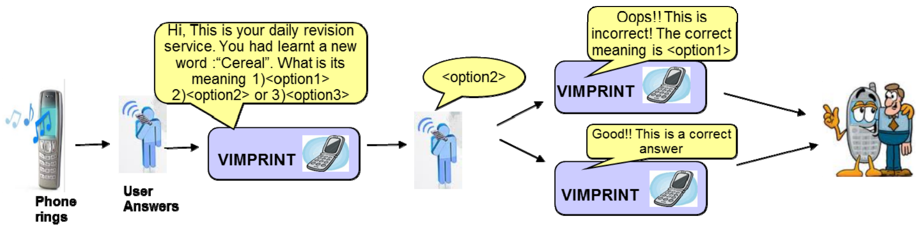


Fig. 2. User Interaction for Revision calls

In Vimprint, word recall is facilitated by proactively placing outbound calls to users to revise words previously seen. A user profile is created for each caller into the application. The user's preferred time for receiving revision calls; words accessed by the user, and actual calling times are maintained in a database. Since receiving a call requires a user's immediate attention, we adapted the Pim Sleur's method to take into account user's preferred time slots for receiving reminder calls. Figure 2 depicts the user interaction for revision calls. At each recall, a multiple choice quiz is rendered where user is asked to select the correct meaning of the word from multiple choices. While this method has been used to teach languages through books and audio courses, none of them are push-based in nature. The figure shows revision for a single word only, but a single revision call is generated on a scheduled day for all the words that need to be revised for that user.

4 VIMPRINT - Implementation

The Vimprint application is a Java based voice application platform that runs on an application server and generates Voice XML (VXML)[2] at runtime. These VXML pages are then rendered by a Voice Platform (Cisco voice gateway as shown in Figure 3). We use the Nuance Automatic Speech Recognition (ASR) Engine (i.e. Nuance Recognizer v9) for word recognition.

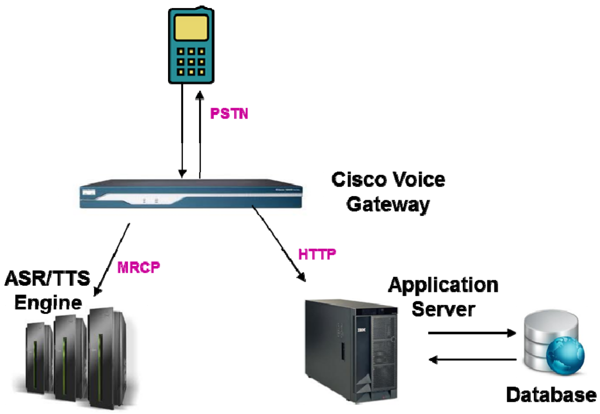


Fig. 3. Vimprint Infrastructure Setup

For each word the user speaks, we use the Nuance's API to return a list of similar sounding words and allow the user to select the word he wants to access. The selected word is then fed into a Vocabulary lookup module which fetches and plays back the details for that word. At the same time, that word is also scheduled for revision calls as per the schedule described above (refer Section 3).

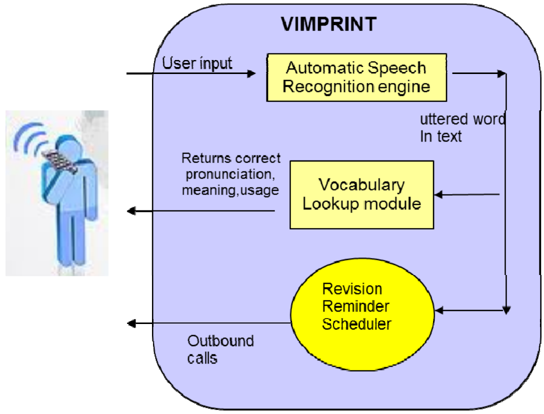


Fig. 4. Vimprint System Design

[2] `http://www.w3.org/TR/voicexml20/`

Our voice application platform that forms the core of Vimprint system has been tested to scale to 200 simultaneous calls on an 8-core cpu, 4 GB RAM Intel machine running Linux and hosting Vimprint system on Apache Tomcat application server and MySQL database server. Nuance was hosted separately on a similarly configured machine.

5 Preliminary User Study

Deployment and evaluation of Vimprint's Vocabulary learning module is a work-in-progress. Before going all out and recruiting users from the actual end segment who are often too busy in earning a livelihood, we conducted a preliminary user study with users outside the target segment. The aim was to test the effectiveness of the technique rather than fine tune the user interface for the actual target user group.

The preliminary study was conducted with 9 subjects over a period of 3 days. Since this is a preliminary test, we recruited users from our social network who are not native English speakers but have undergone formal education in English. Given their educational background, users were exposed to 10 pre-selected English words having good difficulty level. None of the users were familiar with these words nor were these words used in their day to day conversations. Of the 10 words, revision reminders were set for only 5 words. Instead of scheduling reminders for them, the researchers simulated the scheduler by calling the subjects and requesting them to call into the system at the intervals of 1 hr, 5 hrs and 24 hrs from the time they accessed the system the first time. During each revision call, the users were quizzed by the system for meanings of words they had learnt. They were prompted to select the correct option for the word meaning from multiple choices. After the last revision call, the users were quizzed on all the 10 words – 5 for which revision calls were scheduled and 5 for which no revision call were done. Words which were revised were recalled with 93.3% accuracy while the words without revision were recalled with 55.5 % accuracy.

These preliminary results are encouraging and suggest that the classroom effect of regular prodding by teacher on a subject can probably be simulated through a push based mechanism over the phone. This motivates further investigation. Given the limiting nature of this study, an extended user study is being planned to obtain statistically significant results, with our target user group i.e. non-English speaking users with little access to web based learning tools. Although an incoming revision call can be intrusive, but if the learner is comfortable with it, such push mechanism can instill discipline and ensure continuity of learning which is missing from traditional web based learning tools that rely on pull model. Similar model of structured pull cum push based learning is now also being explored and deployed in the web world through the concept of massive open online courses (MOOC) [17].

6 Conclusion

This paper presented Vimprint, a system for exploring telephony voice applications as an alternative means for enabling effective learning and memorization for underprivileged in developing countries where numerous people have no or limited access to internet. It employs a push based interaction over outbound calls in addition to initial pull from the learners, to assist users in memorizing and to keep them disciplined and engaged in the learning process. The paper presented design and implementation of its Vocabulary building module along with the results of a preliminary study that encourages us to explore further with underprivileged users. While we tested the system for learning English, Vimprint itself is language agnostic and can be used to offer vocabulary building for other languages supported by Nuance Speech recognition engine.

Going forward, we are extending this application with additional learning modules and conduct a full scale study towards establishing telephony voice applications as a potential vehicle for structured learning for the underprivileged. For that purpose, we have tied up with an English learning school in the area who take everyday classes for students who may not have had the luxury to study in expensive English medium schools. With this user set, we intend to compare how Vimprint's vocabulary building module fares against classroom based vocabulary building methods.

References

1. Stats, I.W.: World Internet Users and Population Stats, `http://www.internetworldstats.com/stats.htm`
2. Learn English - enableM and OnMobile, `http://www.enablem.com/learn-english-mobilephone.aspx`
3. Boyera, S., et al.: Mobile Web for Social Development Roadmap W3C IG Note (November 2009), `http://www.w3.org/TR/mw4d-roadmap/`
4. Kachru, B.: Indian English: A sociolinguistic profile of a transplanted language. Studies in Language Learning 1(2) (1976)
5. Kam, M., Mathur, A., Kumar, A., Canny, J.: Designing digital games for rural children: a study of traditional village games in india. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 31–40. ACM (2009)
6. Kotkar, P., Thies, W., Amarasinghe, S.: An Audio Wiki for Publishing User-Generated Content in the Developing World. In: HCI for Community and International Development (Workshop at CHI 2008), Florence, Italy (April 2008)
7. Kumar, A., Agarwal, S., Manwani, P.: The Spoken Web Application Framework: User Generated Content and Service Creation through Low-end Mobiles. In: Proc. of W4A (2010)
8. Kumar, A., Reddy, P., Tewari, A., Agrawal, R., Kam, M.: Improving literacy in developing countries using speech recognition-supported games on mobile devices. In: Proc. of CHI (2012)
9. Patel, N., Chittamuru, D., Jain, A., Dave, P., Parikh, T.S.: Avaaj Otalo - A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In: Proceedings of ACM Conference on Human Factors in Computing Systems, CHI (2010)

10. Pimsleur, P.: A memory schedule. The Modern Language Journal 51(2), 73–75 (1967)
11. Padman, R., Li, X., Damir, J., Kumar, A., Khera, P., Kumar, A., Agarwal, S., Nanavati, A.: Engineering a Spoken Web enabled Health Information System (SW-HIS) for Diabetes Management. In: Proceedings of Mayo Clinic Conference on System Engineering and Operations Research in Health Care (2011)
12. Ruohonen, M., Turunen, M., Mahajan, G., Linna, J., Kumar, V., Das, H.: Mobile Phones and Voice-Based Educational Services in Rural India: Project Ruralvoice. In: Ley, T., Ruohonen, M., Laanpere, M., Tatnall, A. (eds.) OST 2012. IFIP AICT, vol. 395, pp. 3–11. Springer, Heidelberg (2013)
13. Saran, M., Supporting, S.G.: foreign language vocabulary learning through multimedia messages via mobile phones. H. U. Journal of Education 38, 252–256 (2010)
14. English Seekho - Tata Indicom, http://www.tataindicom.com/tata-zone-mobilelearning.aspx
15. Thornton, P., Houser, C.: Learning on the Move: Vocabulary Study via Email and Mobile Phone SMS. In: Proc. of World Conference on Educational Multimedia, Hypermedia and Telecommunications EDMEDIA (2001)
16. White, J., Duggirala, M., Kummamuru, K., Srivastava, S.: Designing a voice-based employment exchange for rural India. In: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development (2012)
17. Carey, K.: Into the Future With MOOC's. Chronicle of Higher Education, http://chronicle.com/article/Into-the-Future-With-MOOCs/134080/ (retrieved March 20, 2013)
18. Larson, M., Rajput, N., Singh, A., Srivastava, S.: I want to be Sachin Tendulkar! A Spoken English Cricket Game for Rural Students. In: International Conference on Computer Supported Cooperative Work and Social Computing (CSCW), TX (February 2013)
19. Tata Sky Active English, http://www.tatasky.com/education/actve-english.html

Improving Evaluation Honesty and User Experience in E-learning by Increasing Evaluation Cost and Social Presence

Juha Leino and Tomi Heimonen

University of Tampere, School of Information Sciences,
Kalevantie 4, 33014 University of Tampere, Finland
{juha.leino,tomi.heimonen}@uta.fi

Abstract. While various recommender approaches are increasingly considered in e-learning, lack of studies of actual use is hindering the development. For several years, we have used non-algorithmic recommender features on an undergraduate course website to help students find pertinent study materials. As students earn credit from adding and evaluating materials, some have chosen to evaluate materials dishonesty, i.e. without actually reading them. To improve honesty, in 2012 we coupled 5-star ratings with commenting (previously uncoupled) to increase the cost and complexity of evaluating and gave students individual presence with nicknames (previously anonymous) to increase social presence and enable reputation formation. Our results show that high enough cost of evaluating together with high enough social presence can lead to complete honesty in evaluations and enhance both user experience and student involvement. In effect, designing such e-learning systems includes not only designing the features but also their use, as the two are intertwined.

Keywords: e-learning, recommenders, ratings, social presence, honesty, design.

1 Introduction

Today, plenty of potential study materials can be found on the Internet, ranging from expert columns and scientific papers to video presentations. However, the material quality varies greatly. Also, even if a material is of high quality, it may be too advanced or too elementary for a student at a specific stage of his or her studies [1].

The authors teach an undergraduate-level course on user-centered design (UCD) at a local university. Given that a large majority of college/university students already use online resources to augment course materials and profess readiness to share them [2], the first author (course lecturer) programmed in 2007 the first version of LSRM (Lecture Slides and Reading Materials) for the course website (requires login) and has continued to develop it based on system use and student feedback.

LSRM harnesses the collective intelligence and efforts of the course student community by allowing them to add additional reading materials to complement the materials added by the instructors (the lecturer and the teaching assistant (TA), the

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 597–615, 2013.

second author). In addition, over the years LSRM has provided students with various non-algorithmic recommending features for evaluating the materials as a community of peers to allow the community to guide its members to the most pertinent materials. In a sense, the system functions as a repository of high-quality material links that are annotated with recommending features (tags, ratings, and comments).

The purpose of LSRM is to encourage students to read more widely on UCD to augment learning and to develop a habit of following the field. UCD is a large topic, and it is practically impossible to cover all of its facets exhaustively on a course. LSRM gives students an opportunity to read more on the facets that they find interesting instead of forcing all to read the same materials. Reading these additional materials is voluntary in the sense that students are not examined on them but since 2009, online activity has affected the course grade. Since 2010, students have been required to add two materials and evaluate five to earn full credit for online activity.

A further goal is to help students develop information literacy, i.e. the skills to locate, select, evaluate, and use information from various sources [3]. Information literacy is considered a survival skill in today's information intensive working life and a foundation for life-long learning [3]. By selecting items from the Internet and then getting feedback on them from the evaluations by others, and by evaluating materials others add, students get to hone their information literacy skills.

The design of LSRM has from the start been driven by such contextual factors as short period of use (one semester) and low number of students (below 60 students) and items (below 150 materials). This has rendered such algorithmic approaches as collaborative filtering (CF) impractical [4]. Instead, LSRM has used recommender features that provide value to the community from the first contribution and that allow students to see their contribution to the community immediately. This is grounded on the notion that making visible to users the value of their contributions to the group has positive social outcomes [5]. In 2011, LSRM employed tagging, 5-star rating and commenting, all of which make the contribution immediately visible and useful/usable.

In the past, the system has been plagued by dishonest ratings, i.e. students evaluating materials without even viewing them. While not common enough to cripple the system, as discussed in [6], perceptions of other students not doing their work properly has harmed the user experience. In 2011, the previously used binary rating feature (*Yes* and *No* buttons to respond to the question *Did you find this material useful?*) was replaced with a 5-star rating feature to increase evaluation cost in order to reduce dishonesty. The change almost halved the percentage of dishonest ratings but at the cost of the overall number of ratings falling.

In 2012, evaluation cost was further increased by coupling rating and commenting into a unified evaluation (evaluation title, 5-star rating, and text justification for the rating). The goal was to further increase honesty and to improve the user experience by increasing trust on evaluations, in part by making the thinking behind the rating transparent. Moreover, while in 2011 the system use had been anonymous, in 2012 nicknames were employed to give students individual presence in the system in order to increase social presence—and thus social pressure to add pertinent materials and evaluations—and to enhance the user experience through sociality.

Based on log data, student questionnaire replies, and student interviews, both measures were successful. There were no dishonest evaluations in 2012 and the perceived social presence increased significantly, resulting in positive behavioral changes and enhanced user experience. Also, students reported having had learning and information literacy benefits from the system to a much larger extent than previously.

In addition to discussing how increasing social presence and cost of evaluations can be used to enhance user experience and significantly reduce dishonesty in e-learning, our study contributes to the field by providing a view of actual use by authentic users of a system employing recommender features. While various recommender approaches are today widely considered in e-learning, there is currently a dire need of actual experiences of such systems to guide the development efforts [7].

2 Related Work

Recommender systems (RS) help us deal with large numbers of items in two somewhat overlapping ways, by helping us find salient items (e.g. books we might be interested in) and by helping us make decisions (e.g. which book to buy) [4,8]. RS consist of one or more recommending features, varying from such non-algorithmic approaches as reviews and ratings to heavily algorithmic prediction-computing [4,6].

Recently, the potential of RS has attracted increasing interest in the e-learning research community (e.g. [2,8,9]). However, various domain-specific differences make transferring RS from one domain to another challenging [10], e.g. user interest is not the only determining factor in e-learning, as pedagogical aspects are also an important consideration [1,8]. Significantly, learners recognize this and are also ready to read uninteresting materials that are important for learning [1].

The ability of a recommender system to establish trust with its users is recognized as crucial to the system's success [11]. For this reason, RS have enjoyed more success in low-risk domains; users lack confidence to act on recommendations in high-risk domains [12]. Speculations and examples of dishonest users skewing recommendations have not helped in establish trust in RS [12,13]. In e-commerce the goal for dishonesty is typically to distort recommendations favorably or unfavorably for an item, often for financial profit, whereas in e-learning dishonesty has other motives, such as getting credit without earning it [6,14].

While having users explicitly rate items is a common approach to gathering user preferences, there are few guidelines to selecting rating scales, despite the fact that some scales tend to produce higher and some lower ratings on the same item [9,15]. In e-commerce, contributions appear to come mainly from highly opinionated users, resulting in an unrepresentative sample of user views [13]. Hu et al. [16] suggest that when benefits are not clear, user motivations for contributing can be explained with a *brag-and-moan* model. Similarly, it has been suggested that strongly negative or positive consumption experiences may lead to *expressing positive emotions* or *venting negative feelings* [17]. Consequently, while ratings and reviews have become an important information source in e-commerce, their ability to reflect actual item quality has been questioned [13,16]. Still, altruism and concern for others also appear to be important motivations for contributing in e-commerce [17].

Social presence has been shown to have a positive impact on number of contributions and user behavior [18,19]. However, while RS can provide social texture that can lead to perceptions of others being present, different users require different interface cues to perceive a system as having social presence [18].

While accuracy metrics are important, there is a growing recognition that user experience is, in fact, more decisive [12,20]. RS need to be not only useful and accurate but also pleasurable to use [20]. Consequently, when evaluating recommenders, evaluating user experience is essential [12]. In e-learning, measuring satisfaction is important, as it is closely related to motivation [10]. The problem for researchers is that measuring user experience requires "field studies with long-term users of the system … measuring behavior in a natural context" [12]. Finding or building up and maintaining user communities for research is not easy while commercial systems tend to guard their trade secrets jealously [12]. This has resulted in a dire need for case studies of actual users in a real use context to guide employing RS in e-learning [7].

3 Study Setting and Data Collecting

The 2012 UCD course was lectured in fall semester, and consisted of seven 2-hour lectures (Sept.) and fourteen 2-hour practice sessions (Sept.–early Dec.). The grade was based on design assignment (70%), ten smaller assignments (20%), and online work (full credit required adding two materials and evaluating five) (10%) plus extra 10% for high attendance. In 2012, the student community consisted of 36 students (18 female) while in 2011 there were 37 students (8 female).

In 2012, twenty students (56%), of whom 12 were females, filled out a questionnaire on LSRM. Movie tickets were raffled among the respondents. The questionnaire consisted of five sections: 1) Materials and adding them, 2) Evaluating materials, 3) Tagging materials, 4) Tools, and 5) Other aspects (e.g. social presence). Each section contained evaluative statements using a 7-point interval scale (1= strongly disagree; 7=strongly agree) and open-ended questions. Virtually all evaluative statements allowed commenting, which students frequently did. In 2011, 19 students (51%) filled out the questionnaire that consisted of six sections, as 5-star ratings and commenting had not been coupled and so the questions about them were in separate sections.

The 2012 students who filled out the questionnaire were on average more diligent and motivated than their peers (Table 1). However, the two groups were not that different when it came to viewing materials. Importantly, one student who added no materials or evaluations filled out the questionnaire, thus giving us a view into the thinking of these students, too. Students who filled out the questionnaire represent well the majority of students (81%) who made at least the required number of contributions. In fact, removing respondents who made no contributions (7) eliminates the differences between the groups. The trends for 2011 are similar.

Table 1. Students who filled out the questionnaire vs. students who did not (2012)

	Grade	Motivation	Added materials	Added evaluations	Materials viewed
Students who filled in the survey (avg.)	4.3	4.55	2.6	4.95	22.7
Students who did not (avg.)	3	3.81	1.25	3.13	12.44
Statistical significance of difference	Yes	Yes	Yes	Yes	No
p-value (unpaired t-test)	.0001	.0041	.0012	.0071	.0756

In addition, we interviewed three students (8%) in 2012 to gain deeper understanding of student motivations and views concerning the system and its use. The second author conducted the semi-structured interviews that lasted 30–45 minutes per student, as the first author had designed and built LSRM, something that the students were aware of. When a quote is from an interview, we mention it; otherwise, quotes are from questionnaire replies. In 2011, no interviews were conducted.

As the system collected virtually click-by-clack data of individualized student activity on the LSRM page, we are able to contrast *saying* (questionnaire and interview data) with *doing* (actual activity data), thus reducing the potential of say-do problems.

4 LSRM System Description

The LSRM system was implemented with PHP, JavaScript, and HTML. As most interactive parts were implemented with AJAX (Asynchronous JavaScript and XML), most interactions took place within the current use context. Changing views and navigating by tags, however, reloaded the page. Clicking a material link opened the link in a new browser window.

The LSRM interface in 2012 consisted of a web page that gave different views to the material available (Figure 1). The interface was originally in Finnish, and has been translated into English for relevant parts for this paper.

Because we often compare the student behavior and perceptions in 2012 to those in 2011 to see how the changes in the system affected them, we also discuss to some extent the 2011 design. Moreover, the 2011 design and experiences are relevant because the changes in the system are largely based on student feedback on it.

In 2011, there were no separate views except for navigating with tags. All materials were added to and listed under lectures (most recent lecture on top). Also, while in 2011 all actions had been anonymous, in 2012 all material additions, evaluations, and comments were identified by student (or instructor) nickname.

In 2011, the system allowed students to add materials, rate (5-star scale) materials others had added, comment (*Title* and *Text* fields) all materials (also their own to enable discussing), and tag all items. Rating and commenting were decoupled. In 2012, rating and commenting were coupled into an evaluation (evaluation title, star

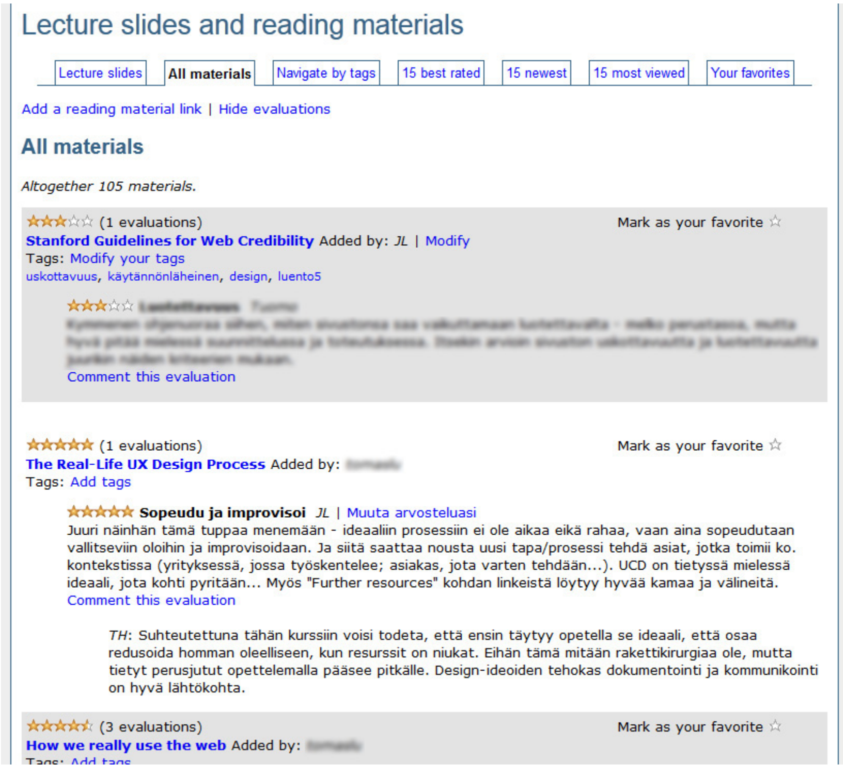


Fig. 1. LSRM interface in 2012—the content is authentic (student nicknames and texts blurred) but the order of the materials has been adjusted for illustrative purposes

rating on 5-star scale, and text explanation for the rating). Stars were given at the precision of full stars while star averages were displayed at the precision of half-stars. Evaluating one's own material addition was impossible. To enable discussing, evaluations could be commented (voluntary in the sense that commenting did not affect grade).

While in 2011 comments were hidden by default and had to be opened for viewing for each material to keep the page length at bay, in 2012 evaluations and comments were displayed by default as a result of student feedback. Tagging materials was also possible, and unlike in 2011, students were asked to tag the materials they added.

4.1 Working Hypothesis behind the 2012 LSRM Design

We based the current design of LSRM on working hypothesis grounded on our experiences and student feedback from the earlier versions of the system. Accordingly, we decided to 1) use nicknames instead of anonymity to increase sociality and trust on both evaluations and materials, and 2) increase the cost and complexity of evaluating materials by coupling five-star rating with a title and comment to increase trust on evaluations and perceived quality of evaluations. Furthermore, with materials getting evaluated more comprehensively, we expected the pressure on students to add good materials to increase and result, in turn, in increased perceived quality of materials.

4.2 LSRM Tool Usability

According to questionnaire responses, the tools LSRM offered were considered easy to use (M=5.60; SD=.82). In 2011, the ease-of-use was evaluated slightly lower (M=5.06; SD=1.16) but the difference is not statistically significant. Consequently, we conclude that usability problems were few and did not significantly color the student perceptions. Also, there were no year-on-year differences in tool usability.

5 Results

The number of material additions and evaluations followed closely what was required for the full credit. Seven students evaluated no materials, 25 students evaluated the required five materials, and four evaluated six (M=4.14; SD=2.09). The lecturer evaluated three. Likewise, eight students added no materials, 17 added the required two, seven added three, three added four, and one added five (M=2.00; SD=1.31). The instructors added 36 materials. The number of materials added and evaluations made correlated strongly, $r(103)=.71$, $p=.01$, meaning that students who contributed in one way also contributed in the other. Only two students commented evaluations made by others (each once) while the instructors commented five.

The average number of honest evaluations per student almost doubled from 2.09 (SD=2.65) in 2011 to 4.14 (SD=2.86) in 2012. The difference is statistically significant, $t(71)=2.28$, $p=.026$. However, this did not come at the cost of activity; the average numbers of evaluations were not statistically different from 2011. Thus, increasing evaluating complexity did not reduce the number of evaluations in 2012.

In 2012, students viewed on average 18.1 materials (range=0–70, SD=17.25). This represents a clear and statistically significant increase from 2011 at $t(71)=3.61$, $p=.001$ when students on average viewed only 7.1 materials (range=0–29; SD=7.06).

5.1 Additional Reading Materials

Students perceived being able to add additional reading materials positively both in 2012 (M=5.70; SD=1.17) and 2011 (M=5.11, SD=.99). Student comments also indicate that the majority of students saw the feature very positively. They especially appreciated how the materials complemented and added to the lectures, giving more information on the topics that interested them and covering a wider spectrum of topics than possible in the lectures: "*You found materials that had not been discussed in the lectures. …you had an opportunity to read on topics you were interested. Also, you could share with others interesting and useful materials that had inspired you.*" Students also emphasized that nobody can find all the good materials alone. Several students also mentioned that having to add materials instilled a good habit of following the field.

When viewed numerically, students perceived the added materials in 2012 much as in 2011. In 2012, students rated the quality of the materials at 5.40 (SD=.60) and in 2011, at 5.21 (SD=.79). The difference is not statistically significant. However, other differences support the idea that students viewed the materials more positively in 2012. While in 2011, some students suspected that others had added materials without selecting them carefully, in 2012 only one student questioned the quality of materials. Most comments saw material quality in positive terms: "*The materials covered a very wide spectrum … It felt that the people who added materials had really wanted to add the articles and not just find quickly 2 articles….*"

In fact, many students reported having read numerous articles to find links that they felt were worth adding in the sense of being "*genuinely interesting and useful*" to other students: "*I wanted to find as high-quality materials as possible, materials that I felt had taught me something so that others also could learn from them.*" Many students professed altruistic motives and wanted others to benefit from their work: "*Even though, as far as I know, the evaluations that the added links got didn't affect the grade, I didn't feel it proper to add a material that I would not have reviewed positively myself.*" At the same time, at least some students were acutely aware that others would evaluate the links they add: "*Of course, there also was something of a 'social pressure' since I knew that others would later evaluate my materials….*"

When selecting links to add, students considered what was useful and relevant for oneself—"*But the good thing was that I had to learn to find articles that are pertinent for me*"—and others: "*It taught me to search for information on a topic and be critical towards it.*" Students felt that selecting items "*forced me to read the materials with care and at the same time think about their good and bad aspects. …it helped learning.*" As one student put it nicely, "*…adding materials in and of itself was beneficial because looking for a suitable article made you reflect on the concepts taught on the course and appraise the relevance of the articles according to them.*" In effect, many student comments show that they had to work on their information literacy skills, learning to locate, select and evaluate materials [3], and that they were aware of learning these skills. These aspects were much less discussed in 2011 comments.

However, not all students went through the trouble of finding pertinent links to add. Two students admitted having taken the easy way out: "*…adding materials was 'compulsory chore'…. I thought that all I've got to do is find some relatively sensible piece of text that has something to do with the course….*" While getting all the students to work hard on finding good materials is probably impossible, using social presence and high-cost evaluations appears to have resulted in most students taking adding materials seriously and thus reaping benefits from it.

5.2 Selecting Materials to View in LSRM

Students viewed on average 18.1 materials in 2012, slightly more than 2.5 times the average in 2011 (M=7.1). Instructor-added materials were on average viewed 5.56

times (SD=2.63) and student-added ones 6.67 times (SD=3.57). The difference is not statistically significant. The most important selecting criterion was the content of the link, the "*interestingness*" of the article. There were two sources for judging this, 1) the title of the link—"*Interesting topics, interesting title in particular. The title should tell what it's all about.*"—and 2) the evaluations: "*I browsed materials and their evaluations to form an impression … and decided based on the impressions of the interestingness of the topic and the quality with which the topic was covered.*"

Also, both the existence and number of evaluations functioned as important heuristics "*A material with evaluations stood out from the mass and I was more likely to check them out.*" The number of evaluations was a heuristic for interestingness of the link for many students quite independent of their valence. Being part of the communal activity appears to be part of this: "*If the evaluations by others repeatedly referred to some aspect [in the article] or brought up an interesting thing, you also wanted to read it. It wasn't just the good evaluation that affected; I also read 'worse' ones.*" In fact, selecting already evaluated materials had social aspects: "*Evaluated materials appeared more attractive…. The good thing about them was that you get to compare your viewpoint to that of the evaluator.*" This is something that making commenting a compulsory part of evaluations enabled. Interestingly, some students also mentioned having viewed unevaluated materials on purpose so that their evaluations would be useful to others, bringing an altruistic aspect to sociality.

In 2012, students judged the impact of evaluations on which materials they read to be on average 5.05 (SD=1.73). While the 2011 average was lower (M=4.00; SD=1.97), the difference is not statistically significant. Still, in 2012, students tended more towards feeling that evaluations had an impact, as 75% rated the effect at 5-7 while in 2011 only 53% did so. In 2012, the question concerned evaluations while in 2011 only ratings. Since there was no statistical difference between the averages of star ratings given in 2012 and 2011 at $t(294)=1.17$, $p=.24$, we conjecture that coupling the star-rating with a comment is the main reason for proportionally higher ratings of effect.

We identified three facets of evaluations that had an appreciable impact on the process of selecting a material for further investigation: The star rating averages, evaluation valence, and the number of evaluations. While star rating averages clearly affected the equation—rating averages correlated positively with the number of viewers, $r(103)=.52$, $p=.01$—some students felt that star ratings did not necessarily tell much while the textual reviews did: "*I don't trust the star ratings, as everybody has their own rating scales in their heads. Comments and reasonings, on the other hand, give you hints about why the article might be worth reading.*" Perhaps consequently, star averages became more important as a selection heuristic towards the end of the course: "*At the beginning there weren't so many materials, so you could read them all or at least glance at them. … Towards the end I browsed the ones with most stars and opened the link if the evaluations aroused my interest sufficiently.*"

The valence of the evaluation also affected the equation for many: "*I selected for reading materials when the title aroused my interest. The ratings and reviews affected if I opened an article or not. If somebody had commented that rather superficial and* *circumspect, then no need to think twice if I opened it or not :)*" In turn, praising evaluations attracted viewings: "*I selected materials based on topic… The evaluations also influenced what I selected; if the article was highly praised, I took a look.*"

However, according to student comments, the existence and the number of evaluations was at least as significant a selection criterion as valence. In fact, the correlation between the number of ratings and the number of students viewing the item is higher at $r(103)=.64$, $p=.01$ than between the average of evaluations for the materials and the number of students viewing it.

That a material ended up being opened did not, naturally enough, mean that it was carefully read. Opening the material was only a step in the process. The next step was to glance at the material, to decide if it really was worth reading. If the article passed this impression-forming glance and the students started to read it, the judging process continued: "*I only bothered to read the article to the end if it gave me some new information or a new viewpoint, i.e. I felt it to be useful.*"

5.3 Evaluating Materials

In 2012, evaluating a material involved an in-depth evaluation (title, star-rating, and text justification) while in 2011, compulsory evaluating consisted of simply giving a star rating while commenting was voluntary. Students viewed the possibility to evaluate materials positively, on average at 5.55 (SD=1.15) in 2012 and at 5.11 (SD=1.59) in 2011. The difference is not statistically significant; increasing the cost of evaluating did not reduce the positivity of student views. As reasons for liking evaluations, students mentioned social aspects and possibility for expressing opinions in addition to evaluations helping in selecting materials for reading.

Many students especially liked the text part of the evaluation: "*…the possibility of textual evaluation was very good, and it was also good that you could comment evaluations! It even led to some discussing.*" However, one student mentioned a negative aspect concerning the text comment: "*You saw what others had praised so you read it with interest, too, but I myself didn't like to evaluate. Especially since I'm not a professional or somebody who knows a lot, so my comments may have appeared pretty bad for somebody who knew more.*" Two interviewed students touched the same theme, noting that it was easier to comment in LSRM than on the Internet because there are so much more knowledgeable people on the Internet. For them, a smaller community with the members more or less at the same level of knowledge made it easier to comment. Consequently, a small, closed community of peers can create a safer environment to encourage participation. This finding is in line with the idea that the sense of community is connected to a feeling of *membership* that includes boundaries that provide members with emotional safety [21].

Star-ratings were seen in a more problematic light: "*Comments enrich and give new viewpoints, stimulate discussion. Star-ratings I found somewhat unnecessary.*" One reason for disliking star-ratings was the difficulty in deciding the appropriate ratings: "*Occasionally it was hard to give stars because even if the material was really useful for me, it's not necessarily that for everybody so you don't want to rate it too highly, either.*" The problem was exacerbated by the fact that students were not told which specific aspect to rate. This was a clear design mistake; the interface should have made the rated aspect clear.

Although students in 2012 viewed on average slightly over 2.5 times more materials than in 2011, the number of evaluations on average was statistically the same in 2012 (M=4.14, SD=2.09) as in 2011 (M=3.89, SD=2.75). Also, on average materials had statistically as many ratings/evaluations in 2012 (M=1.45, SD=1.54) as in 2011 (M=1.8, SD=1.96). In effect, in 2012, students therefore read more materials but evaluated fewer in relation to the number read than in 2011. However, significantly, on average students made more *honest* evaluations, i.e. viewed the material before evaluating it, in 2012 (M=4.14, SD=2.09) than in 2011 (2.86, SD=2.65). The difference is statistically significant, $t(71)=2.28$, $p=.026$.

In fact, in 2012, not a single dishonest evaluation was made: The use log data shows unequivocally that on each occasion, the student had opened the link before evaluating it. Increasing social presence and the cost and complexity of evaluations removed dishonest attempts to get points without earning them entirely. Also, since the number of evaluations per student did not decrease, increasing the cost of evaluating did not reduce the number of evaluations. In effect, more students were motivated to do the required work when the 2012 design was used.

Not only did increasing cost and complexity of evaluations in comparison to ratings result in complete honesty but it also resulted in perceptions of honesty. Not a single student mentioned suspecting dishonesty in the 2012 feedback while such suspicion was entertained in the 2011 feedback (when dishonest rating in fact took place). While social presence likely played a significant role in this, too, given that changing the rating scale from binary (2010) to 5-star (2011) reduced dishonest rating almost by half (from 43% in 2010 to 26% in 2011), we conjecture that needing to write a textual justification for the rating made cheating simply too difficult: "*You had to read the materials to be able to evaluate it.*" Since student perceived others as doing the required work, they also ended up reciprocating by doing their own share.

The care with which materials were read before they were evaluated increased clearly. When we look at the time periods that passed between opening the link and adding the evaluation (when evaluating took place within the same session, as it typically did) for honest ratings/evaluations, we notice that the reading times almost doubled in 2012. If we examine the reading times that were shorter than 15 minutes (to filter out sessions that may have included other activities), there is still a clear statistical difference between 2012 (M=409 seconds, SD=242.33) and 2011 (M=231 seconds, SD=203.94), $t(164)=5.133$, $p < .001$. The materials were clearly read longer in 2012 before they were evaluated. In fact, in 2012, only in two cases (1.8%) did a student evaluate a material after reading it for less than a full minute while in 2011, there were 21 (23.3%) such cases. Still, having to write a textual evaluation in addition to clicking a rating must also have been a partial reason for the increased time between open a link and evaluating it.

In 2012, only four students (21%) said that they had rated all the materials they had read while in 2011, eight students (42%) said the same. In both years, the main reason for not rating a viewed material was the same; students felt that they had not read the material carefully enough to rate it.

While in 2011 some students did refer to social factors as a rationale for not evaluating, it was in 2012 that social aspects were mentioned repeatedly in this context. Sociality inherent to the system clearly affected evaluating behavior: "*There were social aspects to evaluating, so I did not want to write an evaluation that just said 'nice one' or 'interesting article' but something more. For this reason I wanted to evaluate only articles on which I had a clear opinion and something a bit deeper to say—something that might inspire others to comments and something that others could comment.*" Besides a certain social pressure, there also was a sense of moral duty towards others: *"…I would have felt wrong about evaluating a material that I had not read entirely.*" In effect, as with material additions, certain altruistic motivations were evident in many student comments concerning evaluations.

Also, a few students mentioned that they did not evaluate some materials they had read "*because I had nothing new/significant to add to the comments by others.*" While no student in 2011 mentioned thinking twice about rating a rated item, coupling ratings and comments made students feel that they had to have something significant to say, something that had not already been said: "*Somebody else might have already said what was essential in his or her comment.*"

Another reason for students not to evaluate a material they had read in 2012 was that "*the materials had not aroused any big emotion*." Mediocre, bland articles simply did not garner evaluations: "*I evaluated materials based on whether they stirred up thoughts or not. I selected for evaluating only materials on which I had some kind of an opinion. In the evaluating phase I simply skipped lackluster articles altogether.*" If students did not have something to say about the material, they did not evaluate it.

Consequently, students read articles based on personal interests and need. If they read the whole article and felt that it was "*useful*" and "*interesting*" and they had something to say about it (that somebody else had not already said), they probably evaluated the material. As a result, students largely ended up evaluating good articles: "*I didn't really bother to read materials that I found worthless with the first glance, so I ended up choosing for evaluation only good materials.*"

Usefulness was the most important criterion for students when they evaluated materials. However, how clearly written and presented and how illustrative the article was also affected the evaluation. Moreover, students appreciated learning something new from the material. "*Usefulness and practicality, can I use the material in future in studies and at work. Also if I learned something new and if the materials was relevant to the course and its content.*"

The above factors largely explain why over 50% of the star-ratings were four stars and 70% 4–5 stars in 2012. There is no statistical difference between the average star-ratings in 2012 (M=3.82, SD=.83) and 2011 (M=3.70, SD=.86), $t(294)=1.17$, $p=.244$ (Table 2).

Table 2. Distributions of star ratings in 2012 and 2011

	1 star	**2 stars**	**3 stars**	**4 stars**	**5 stars**
2011	0 (0%)	13 (9%)	42 (29%)	64 (44%)	25 (17%)
2012	4 (3%)	1 (1%)	41 (27%)	79 (52%)	27 (18%)

In 2012, three students gave 1-star evaluations (two once and one twice). Two materials ended up having one star as the average of its ratings (both had two 1-star evaluations). Based on the evaluation texts, it appears that the students giving the 1-star evaluations felt that the materials should not have been added to the system. Interestingly, it was only when nicknames were used that 1-star ratings were made; in 2011, no 1-star ratings were made. We conjecture that a heightened sense of social presence/sociality resulted in people showing disapproval for substandard materials. Most appeared to have communal, even altruistic motivations, and they probably expected the same in return from the other members of the community. Thus, while students mostly ended up selecting good materials to evaluate, when the experience was strongly negative, they were ready to *vent negative feeling* [17], even if the comments connected to 1-star evaluations were still quite polite and matter-of-fact in tone.

Several student comments underline that students were aware of the benefit they got from reading materials and that they understood that evaluating materials did make them to read them more carefully: "*evaluating materials involved reading a lot of materials when 5 evaluations were required. 5 felt a lot but afterward I felt it was useful that it made me read so many articles.*" Students were also clearly aware of the information literacy benefits that evaluating materials and reading evaluations by others brought: *"Evaluating materials increased the teaching value of the articles. By reading the evaluations by others we got good feedback on how to apply scientific texts in studying. Also, finding and reading articles in our field is very important, especially for working life.*" This contrast with the 2011 comments where students focused on the ability of rating to guide them to better materials and to warn them against bad ones but did not discuss much the benefits of reading and evaluating.

Another social aspect of evaluations was curiosity of how others evaluated the materials one had added: "*Of course I checked out what kinds of evaluations the materials I had added had gotten.*" One interviewed student in fact mentioned she had just before coming to the interview (interviews took place after the course) checked if there had been any new evaluations on her materials. Another interviewed student said that the system gave him a feeling that the materials he had added were useful to others and that one of them had had "*a fair amount*" of evaluations. What others said clearly mattered to and interested students.

5.4 Perceived Social Presence and Its Impact

While contributions had previously been anonymous, in 2012 students were asked to choose a nickname for LSRM when registering to the course. The purpose was to increase perceived social presence and to allow reputation formation by giving students individual presence in the system. In effect, there was a significant effect for social presence, $t(36)=2.06$, $p=.047$, with students reporting on average higher social presence in 2012 (M=3.53, SD=1.22) than in 2011 (M=2.75, SD=1.15).

Given that we were using a 7-point scale, average perceived social presence of 3.53 might not seem very high. However, LSRM is in a sense competing against such popular social systems as Facebook and instant messaging when it comes to how social students perceive it. In that sense, the seemingly low average of 3.53 may in fact actually indicate a relatively high perceived social presence for a system that does not support real-time presence. In fact, some comments from students who professed not to have felt others as present emphasized the slow rhythm of interaction: "*I didn't find much [social presence] because there wasn't that much activity and I didn't get announcements of e.g. that my materials had been evaluated etc. The thought of there being other students affected so that I wondered what others thought about my evaluations….*" One interviewed student encapsulated the general perception by saying that the feeling of sociality was clear but not very strong and that the level was appropriate to the system, as it was about as much as can be archived "*unnaturally*."

Student comments indicate that the perceived social presence had a clear and positive impact on behavior: "*Sociality in the service affected my actions significantly: It affected so that I wanted to select as suitable articles as possible to add to the page and that I wanted to say something more deep than just 'quite nice' in the evaluation.*" Repeatedly, students mention having tried to find materials that would be useful to others and making evaluations that would help others. While there were similar trends also in 2011, altruistic aspects are more emphasized in the 2012 comments.

Some comments clearly connected nicknames to reputation: "*When adding materials I thought that I can't add just any odd stuff … because all the other students see them. The fact that my name was connected to the materials and evaluations I added also made me think twice what to say.*" Students appear to have felt that through the nickname they had an individual presence and reputation in the system, and that affected their behavior positively. Having an individual presence in the systems also made students consider their self-image in relation to the community, as one interviewed students explained: "*Also, building my self-image influenced it; I didn't feel it satisfactory for myself to put there something that I wouldn't want to read myself.*" The student comments indicate that most wanted to be responsible members of the community, and in this sense, the achieved social presence was high enough.

While some students professed not having felt the presence of others, their comments show that their actions were nevertheless influenced by awareness of others, e.g. "*I didn't really feel others to be present that much. Still, the thought that others see what I add there affected what materials I added and what kind of evaluations I made. I.e., I did my job with care.*" The impact of this awareness on student behavior appears to have been larger than the numeric evaluation of social presence indicates.

Social presence had both activating and experience-enhancing impact: "*…the presence of others there activated me, too, and it was great to see that others actually read and evaluated materials.*" In particular, student comments underline that perceiving others as present improved the user experience, e.g. "*The presence of other students affected positively because [that way] you knew that somebody else is also reading these comments and not just the teacher alone.*"

The positive effects of increased social presence appear at least partially attributable to increased social pressure that drove students to do more than just bare minimum for earning points: "*[I read the materials I added] very carefully indeed exactly because of the sociality connected to evaluating materials. There was some 'social pressure' involved in evaluating articles because other students could read your evaluations, respond to them, disagree and comment, respond to the evaluation…*"

The effects of social presence are likely intertwined with the effects that the emerging sense of community had. Besides membership, using LSRM also had many other elements that McMillan and Chives [21] suggest as contributing to the sense of community, including *personal investment* (added evaluations and materials), bi-directional *influence* (students affecting the community and vice versa) and *integration and fulfillment of needs*.

5.5 Social Presence of Instructors and Its Impact

Students felt that the lecturer's presence had a positive impact in two ways. First, he maintained a feeling of activity in the system, e.g. "*A lot of added materials and evaluations on materials [by the lecturer]. I feel that it was good that the lecturer kept the page active when it occasionally got silent.*" Second, the lecturer's presence brought positive social pressure: "*The presence of the lecturer did encourage investing in the materials. I didn't have the nerve to add just any old dude's blog there and added instead content by recognized sites or known experts.*"

Some students mentioned having been nervous about evaluating materials added by the lecturer. In fact, student-added materials (M=1.88, SD=1.54) did get on average more evaluations that instructor-added materials (M=.61, SD=1.18), $t(103)=4.34$, $p < .001$. Interestingly, there was no statistically significant difference between the average star ratings given to instructor-added materials (M=3.95, SD=.72) and student-added materials (M=3.79, SD=.84). The scale, in a sense, was the same, meaning that the 16 students who evaluated instructor-added materials (in contrast, 30 students evaluated student-added materials) did not give the materials special treatment.

Comments and evaluations from the lecturer were warmly welcomed: "*I especially liked how the lecturer commented on some evaluations and gave his own examples (e.g. on a material that I had added).*" The evaluations and comments that the instructors added were mainly positive or, in one case, only mildly challenging.

6 Discussion

Using Nicknames Increases Sociality and Trust on Materials and Evaluations. Using nicknames clearly increased social presence, as evidenced by the statistically significant increase in student evaluations and student questionnaire replies. Even the students who did not feel others as present described how the idea of other users affected their behavior positively. Many students reported altruistic motivations, and their comments show that they felt a certain sense of duty towards others. Also, students had a sense of individual presence in the system, which created social pressure that also affected their actions positively. Overall, students perceived others as taking online activity seriously, and this motivated them to approach it with due diligence.

Increasing Evaluation Cost Engenders Trust on Evaluations. Students clearly trusted better that student evaluations were properly made than in 2011. How much of this is attributable to nicknames and how much to increased evaluating cost is an open question, but considering that in 2011, increasing the rating cost resulted in a significant improvement in honesty, we conjecture that coupling commenting and rating affected it significantly and also encouraged reading materials more carefully.

Requiring More thorough Evaluations Increases Pressure to Add Good Materials. There was less questioning of the motives of the students adding materials in 2012, indicating that at least some aspects of the perceived quality had improved. Also, many student comments show that students did approach finding materials to add very seriously, in part because they knew that they would be evaluated. We conclude that the more complex evaluations made students more careful about the links they added but that the effect is again intertwined with the effects of the nickname use.

Overall, our working hypothesis concerning nicknames and evaluation cost worked out well. Using nicknames and a more complex evaluating approach removed dishonesty entirely from evaluations. Importantly, this was accomplished without the number of contributions falling; in fact, the number of honest contributions increased. This is a significant improvement to the system and gives other e-learning practitioners practical tools and approaches to root out dishonesty.

6.1 A Word of Caution: Use and Interface Design Are Intertwined

When applying our results in e-learning, and particularly in other contexts, it should be noted that our results are subject to specific contextual factors. Our system is designed for formal e-learning where compulsoriness can be used to encourage contributions, students form a small, closed community of peers, and the use period is short. In informal e-learning, for instance, using compulsoriness may not be possible or even advisable. Also, designing an e-learning space must go hand in hand with designing its use (compulsoriness, regulations etc.), as the two are intertwined. For example, our system used compulsoriness and high evaluation cost to improve evaluation/rating honesty. However, if there had been no compulsoriness, this design would likely have failed. With voluntary evaluating, it would have been advisable to lower the evaluation cost to encourage contributing; after all, there would have been little motivation for dishonesty. In 2011, while there were dishonest ratings (compulsory), there was no dishonest commenting (voluntary).

6.2 Enhanced User Experience

The 2012 use and interface designs led to students approaching their work more honestly and with more gusto. The space was significantly more social due to the tools bringing sociality and providing social texture and the nicknames providing sociality through individual presence. Student trust on materials and especially evaluations increased; students saw others as doing their work properly, which led to altruistic motivations and a sense of duty towards others, leading to deeper involvement. In effect, this perception was justified by actual changes in honesty and due diligence, as most students did work hard to add links that were meaningful and tried to make evaluations that would be useful. Also, in 2012, many students were more aware of the benefits—in particular, improved learning and information literacy skills—they accrued from using the system with due diligence. Seeing benefits in turn encouraged using the system, leading to a virtuous circle: Positive behavior led to positive experience and perception that in turn encouraged positive behavior.

6.3 How to Develop LSRM Further

Further increases in evaluation cost do not appear necessary, as 100% honesty was already reached with the current approach. Nevertheless, the aspect to be rated needs to be made clear in the interface so that all students are rating the same aspect. The most important criterion for students, *usefulness* of the materials, is the obvious candidate. While increasing the cost of adding materials by requiring a description might seem a logical step to induce further trust on materials, it may prove problematic; in an interview, one student said that this would be a "*miserable feature*" that would only lead to marketing one's materials instead of objectively describing them.

In fact, the most promising approach to encourage positive behavior and improving the user experience further appears to be enhancing sociality and individual presence in the system. First, the system needs to be more connected to students' everyday lives. LSRM should make it possible for students to subscribe to email notices so that they can maintain awareness of any development in the system that concerns them. In addition, the system should incorporate a private group in Twitter and Facebook, to mention two obvious candidates. This way, the information about new materials and evaluations would reach students without them having to log in the system. Making groups private is important, as the community being small, closed, and consisting of peers (same level of knowledge) were important factors for students.

The second approach to increasing sociality is to enhance the sense of individual presence by allowing viewing material additions, evaluations and comments by individual students. This would also increase social pressure, as one could not hope to be hidden in the mass of materials. Interviews gave indications that students would not find this intrusive. In all likelihood, this would further increase the care with which students add and evaluate materials.

7 Conclusion

We managed to enhance students' user experience and remove dishonesty from additional reading material evaluations by increasing the evaluation cost and by replacing anonymity with nicknames, thus giving students an individual presence.

This study also contributes to the field by providing much-needed experiences of using recommender features in e-learning in a genuine use context.

However, when applying our results in other systems and contexts, one should bear in mind that the results were obtained within formal e-learning context and that, consequently, contextual factors may limit their applicability elsewhere.

While we hope to develop LSRM further to the directions outlined here, we also encourage other practitioners to report their experiences of RS in various educational contexts. It is important that practice and theory go hand in hand in employing RS in e-learning instead of theories being developed independent of the ground realities.

References

1. Tang, T., McCalla, G.: Beyond Learners' Interest: Personalized Paper Recommendation Based on Their Pedagogical Features for an e-Learning System. In: Zhang, C., Guesgen, H.W., Yeap, W.-K. (eds.) PRICAI 2004. LNCS (LNAI), vol. 3157, pp. 301–310. Springer, Heidelberg (2004)
2. Hage, H., Aïmeur, E.: Harnessing Learner's Collective Intelligence: A Web 2.0 Approach to E-learning. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 438–447. Springer, Heidelberg (2008)
3. Kiliç-Çakmak, E.: Learning strategies and motivational factors predicting information literacy self-efficacy of e-learners. AJET 26(2), 192–208 (2010)
4. Schafer, J.B., Konstan, J., Riedl, J.: Recommender Systems in E-commerce. In: Proc. the 1st ACM Conf. on Electronic Commerce, pp. 158–166. ACM, New York (1999)
5. Rashid, A.M., Ling, K., Tassone, R., Resnick, P., Kraut, R., Riedl, J.: Motivating Participation by Displaying the Value of Contribution. In: Proc. the SIGCHI Conf. on Human Factors in Computing Systems, pp. 955–958. ACM, New York (2006)
6. Leino, J.: Case Study: Material Additions, Ratings, and Comments in a Course Setting. In: Santos, O.C., Boticario, J.G. (eds.) Educational Recommender Systems and Technologies: Practices and Challenges, pp. 258–280. IGI Global (2011)
7. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., Koper, R.: Recommender Systems in Technology Enhanced Learning. In: Kantor, P., Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 387–415. Springer, Heidelberg (2011)
8. Santos, O.C., Boticario, J.G.: Modeling Recommendations for the Educational Domain. Procedia Computer Science 1(2), 2793–2800 (2010)
9. Ghauth, K.I., Abdullah, N.A.: Learning Materials Recommendation Using Good Learners Ratings and Content-based Filtering. Journal of Educational Technology Research and Development 58(6), 711–727 (2010)
10. Drachsler, H., Hummel, H.G.K., Koper, R.: Identifying the Goal, User model and Conditions of Recommender Systems for Formal and Informal Learning. Journal of Digital Information 10(2), 4–24 (2009)
11. Pu, P., Chen, L.: Trust-inspiring Explanation Interfaces for Recommender Systems. Knowledge-Based Systems 20, 542–556 (2007)
12. Konstan, J., Riedl, J.: Recommender Systems: From Algorithms to User Experience. User Model. User-Adap. Inter. 22, 101–123 (2012)

13. Talwar, A., Jurca, R., Faltings, B.: Understanding User Behavior in Online Feedback Reporting. In: Proc. the 8th ACM Conf. on Electronic Commerce, pp. 134–142. ACM, New York (2007)
14. Lam, S.K., Riedl, J.: Shilling Recommender Systems for Fun and Profit. In: Proc. the 13th Int. Conf. on World Wide Web, pp. 393–402. ACM, New York (2004)
15. Sparling, E.I., Sen, S.: Rating: How Difficult is It? In: Proc. the 5th ACM Conf. on Recommender Systems, pp. 149–156. ACM, New York (2011)
16. Hu, N., Pavlou, P., Zhang, J.: Can Online Reviews Reveal a Product's True Quality? In: Proc. the 7th ACM Conf. on Electronic Commerce, pp. 324–330. ACM, New York (2006)
17. 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)
18. Leino, J., Räihä, K.-J.: User Experiences and Impressions of Recommenders in Complex Information Environments. IEEE Data Engineering Bulletin 31(2), 32–39 (2008)
19. Nov, O., Ye, C.: Why Do People Tag? Communications of ACM 53(7), 128–131 (2010)
20. McNee, S.M., Riedl, J., Konstan, J.: Being Accurate Is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 1097–1101. ACM, New York (2006)
21. McMillan, D.W., Chavis, D.M.: Sense of Community: A Definition and Theory. Journal of Community Psychology 14(1), 6–23 (1986)

Interacting with Augmented Reality: How Does Location-Based AR Enhance Learning?

Ruobing Li[1,*] , Bo Zhang[1], S. Shyam Sundar[1] and Henry Been-Lirn Duh[2]

[1] Media Effects Research Laboratory, College of Communications, The Pennsylvania State University, University Park, PA 16802
{rul148,buz114,sss12}@psu.edu
[2] Human Interface Technology Lab Australia, University of Tasmania & Keio-NUS CUTE Center, NUS
duhbl@acm.org

Abstract. Augmented Reality (AR) can provide additional information about mediated events, but can it enhance our learning and comprehension? We performed a user study of a location-based AR application in order to answer this question. A 2-condition (AR modality vs. non-AR digital book modality) between-subjects experiment with 36 dyads of secondary school students in Singapore was conducted to examine how the use of AR modality in an educational context impacts students' learning performance. Data from the experiment showed that location-based AR improved students' learning performance by catching their attention and enhancing their ability to elaboratively process the information they encountered. Theoretical and practical implications are discussed.

Keywords: Location-based AR, modality, learning, transportation.

1 Introduction

Augmented Reality (AR) development is booming today, providing a contextual layer of extra information (e.g., images, locations, sound, etc.) to add on to user perceptions of the physical environment. Numerous applications have adopted the AR technology into their design, such as in the fields of gaming, entertainment, tourism, marketing, and social networking. However, as the industry is busy developing new apps to improve their design, little is known about the psychological effects of AR, and related user-experience outcomes. For example, AR is being touted as a promising tool for enhancing education, but we do not know how the use of an AR app compared with other digital educational tools can impact students' learning performance. Therefore, our study investigates the impact of an AR game, "the Jackson Plan," on students' learning performance, and attempts to explore the psychological factors that may mediate the relationship.

* Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 616–623, 2013.

1.1 Augmented Reality

Augmented Reality (AR) is a new modality of presenting information, whereby a contextual layer of computer-generated virtual content is superimposed on a mediated representation of the real physical world. AR blends the real and virtual worlds seamlessly, constantly changing the ways we experience and interact with the physical environment. With the arrival of mobile media, this technology is on the verge of being adopted ubiquitously in our daily lives to enhance what we see and feel. In 2010 alone, AR mobile applications were downloaded more than eleven million times. By 2015, the number of downloads is estimated to exceed one billion per year [1]. Among these applications, for example, AR has been designed as a tourist guide on mobile devices [2], a shopping assistant in e-commerce websites [3], and a way to promote learning in an educational environment. According to the 2011 Horizon Report, AR technology will be widely employed in US universities in the next two to three years, as a way to assist teaching inside or even outside the classroom [4]. These educational AR applications include AR books, AR gaming, object modeling, skill training, etc., which aim to enhance students' learning.

1.2 Psychological Effects of AR

Due to its unique characteristics, AR may have a psychological effect on users in various ways. The most representative and appealing feature of AR is its ability to superimpose extra information into the real world based on sophisticated display technologies, which, on the one hand, increases the viewability of the interface and contextualizes the information, but, on the other hand, runs the risk of demanding too much user attention to encode the supplemental information. Limited Capacity Theory [5] would argue that cognitive resources directed towards encoding the message come at the cost of storing the contents of the message for later retrieval, thereby resulting in poorer memory, and therefore inferior learning outcomes. While AR certainly runs this risk of privileging encoding over storage, many AR games and apps are tied to real-world locations, allowing users to view the world through a web camera on their devices with additional information superimposed by the app, which may provide users a sense of presence, or physically being in the environment [6], thus leveraging their engagement with the content (i.e., mental "transportation" into the narrative [7]), eventually enhancing users' cognitive information processing and memory. Under this alternative view, AR will serve to enhance the user's "perceptual bandwidth" [8] such that the user is able to process more information more efficiently and thereby achieve an enhanced mental representation of mediated reality. Given these potential psychological paths to user outcomes, how AR can influence users' learning performance is unclear. Therefore, we experimentally examine how the use of AR modality in an educational context impacts students' learning performance as well as their user experience.

2 Method

2.1 Design and Participants

A 2-condition, between-subjects (location-based AR modality vs. non-AR digital book modality) experiment was conducted to answer the research questions. Participants were recruited from 3 Secondary One history classes in a public school in Singapore. The students were selected by a teacher to ensure social and cognitive homogeneity. In all, we had 72 students participating in the study. Since the procedure required students to work in groups of two, we had a total of 36 dyads for random assignment to the two experimental conditions. None of the students had learned about the historical event described in the application, which was a history class chapter for their grade, before the study, and all of them had some experience of playing games on handheld devices.

2.2 Stimulus Material

Two applications containing the same interactive game were developed for this study, one of which contained AR features while the other one was a digital book. The content presented by the two different modalities was identical. An Apple iPad 2 was used in both conditions as the primary device via which they experienced the educational content. The content followed a narrative structure. In the story, “The Jackson Plan” map, which was the original city plan map of Singapore, was stolen, and the participants were asked to locate the missing map by talking to different original immigrants (non-player characters or NPCs, see Figure 1) and collecting clues from their conversations. The story was created based on information about the founding of modern Singapore and the life conditions of its original immigrants. As the original immigrants’ trading center, Singapore River was chosen as the physical corresponding location for the location-based application. While students were walking, a map of Singapore River (see Figure 2) would show up on the iPad screen, on which students’ current location and their next destination were indicated. There were 6 destinations in all, representing the starting point, ancient trading spots of British, Chinese, Indians and Malays, and the final destination. Once the student arrived at a spot, the location-based service would automatically trigger a NPC’s appearance on the screen, and then the NPC started to talk to the student. Besides, four mini-games were embedded in the narrative story to help students understand and learn different immigrants’ specific trading activities. Similarly, the mini-games were also triggered by the GPS. AR features were embedded in several steps. For example, when participants arrived at the destination, which was the statue of Raffles, they needed to capture information on the statue using the iPad to trigger a newspaper page on the iPad screen indicating who had stolen the map. While for the digital book application, students just needed to tap the iPad screen to proceed from screen to screen and get the newspaper page at the end, thus featuring no interactions with objects in the real world.

2.3 Procedure

Each dyad of participants was randomly assigned to one of the two conditions. In the experimental condition, students were taken to the Singapore River to complete their session (see Figure 3). A moderator was assigned to accompany each dyad of students. After briefly introducing the game, the moderator started the game and gave each dyad an iPad. The 18 dyads of participants in the control condition completed their session in a classroom in their school. After completing the game, each participant was asked to complete two paper-and-pencil questionnaires. Then, researchers bought each participant a meal, and asked them not to discuss the experiment with others.

Fig. 1. Non-player character

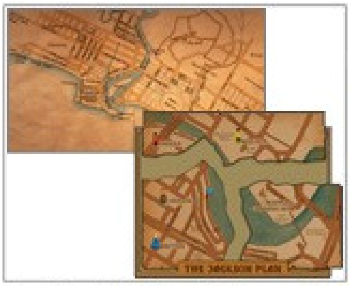


Fig. 2. Real artifact and game maps

Fig. 3. Outdoor AR players vs. indoor non-AR players

2.4 Measurement of Learning and User Experience (UX)

Two questionnaires were used to measure participants' learning and user experience. One questionnaire contained 20 multiple-choice questions testing their ability to recognize factual information and knowledge embedded in the game, such as information about the history of modern Singapore's founding and the trading activities of original immigrants. Another questionnaire contained 40 items which participants rated using a 5-point Likert-scale ranging from 1 (not at all) to 5 (very much). Learning motivation was measured by 11 items, two of which pertained to intrinsic goal orientation (r=0.60, e.g., the most satisfying thing for me in playing "The Jackson Plan" is trying to understand the historical content as thoroughly as possible), 5 referred to task value (α=0.84, e.g., I think learning history in playing "The Jackson Plan" is useful for me to learn), and 4 to self-efficacy (α=0.86, e.g., I'm confident I can understand the most complex materials presented by "The Jackson

Plan"). Learning strategy was measured by 7 items, among which elaboration was measured by 4 items (α=0.85, e.g., when playing "The Jackson Plan", I try to relate the material to what I already know) and peer learning was measured by 3 items (α=0.75, e.g., when studying for history, I will try to explain the material to a classmate or a friend). In addition, game engagement was measured by 7 items (α=0.80, e.g., I feel my mind wandering when I am playing "The Jackson Plan"). Telepresence was measured by 3 items (α=0.80, e.g., when I am playing "The Jackson Plan", I have a sense of "being there"). Perceived novelty of the application was measured by 5 items (α=0.77, e.g., the way in which "The Jackson Plan" was presented is novel). Transportation was measured by 7 items (α=0.84, e.g., I was mentally involved in the story while playing "The Jackson Plan"). Aside from these self-report measures, the applications in both conditions automatically recorded the time that participants spent on the application.

3 Results

We first conducted a MANOVA by considering all the dependent variables. The results showed a significant effect for modality, Wilks' Λ=.64, F (10, 61)=3.45, p<.01. Univariate results confirmed a main effect of modality on intrinsic goal orientation, F (1, 70)=7.20, p<.01, task value, F (1, 70)=9.74, p<.01, self-efficacy, F (1, 70)=5.57, p <.05, elaboration, F (1, 70)=11.50, p<.01, telepresence, F (1, 70)=11.00, p<.01, transportation, F (1, 70)= 8.91, p<.01, and learning performance, F (1, 70)=10.76, p< .01. On all variables, participants in the location-based AR condition scored higher than their counterparts in the digital book condition.

Since participants in the location-based AR condition spent much longer time (M=29.51, SD=4.73, counted in minutes) to complete the whole game than their counterparts in the digital book condition (M=9.40, SD=2.78), the analyses were repeated with time as a covariate. In addition, perceived novelty was also controlled to rule out its effect. Results of ANCOVA analyses showed that, with time and perceived novelty controlled, participants in the AR condition still reported higher level of task-value (F (1, 58)=4.06, p<.05), more self-efficacy towards learning history from the application (F (1, 58)=4.66, p<.05), more elaboration (F (1, 58)=3.95, p=.05), more engagement (F (1, 58)=5.03, p<.05) while experiencing the game, and showed better learning performance than those in the non-AR condition (F (1, 58)=3.96, p<.05).

Multivariate regression was performed to test the relationships between measured variables. The results showed that with time and perceived novelty controlled, participants' amount of elaboration during playing the game was significantly correlated with their leaning performance (b=.85, t (51)=2.04, p<.05). Similarly, participants' amount of elaboration (b=-.35, t (52)=-2.00, p<.05), engagement level (b=.24, t (52)=2.52, p<.05), and telepresence (b=.38, t (52)=3.31, p<.01) were significantly correlated with their transportation level.

A series of mediation tests using the PROCESS Macro [9] was conducted to explore the mechanism underlying the effect of AR on students' learning performance and transportation level. The results showed that the modality influenced participants' learning performance through a three-step mediation model. The two modalities led to different levels of engagement, and participants' engagement further influenced their amount of elaboration, which in turn affected their learning performance, *b*=.32, 95% C.I. from .02 to 1.06, *SE*=.24 (see Figure 4). The modalities influenced participants' transportation level through two paths: firstly, engagement level of participants mediated modality's impact on transportation, *b*=.28, 95% C.I. from .01 to .92, *SE*=.21; secondly, modality influenced participants' engagement, and their engagement level influenced telepresence, which then in turn influenced the transportation, *b*=.30, 95% C.I. from .02 to .88, *SE*=.20 (see Figure 5). However, transportation did not predict participants' learning performance. Nor did it mediate modality's effect on learning.

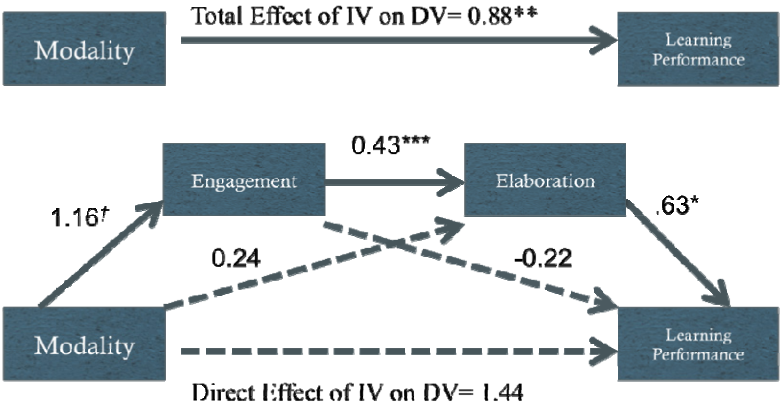


Fig. 4. Regression coefficients (†p < .10, *p < .05, **p < .01, *** p < .001; solid lines indicate a significant indirect path). IV refers to independent variable and DV refers to dependent variable, which are modality and learning performance respectively in this model. The modality variable had two values; with 0 representing digital book condition and 1 representing the location-based AR condition.

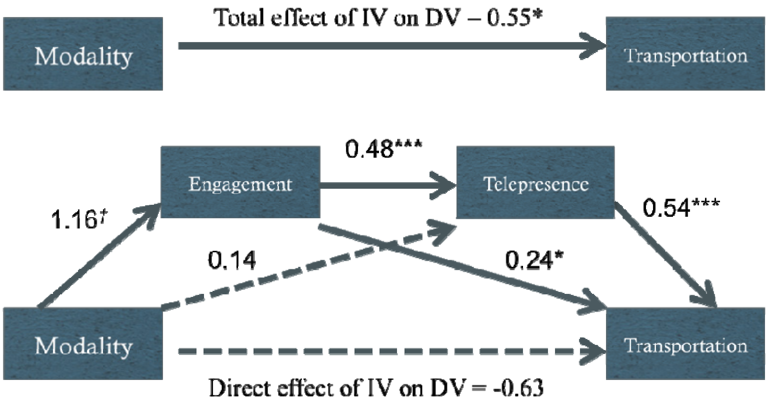


Fig. 5. Regression coefficients (†p < .10, *p < .05, *** p < .001; solid lines indicate a significant indirect path). IV and DV are modality and transportation respectively in this model.

In summary, this study found that location-based modality could enhance students' learning performance by making them engaged in the task and facilitating them to elaborate on the information they encountered. Similarly, location-based modality was also more likely to transport users away from the physical world and make them immersed in the virtual narrative world. Users' enhanced engagement and telepresence mediated this process.

4 Discussion

Our findings show that location-based AR modality improves users' learning performance by enhancing their engagement and ability to elaboratively learn information from the application. The elaboration variable mainly referred to students' ability to put together information from different sources (e.g., narrative, mini games) and make connections between new information and their prior knowledge, history and the game event, virtual event and real world activities. The present study provides practical evidence that location-based AR modality, by enabling users to augment the real world environment with virtual features, generates greater engagement with the content, thereby improving learning performance. Therefore, AR holds promise a potentially useful avenue for educators to fully engage students and achieve desirable learning outcomes.

Our study also found that location-based AR modality had an effect on users' transportation level by leveraging their level of engagement and strengthening their feeling of presence in the virtual world. Since transportation is an important determinant of narrative's persuasive effect on audience, AR can serve a persuasive function [10]. Digital content designers who want users to be immersed in their content should try to use location-based AR modality to present the content, so as to provide users a feeling of actually being in the narrative world. This can be particularly useful for digital story-telling and other forms of informational, as well as entertainment, narratives delivered via devices that can support AR. It can also help marketers persuade users about their products and services by affording greater transportation into the narratives constructed.

A common concern with any new technology is that the technology's effect may be caused by the novelty effect, which will fade away as the technology becomes common. However, in our study, participants in the digital book condition reported slightly higher perceived novelty (M=3.54, SD=.76) of the application than their counterparts in the location-based AR condition (M=3.51, SD=.61). Therefore, we can rule out this alternative explanation in this study.

A limitation of this study is that although students worked in pairs, their data were collected and analyzed individually, without controlling for their dyad membership. Due to a technical failure in the study protocol, information about which two participants worked together was not captured by study administrators. Another, more important, limitation of this study is that while participants in the experimental condition completed their session in the outdoor environment, those in the control condition completed all the procedures in a classroom. Therefore, one may argue that it was the actual historical site, but not the AR technology that made the experimental stimulus outperform the stimulus of control condition. However, based on several factors, we believe this alternative explanation is unlikely. First, the Singapore River is totally different from what it looked like 200 years ago. The area used to be a trading center, but now all the old buildings were replaced or used as modern restaurants, so study participants were not able to get any relevant information about the history of Singapore or the immigrants' trading activities by simply being in the environment. Therefore, it is safe to argue that the real-world environment is not likely to influence students' learning performance. Besides, in terms of the transportation level, the real world environment is actually more likely to distract students from concentrating on and immersed in the narrative. Students had to walk for several minutes before they could continue to get new information from the application, so the whole story was not continuous for them. In addition, while they were walking along the river, a lot of students were attracted by lobsters and crabs in some restaurants' aquariums. Each of these two situations could decrease outdoor students' transportation compared to a classroom, but they still reported much higher level of transportation. Therefore, the location-based AR modality appears to be quite powerful in inducing transportation into the narrative despite the presence of formidable environmental distractions. This bodes well for the future of AR apps in the domain of education, especially in distracting out-of-school environments, because the technology itself appears to be quite involving and therefore conducive to learning.

References

1. Juniper Research Mobile Augmented Reality (2011), http://juniperresearch.com/reports/mobile_augmented_reality
2. El Choubassi, M., Nestares, O., Wu, Y., Kozintsev, I., Haussecker, H.: An Augmented Reality Tourist Guide on Your Mobile Devices. In: Boll, S., Tian, Q., Zhang, L., Zhang, Z., Chen, Y.-P.P. (eds.) MMM 2010. LNCS, vol. 5916, pp. 588–602. Springer, Heidelberg (2010)
3. Zhu, W., Owen, C., Li, H., Lee, J.-H.: Personalized In-Store E-Commerce with PromoPad: An Augmented Reality Shopping Assistant. Electronic Journal for E-commerce Tools and Applications 1(3), 1–19 (2004)
4. New Media Consortium (NMC) The 2011 Horizontal Report (2011), http://wp.nmc.org/horizon2011/
5. Lang, A.: The limited capacity model of mediated message processing. Journal of Communication 50(1), 46–70 (2000)
6. Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. Journal of Computer-Mediated Communication 3 (1997), http://www.ascusc.org/jcmc/vol3/issue2/lombard.html (retrieved June 14, 2002)
7. Green, M.C., Brock, T.C.: The role of transportation in the persuasiveness of public narratives. Journal of Personality and Social Psychology 79(5), 701–721 (2000)
8. Reeves, B., Nass, C.: Perceptual user interfaces: perceptual bandwidth. Communications of the ACM 43(3), 65–70 (2000)
9. Hayes, A.F.: PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling (White paper) (2012), http://www.afhayes.com/ (retrieved)
10. Green, M.C., Brock, T.C.: The Role of Transportation in the Persuasiveness of Public Narratives. Journal of Personality and Social Psychology 79, 701–721 (2000)

The ReflecTable: Bridging the Gap between Theory and Practice in Design Education

Jonathan Hook[1], Thomas Hjermitslev[2], Ole Sejer Iversen[2], and Patrick Olivier[1]

[1] Culture Lab, Newcastle University, Newcastle upon Tyne, United Kingdom
[2] Centre for Advanced Visualisation and Interaction, Aarhus, Denmark
{jonathan.hook,patrick.olivier}@ncl.ac.uk
{thomashj,oiversen}@cs.au.dk

Abstract. The ReflecTable is a digital learning environment that explores how design games and video-led reflection might be combined to bridge the gap between the theoretical and practical components of design education. The concept seeks to leverage the qualities of exploratory design games and video to inspire design students to critically reflect upon the relationship between their evolving design practices and the theories and techniques they are taught in lectures, by allowing them to capture, review and reflect upon short videos of a design game. In this paper, we present the ReflecTable design and nine studies conducted during the course of its development. The studies suggest that the ReflecTable has the potential to support design students in understanding how theoretical concepts and methods relate to different design situations and their own evolving design practices.

Keywords: Design, design games, video-led reflection, education, reflection, reflective practicum, off-loop reflection, learning-by-doing.

1 Introduction

Educating students in the skilled and creative practice of design is a challenging task. Typically, the situations and challenges designers face are multifaceted, ill defined and, therefore, not easily addressed by the straightforward application of theories or guidelines [27, 29]. Consequently, it is widely accepted that approaches to educating designers that focus upon the teaching of theoretical concepts alone are unlikely to equip students with the skills required to respond appropriately to the challenges faced by design practitioners. Instead, design educators characteristically seek to train their students to dynamically adapt and employ a repertoire of theories, techniques and skills in response to their reflective appreciation of an individual design situation; a competency that Donald Schön refers to as a designer's "artistry" [31, p. 13].

The artistry of design practice is not something that can be easily taught, as it hinges on the student developing an understanding of how the particular theories and techniques that they have been shown in lectures and tutorials relate to their own personal experiences of designing in response to a range of different situations and challenges. Purely theoretical education has been found to be ineffectual at facilitating

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 624–641, 2013.

this kind of learning, as novice designers do not have the necessary experience of design practice in order to develop these contextual and holistic understandings of what they are taught [31, p. 82]. However, the alternative, a focus on practical education through design projects, has also been found to fall short, in our own and others' experiences. While learning-by-doing allows students to develop personal experience of design practice, students are often given little opportunity or motive to reflect upon the relationship between the theories and techniques they are taught and the situations they encounter during their design projects [37].

Fig. 1. The ReflecTable

Our response to this challenge, the ReflecTable (Figure 1), is a digital learning environment that seeks to support students in reflecting upon a condensed design process, which is completed while playing a group design game around a tabletop. Students playing the ReflecTable game are given a number of small buttons that can be pressed to capture short video clips of noteworthy moments in their design process. Following the completion of this game, the students are guided through a session of structured reflection on their design process, which involves reviewing and annotating the video clips captured using a bespoke tablet PC application. By leveraging qualities of both design games and video-led reflection, the ReflecTable seeks to provide design students with an environment that combines learning by doing with critical reflection about the relationship between design practice and theory.

In the remainder of this paper, we review previous work relating to the use of digital technology to support design practice and education, describe the design of the ReflecTable in detail and, finally, reflect upon our initial experiences of using the ReflecTable as part of the education of undergraduate design students. These experiences illustrate the design's potential as a means to support students in reflecting upon and, therefore, developing further understandings of the relationship between their evolving design practices and the design vocabulary and procedures (i.e. theoretical components) that they are taught as part of their education.

2 Related Work

2.1 Interactive Systems for Design Practice and Education

The use of digital technology to support design practice and education can be seen to be an important concern of the HCI research community. A range of previous research has led to the development of a number of novel interactive systems that support both the practices of designers and the education of design students.

Several interfaces have been developed by HCI researchers that seek to enhance the existing practices of designers through the application of digital technology. A number of hybrid interactive surfaces [20] have been proposed that aim to seamlessly integrate the workflows and physical artifacts of the traditional design studio with advantages of digital technology. These interfaces have used approaches such as computer vision [17, 21], tangibles and digital pen and paper technology [14, 15] to digitally augment the physical artifacts used during traditional design meetings (e.g. post-it notes) with functionality such as saving and versioning and support for the transition to later, digital, stages of the design process. Additionally, a number of designs have been proposed to afford lightweight integration of digital technology and digital media into design meetings [33, 34] and the configurations of artifacts and space witnessed in the designer's traditional studio [1].

HCI researchers have also explored the development of interfaces that enhance collaboration and awareness amongst design teams. The Cooperative Artifact Memory system was developed to allow designers to share digital information about artifacts from a design process with their collaborators such as annotations, messages and web-links using a mobile phone application [38]. Blevis et al. developed an interface that aimed to scaffold collaborative decision making in respect to both physical and digital design materials, which were associated together using barcodes [2]. Additionally, work by Everitt et al. explored how the use of hybrid interactive surfaces (such as those mentioned previously) might support collaboration amongst designers in different geographic locations [10].

Of most relevance to the ReflecTable, however, are previous research projects that have explored how digital technologies might support designers, and design students, in documenting and reflecting on their design processes and practices. Nakakoji et al. proposed a tool that indexes videos of design meetings based upon sketches drawn on a whiteboard. Their system aimed to allow designers joining existing projects and teams to more easily use videos of design meetings to develop understandings of, and reflect upon, design decisions made previously by the existing design team members [25]. Geyer and Reiterer explored the use of the commercial note taking software Evernote to support design students documenting longitudinal design projects [13]. Swan, Tanase and Taylor explored the Digital Scrapbook system, which automatically collects digital materials from, e.g., design students' blogs and Flickr pages [35]. In both of these studies, providing students with a mechanism to collate, document and review design project materials was found to inspire and support reflection on their design practices. In the latter case, the students' reflection on their practices using the Digital Scrapbook system was found to make students aware of the

"Processional" nature of the design processes they conducted. Additionally, in the context of design research [39], Dalsgaard and Halskov developed the Project Reflection Tool, which allows design researchers to document their practices as a series of events and notes, with the aim of supporting reflection on both design practice and broader research questions [8].

A number of research projects have looked beyond the possibility of simply supporting the documentation of design processes, to explore instead how the designer's active manipulation of digital design materials might inspire reflection on design practice. The Amplifying Representational Talkback system aimed to inspire reflection during the design process by allowing designers to experiment with different two-dimensional arrangements of their digital design materials [26]. Additionally, the Freed system enabled design students to create multiple spatial representations of the collections of digital materials that are amassed during longitudinal design projects. This process was found to inspire design students to reflect on their design practices throughout the course of a design project, rather than simply upon its completion [24].

2.2 Using Design Games and Video to Inspire Reflection

In the following sections, we review a range of additional work on design games and the use of video as a means to inspire reflection, which influenced the ReflecTable design. The idea of playing games to study design practice derives from research in architecture and planning from the late eighties. At MIT, Harbraken and Gross [16] developed Concept Design Games to study how design decisions and choices are made in a process of cooperation and negotiation among multiple stakeholders in interdisciplinary design projects. By identifying similarities between design collaboration and simple board games, Harbraken and Gross envisioned their own board games that isolated and focused on certain aspects of the design process. Games as a method to understand design was further developed in Participatory Design literature as a way to understand user collaboration in design [9], to support design practitioners in their collaborative design processes [18] and to develop a repertoire of gaming possibilities for designers to be aware of when creating their own exploratory design games [3, 4].

To some extent, games have also been used in design education. Buccerelli [6] identified gaming as a way to teach engineering design students about different roles in the design process. Iversen et al. discussed the educational benefits of design games [18]. They envisioned design games as a way to train design collaboration in a "safe" environment. With respect to this safe environment for design team training, design games have similarities with the notion of "Serious Play" developed by LEGO, where LEGO bricks are used to explore real life challenges in business development [28].

The value of design games as a means to support design education has, however, been questioned in recent Participatory Design literature. Törpel critically argued that what students learn from a design game might be limited by the different scales of students' design projects and real life design practice: "What in relatively small education projects can be experienced on a small-scale level could become substantial factors of success or drawback if practiced on a larger scale in professional design projects" [36]. This challenge has been addressed by combining design games with video material from real life design situations. Buur and Søndergaard [7] developed an augmented board game in which video from real life design work was made accessible to the participants. In this augmented space, qualities from the design game were combined with real life design situations captured on video and brought into the design game using augmented paper sticks. Iversen & Buur [19] developed a Video Design Case in which design practitioners could reflect on their own design skills by collaboratively working with video clips from an industrial design case. The Video Design Case made design practitioners reflect on their own practice by mirroring their own design with the design activities displayed in the case material. Iversen & Buur found that video is a valuable approach for fostering reflection on design practice; however, the extent to which designers can familiarize themselves with videos of other designers' practices and cases may limit the reflection that can be spurred by the approach.

The ReflecTable extends this existing body of research by exploring how a novel system for video-led reflection on design practice might be combined with a gaming approach to reflection on design to address the particular challenge of bridging the gap between the theoretical and practical components of design education.

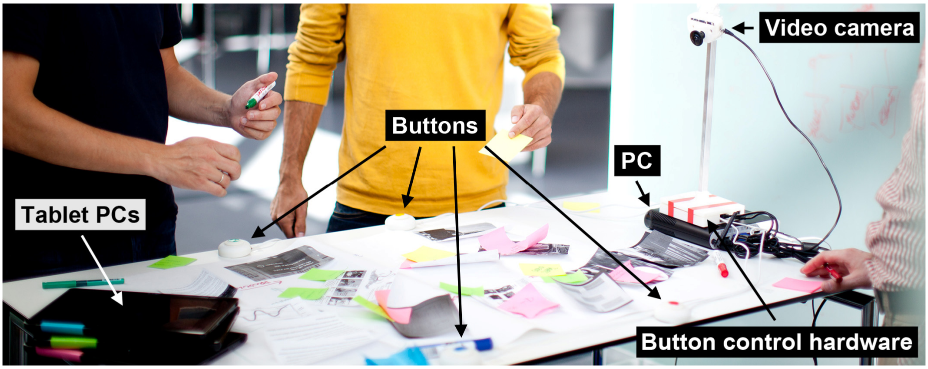


Fig. 2. The key components of the ReflecTable system

3 The ReflecTable

The ReflecTable is a digital learning environment that seeks to leverage qualities of both design games and video as means to inspire designs students to critically reflect upon the relationship between their evolving design practices and the theories and techniques they are taught in lectures. This is done by allowing them to capture, review and reflect upon short videos of their own design process in a design game.

The ReflecTable (Figure 2) comprises three key elements: a paper-based design game that is played around a tabletop between groups of three to four design students, a video capture system that allows students to capture short video clips of noteworthy moments during their game by pressing a number of buttons and a tablet-based interface that allows students to review and reflect upon these videos upon the conclusion of the game.

The design game is comprised of two stages and six rounds (see Table 1 for an overview) and is played around a tabletop between groups of three to four design students. This game moves between divergence, utilizing turn taking inspired by Harbraken and Gross' silent game [17], together with convergence to emulate the phases of a real life design project [22]. In the following sections we describe the course of a typical ReflecTable session.

Table 1. An overview of the rounds of the design game

Round no.	Description
Stage A: Designing using paper, pens and post-its	
1	Diverging by individually annotating artifacts – i.e. pictures, notes and sketches – from their design domain.
2	Converging by drawing and naming relations between artifacts.
3	Transcending by sketching a design proposal based on the work in rounds 1 and 2.
Stage B: Reflecting on the process using video	
4	Browsing video and formulating a research question about the group's design process in stage 1.
5	Diverging by individually annotating video clips on the tablets.
6	Converging by relating and grouping clips on a shared tablet to answer the question from round 4.

3.1 Stage A: Designing on Paper

The aim of the first half of the design game is to co-develop a design in response to a design situation shared by the players (e.g. a brief set during a group project). The stage comprises three rounds, which each mimic a phase in a condensed design process. A game facilitator guides the group throughout the game, offering coaching to the players if, for instance, they become stuck, or need further advice.

During this first half, a video camera is used to record an image of the table surface and a microphone records the discussion between players and the facilitator. Each player is given a button, which can be pressed at any time during the game, to record the previous 15 seconds of video. The players are instructed to press this button when something noteworthy has happened. For instance, a player might press their button when a particularly interesting idea arises or when s/he feels that something is worth remembering.

The group is asked to bring five artifacts (per player) to the game, which must relate to the design situation and be small enough to be placed and manipulated on the tabletop. In our experiments so far, people have brought a wide range of different artifacts including pictures from field studies, vision statements and scenario descriptions.

In the first round, the group is instructed to place their artifacts on the table for individual exploration by each of the players. The players are given post-it notes (a different color for each player) and are asked to write comments about any of the artifacts that they find to be in any way remarkable. At the end of the round, the players present these annotations to each other. The purpose of this round is to allow players to develop divergent interpretations of the artifacts, which will be used as the basis of a shared understanding of the design situation developed in later rounds. Consequently, the players are asked to complete this task without communicating with the other players. The players are given about 10 minutes to complete this round.

In the second round, the players are asked to collaboratively name and draw relationships between the artifacts, which have now been augmented with post-its. The players are instructed to arrange the artifacts and their comments on the table surface, so that they illustrate relationships and shared understandings of the materials and design situation developed by the group. The purpose of this round is to converge the players' individual interpretations of the design situation into a shared understanding, upon which they might base a design. The players are given about 20 minutes to complete this round.

The third round of the game involves transcending the ideas and relations discussed in the previous round, in order to propose a concrete design response. Firstly, the players are asked to collaborate in identifying three potential design openings, i.e., places in the network of relations between artifacts and notes that they feel have the potential to be the starting point for a promising design. Then, the players are given exactly 15 minutes to sketch and write-up (informally) a proposal for a concrete design response to the situation, based upon the ideas developed during the previous rounds.

3.2 Stage B: Structured Reflection on the Design Process

In the second half of the game, a tablet-based interface is used to scaffold reflection amongst the players about the design process conducted during the first stage. Each of the players is given a tablet PC (Acer Iconia Tab W500) with a bespoke interface that allows them to review, annotate and, finally, collate and compare the videos captured by pressing the buttons during the design game. Like the first stage, the second stage is structured around three rounds. The facilitator offers coaching and guidance throughout the three rounds.

In the fourth round, the first round of stage B, the interface displays a Panopticon video rendering, which allows the players to quickly view footage of the entire design game at a glance (Figure 3). This Panopticon rendering works by dividing a video into a number of clips of equal length, which are then arranged in a grid sequentially from left to right, (akin to a cartoon strip). All of these clips are played concurrently in a continuous loop. As they play, the videos slowly cycle towards the right, such that when a video has reached the end of its loop it is in the starting position of the video to its right. This movement of video clips allows the viewer to watch the video continuously, from a selected point, by simply following the passage of one square in the grid of videos.

This Panopticon video is overlaid with a number of color-coded markers, which each highlight a point in the video where a particular player pressed their button in order to indicate a noteworthy event. The players are able to touch these markers to view a larger version of the 15 seconds of video recorded before that particular button press, which is presented, with sound, in a pop-up window. Using this interface, the players are asked to collaboratively review the video clips created during the first stage. While reviewing these video clips, the players are asked to formulate a research question about their design process that they would like to explore during the remainder of the ReflecTable game. In the studies we have conducted so far, players have sought to answer questions such as "what in the process made us change direction?" and "can you see traces of our current design in this workshop?"

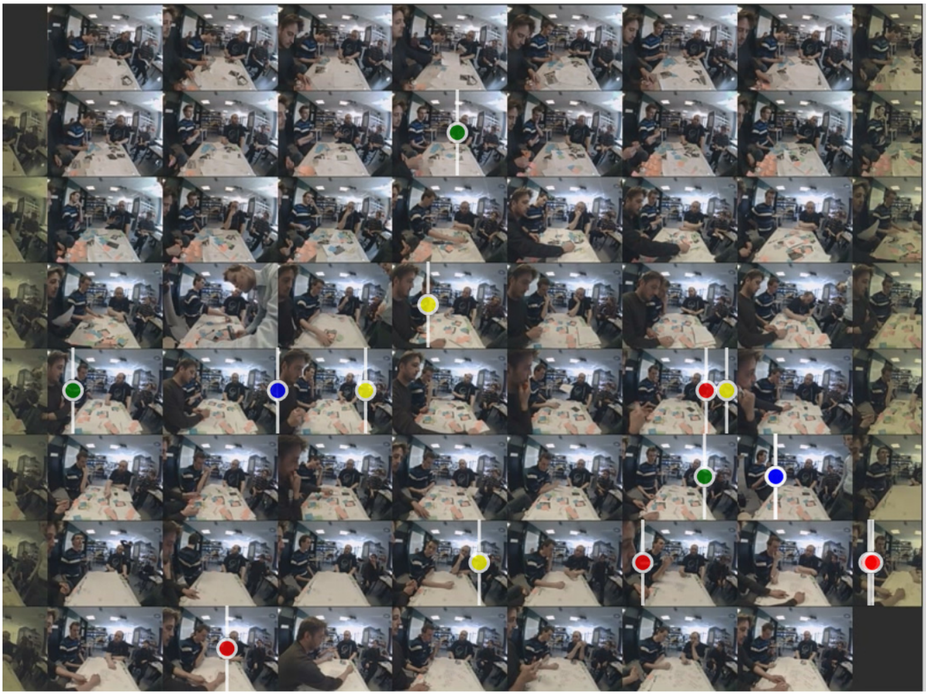


Fig. 3. A Panopticon video of a design game

In the fifth round, the participants begin to explore the video clips from the design game, with the aim of answering their research question. During this round, the tablet interface is augmented with an additional text box and virtual keyboard, which enables the players to annotate selected video clips. Each player is asked to choose three video clips that offer insight into their research question, using the annotation functionality to note the rationale for each selection. As the purpose of this round is to allow the players to propose potentially *divergent* interpretations of each of the video clips, this annotation process is completed individually, with players coming together at the end of the round to present their selections to the other group members.

During the final round, the players collaborate around one of the tablets, to develop an answer to their research question by drawing and naming relations between the clips annotated in the previous round. The single tablet displays an interface (Figure 4) that shows each of the clips, surrounded by any annotations made by the players. Annotations are represented by *virtual* post-it notes, which are colored in order to identify the player who wrote them. These annotations, and the videos to which they relate, can be moved, rotated and resized on the interface using simple multi-touch gestures. Furthermore, a mode can be selected whereby the players are able to write and draw on the background of the screen, to illustrate connections between particular clips and annotations. This final activity seeks to scaffold the players in *converging* their individual understandings of the clips into an answer to their research question, which in turn represents a shared understanding of the design process conducted by the group.

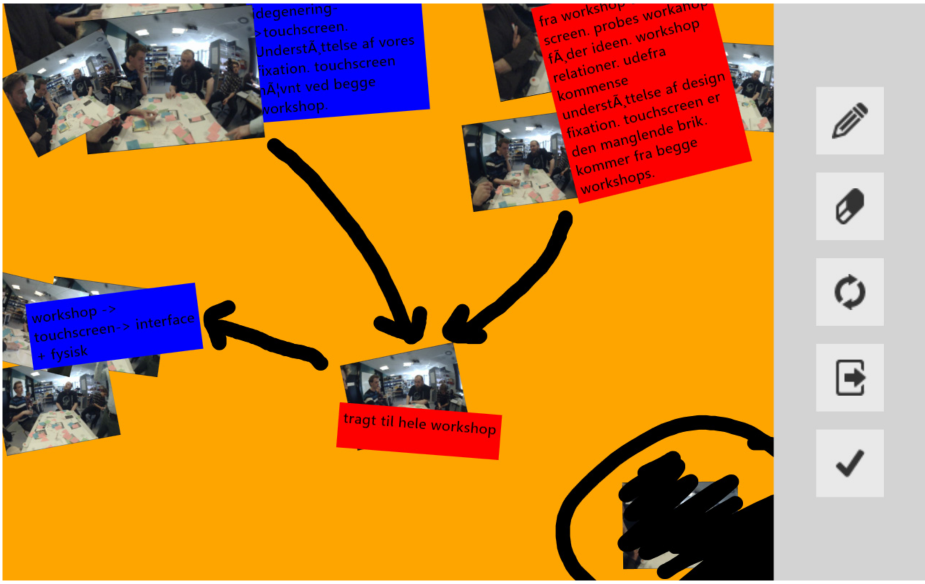


Fig. 4. The interface used in the last round of the game

4 Development and Exploratory Study

The ReflecTable was designed to explore how design games and video-led reflection might be combined to bridge the gap between theory and practice in design education. We used a research-through-design approach [39], whereby the iterative design and development of the ReflecTable game and hardware platform was used to scaffold our investigation of this research question. This design and development process occurred over the course of nine ReflecTable games, which were each played with three to four players. In the first six of these games, we focused on developing the game rules using a mock-up of the final setup. In the final three games, we tested the finished prototype of the ReflecTable setup, as described in the previous section. We documented each of these games using video, pictures and observational notes.

In order to explore if the evolving design game was supportive of students' reflection upon the relationships between design theory and practice, before commencing development of the final prototype, we studied two of the initial six games in further detail (games 2 and 3 in the series). As these games were played prior to the development of the final prototype, we used a mock-up system that allowed students to press buttons to record video clips. However, manual video editing was required to produce the video media for the second stage of the game. This meant that two time slots had to be scheduled and, as such, the two stages of the game were played two to three weeks apart. In the following sections, we describe the findings of this exploratory study. We present these initial exploratory findings as they provide valuable insight into the kinds of learning supported by the ReflecTable game, which was not changed substantially from the version used in this study to the final design.

The study was completed with two groups of undergraduate interaction design students, who were in their first semester. The design education of these students had been based primarily upon the work of Schön on reflective design practice [30, 32] as well as the theories of Löwgreen and Stolterman [22], and Nelson and Stolterman [27]. To explore the learning outcomes from these two games, we gathered two kinds of qualitative data from the participants. On completion of stage B, we asked the participants to write three post-its about the group's outcome and present these to the other players. We then conducted short semi-structured interviews [12] (six minutes on average) with each participant about their experiences of playing the game and what they thought they had learned. These interviews included a targeted question that sought to explore whether the students could relate their experiences of playing the ReflecTable game with concepts from Schön's theories of design. By asking the students to comment upon the relationship between their design process and one of the theories they had been taught, our intention was to gain insight into whether the game had supported them in learning to relate their experiences of practice to design theory. The questions asked were as follows:

1. What did you personally gain from playing the game?
2. What did you learn about designing by playing the game?
3. Can you recognize any of Schön's concepts in the game?
4. What do you think our intention is with this game?

We recorded and transcribed these interviews and, subsequently, analyzed them using two different coding procedures[1]. Firstly, an inductive approach was used, where we coded and grouped students' comments into themes and, secondly, a theoretical approach was used where we identified terms from design theory – primarily Schön's theory – in the students' responses [5]. Table 2 shows an overview of the design concepts that we identified in the interviews as either an explicit

[1] The excerpts presented from these transcripts have been translated from Danish to English.

Table 2. Design theory concepts identified in the interviews

Design terms	Players							
	2-1	2-2	2-3	2-4	3-1	3-2	3-3	3-4
Conversation with the materials		A		A(x)	B			A
Design judgment		A		B		B	B	
Learning-by-doing	A		A	A	A	B		A
Professional artistry	A				A(x)			
Reflection-in/on-action			A(x)	A(x)	A(x)		A(x)	A(x)
Reflective practice					A	B(x)		A
Repertoire			B	B	A		B	A
World-making							A(x)	A(x)

statement (marked with '(x)') or an implicit description. An 'A' indicates that a concept was mentioned after the topic of Schön's theory had been broached in discussion (e.g. question three had been asked), while a 'B' is used where the players mentioned concepts beforehand.

4.1 Learning Outcomes

Schön notes that it is hard to know exactly whether a student has been taught the "artistry" of design (i.e. the skill of being able to understand and apply a repertoire of theories and approaches sensitively and appropriately in practice) that the ReflecTable seeks to teach. However, he describes four dimensions of learning [31, pp. 168-170] that represent competencies that a student might possess in varying degrees. In the following sections, we discuss the students' responses to the semi-structured interviews in terms of these dimensions. We refer to the students using a system where the first number denotes the game/group and the second the player (e.g. "player 2-4" or "player 3-2").

Understanding Vocabulary. The first dimension that Schön described concerns the student's ability to know and use the vocabulary of their field. The student should not only be able to name the concepts that they have been taught, but demonstrate an understanding of the processes to which they refer and their relationship with design practice. The results of the study suggest that the ReflecTable can help students to gain a better understanding of the relation between the design vocabulary that they knew from lectures and their own design practices. In the interviews, all of the participants described how they had encountered one or more design concepts during the game. Table 2 illustrates how some students, such as player 3-2, were very good at using the vocabulary offhand, while others, such as player 3-4, were very good at identifying the concepts when prompted to do so.

The most prominent design concepts that students demonstrated an understanding of when discussing the ReflecTable game were those of reflective conversation with materials and learning-by-doing. For example, player 3-1 discovered that designing is easier when you work actively with the artifacts or, in Schön's words, have a "reflective conversation with the materials of the design situation".

Researcher: What did you personally gain from participating in this design game?
Player 3-1: Well, I probably realized, that I'm better at thinking creatively when I'm doing something [...] Because I've sometimes wondered 'Am I even thinking creatively enough' and stuff like that, when you're sitting by your computer and 'now, you should write a design proposal'. And then you try something like this, and you feel that, 'yes, you can do it'.

These are important insights because the metaphor of conversation embodies some of the most central pragmatic points in both Schön's theory of design and many of the other theories that the students are taught as part of their course. We believe that the students' ability to externalize connections between their practice and design theory in this way suggests a stronger understanding of the design vocabulary and its connection to design practice, as a result of playing the ReflecTable game.

Appropriating Procedures. The second dimension concerns a student's ability to apply design procedures sensitively and appropriately. Schön states that a competent student should not only know how to apply discrete techniques, but also to combine them and tailor them to the specific demands of a design situation. Every designer has what Schön terms a "repertoire" of design techniques, which they can apply appropriately and fluidly in their practice. Interestingly, students were spurred to reflect upon the role that the ReflecTable game itself, as a design method or technique, might play in their wider and future design practices. Five of the students described how particular aspects of the design game (e.g. the convergence and divergence of ideas in the group) could be put in their toolbox for later use. For example, one student explained that he felt his group could appropriate aspects of the ReflecTable game for use in their future design practices.

Player 3-3: All this you have shown us here is naturally a part of your own project, but it is also something that we can take parts of and take them out of their context, break them down a bit and use them as a tool in themselves. For example, we take a lot of pictures and group them, kind of like a new twist on the inspiration card workshop (see [23]) that we already know.

Students were able to highlight a range of qualities of the ReflecTable design game that might be appropriated in their future practices. As for the ReflecTable itself, three of the four students in group 2 commented on the value of ReflecTable as a learning-by-doing educational tool, while group 3 saw it as an idea generator because the process of playing and reflecting upon the game had helped them to come up with a new idea for their project.

Generalizing Cases and Critical Stance. The third dimension describes a student's ability to generalize their experiences of design practices, as insight that will inform how they address future design situations, while the fourth dimension concerns a student's ability to critically appraise the traditions of the field to which they are being inducted. The students did not clearly exhibit such competencies when playing the ReflecTable game during our studies. However, we hypothesize that this might be because the students had only just begun their education and, as such, had limited practical and theoretical knowledge upon which to base discussions, reflections and critical evaluations of theory. We expect that more experienced students would have been better equipped to tackle these higher order issues when playing the ReflecTable game.

5 Discussion: Scaffolding Reflection on Design

The ReflecTable was designed to bridge the gap between the theoretical and practical components of design education, by stimulating design students to reflect upon a condensed design process conducted during a design game. Our initial studies of the ReflecTable suggested that it supports design students in understanding how the theoretical concepts (vocabulary) and methods (procedures) they are taught in lectures relate to different design situations and their own evolving design practices. In this section, we reflect further upon our experiences of using the ReflecTable throughout the course of all of the nine design games (including three using the final prototype) to examine how the different components of the ReflecTable contributed to the students' reflection on design practice. Consequently, we uncover insights that will guide and inform those wishing to design future interfaces that support reflection in design education and, potentially, other related domains.

5.1 Using 'Real Life' Design Artifacts and Situations

The ReflecTable game used artifacts and materials from real design challenges and situations faced by the players. These artifacts were found to form the basis of much of the discussion amongst participants, by representing and highlighting different perspectives and experiences upon the shared design problem. We found that the incorporation of real life materials and situations into the learning environment heightened the players' engagement with the game because they had a real stake in the project they were working with. Conversely, in situations where participants had already completed the idea generation phase of their project, we found that they tended to be less engaged with the game as its outcome would have a lesser affect on their work.

Most crucially, the use of artifacts drawn from real design projects, allowed students to reflect upon aspects of their design processes, which might not have otherwise been broached in the limited setting of the design game. For instance, artifacts from field studies served to ground the game in an actual design setting, highlighting issues that might have been overlooked had the game been framed as a simple discussion of the problem at hand. Consequently, we argue that the use of real life artifacts and design situations proved to be an invaluable quality of the ReflecTable, as it situated the design game, and subsequent video-led reflection, in the context of the players ongoing, real life design practice. Törpel [36] has argued that the educational value of design games might be limited by the difference between experiences developed while playing design games and those required of real world design practice. We believe that our use of students' own real life materials and situations to foster reflection on wider design practice, rather than just the condensed design process completed during the session, stands out as a potential way to rectify this potential limitation of design games.

5.2 The Buttons

Players were given buttons that could be pressed during the ReflecTable game, in order to capture the previous 15 seconds of video, for review in the second stage. By providing this lightweight video-capture system, our intention was to build upon previous work that leverages design documentation as a spur for reflection on the design process, by allowing design students to capture video of actual momentary experiences of designing for later reflection, in addition to artifacts and notes. Furthermore, by providing a mechanism to capture noteworthy moments in the design game, we sought to frame stage A itself as a reflective process, rather than simply an activity to be reflected upon during stage B. That is to say, the simple action of pressing the buttons was designed to inspire the students to engage in a higher form of reflection about what was interesting and important about their design process, while it was still ongoing.

Throughout the studies, the students were found to use the buttons as we intended to mark moments that they wished to remember and discuss later in the session. Often, button presses would occur collaboratively, with a press proceeded by a discussion of the relative interest of an event. In some cases, the players' use of the buttons took on an unexpectedly performative quality, as the visible pressing of a button acted as an expression to the remainder of the group that an interesting or important event had occurred.

Schön describes "reflection-in-action" (i.e. reflection which occurs during the course of an activity such as designing) as often being spurred by an unexpected or surprising interruption of a normally automatic activity [32, p. 56]. We observed that the simple task of pressing a button when a noteworthy event occurred, and observing and discussing the button presses of others, interrupted the design process in such a way that scaffolded both individual and collective reflection-in-action. Therefore, we hypothesize that the buttons might have acted as more than simply a means to record video clips of noteworthy events from the design game, as the source material for later reflection, but rather, were an important stimulus for the students' reflection on the relationship between theoretical and practical aspects of design observed during the studies.

5.3 Video-Led Reflection

In the second stage of the ReflecTable game, the students were asked to formulate and explore a research question about their design practice, using the video clips captured of the condensed design process conducted during the previous stage. These clips were explored using a Panopticon video rendering, which allowed the players to see video footage of the moments surrounding each button press, the temporal relationships between button presses and, finally, a video overview of the whole design game, all at a glance on one interface. Consequently, we intended to provide the players with a powerful tool that allowed them to not only browse and view individual video clips, but also to evaluate, discuss and reflect upon them in the wider context of the whole design game.

Reviewing video clips in this way was found to offer participants a practical way to discuss each button press in the detailed context within which it occurred; therefore, fostering situated and holistic consideration of the key moments in the design process. By combining this post hoc reflection with the reflection-in-action fostered during the initial stage of the design game, the ReflecTable extends previous work that has sought to support designers in reflecting either in [24, 26] or on [13, 25, 35] the design process, but not both. We found that this combination of reflection in and on action, as part of a single design game, led to a valuable two stage reflective process that allowed participants to revisit initial moments of reflection (spurred by both playing the design game and the pressing of buttons) for further detailed and prolonged examination. Furthermore, by allowing students to reflect upon videos of their own practices, the ReflecTable also responded to the concerns of Iversen and Burr [19] that designers might struggle to empathize with, and hence reflect upon, videos of other designers and design cases that they are not involved in.

The participants who tried the prototype setup were generally enthusiastic about the potential of the Panopticon video rendering. However, we found that the Panopticon experience was limited by the horizontal camera position, which prevented participants from using it to examine the materials on the table in detail (i.e. to read the contents of post-it notes). We believe that using a top down camera might have rectified this problem. Furthermore, the initial design of the interface only allowed players to view the video clips captured during the session in high resolution. Therefore, players could not explore the context of a button press in detail, beyond the initial 15 seconds of video captured. While this issue could be resolved easily by allowing students to use the Panopticon rendering to explore the entire video as well as button presses, it is anticipated that such changes to the design might dilute the meaning of a button press during the design game and, therefore, its ability to provoke reflection amongst players.

5.4 The Frame and the Facilitator

The facilitator played a central role in the ReflecTable, guiding the students through the game, explaining the rules, answering questions and encouraging the players when they became stuck. The students found that, in these ways, the facilitator played an important role in supporting their learning during the game. For example, player 2-4 described the value of the “gentle provocations” that the facilitator gave to the group during the session, which, when combined with the frame of the design game, guided the players to a successful outcome.

Player 2-4: ... it is like 'now, you read some texts' and then we have some practice classes, where we have Stine [an older student] walking around and coaching a bit. But she only comes around to the group once and she is does not stand by us and provokes us gently and sets up some frames [for the group work].

The idea that the facilitator can observe the players designing, while playing the ReflecTable game, and interject to provide such gentle provocations (i.e. guidance and advice) highlights potential parallels between the facilitator's role during a ReflecTable session and Schön's notion of a coach. Schön described how the relationship between student and coach depends on an environment where the student "tries to do what she seeks to learn and thereby reveals what she understands or misunderstands", so that the coach might respond with appropriate guidance [31, p. 162]. By fostering reflective discussion amongst players about the design process conducted, the ReflecTable was found to push students to externalize their understandings of design practice. Consequently, the ReflecTable was found to act as a valuable mechanism to support the facilitator in appreciating both individual players' and whole groups' understandings of particular theoretical concepts; therefore, supporting his ability to provide informed and appropriate coaching to the students during the game.

However, we became concerned during the studies that the facilitator might have had too strong of an influence on the students' design practices. That is to say, we were concerned that rather than freely and intuitively applying their design practices during the game, students might have been preoccupied with *living up to* their perceived expectations of the facilitator (i.e. following their advice and the rules of the game unquestionably). For example, one player commented *"Take a pen; I think he* [the facilitator] *would like you too"*. Therefore, we argue the presence of the facilitator, as coach, might in the case of some students actually prevent them from becoming immersed in the processes of reflection on design practice that the ReflecTable seeks to inspire.

6 Conclusions

In this paper, we have explored how the design of a digital learning environment might bridge the gap between the theoretical and practical components of design education. Our response to this challenge, the ReflecTable, builds upon previous work by combining video-led reflections and design work in a game set up, which enables students to comprehend both design theory and design practice in an interplay between understanding and doing design. The results of an initial exploratory study of the ReflecTable suggest that it has the potential to support design students in understanding how theoretical concepts and methods relate to different design situations and their own evolving design practices. Furthermore, our findings illustrate a range of valuable qualities of the ReflecTable design that might be drawn upon when developing future tools (e.g. the value of video as a means to combine both reflection-in and -on the design process).

In the future, we intend to extend this research by conducting a series of more extensive studies that compare the ReflecTable with other approaches to educating design students. We believe that such studies will help us better ascertain the effectiveness of both the ReflecTable and the combination of games and video-reflection as means to support design education.

References

1. Binder, T., De Michelis, G., Gervautz, M., Jacucci, G., Matkovic, K., Psik, T., Wagner, I.: Supporting Configurability in a Mixed-Media Environment for Design Students. Pers. Ubiquit. Comput. 8, 310–325 (2004)
2. Blevis, E., Lim, Y., Ozakca, M., Aneja, S.: Designing Interactivity for the Specific Context of Designerly Collaborations. In: Ext. Abs. of CHI 2005, pp. 1216–1219. ACM (2005)
3. Brandt, E., Messeter, J.: Facilitating collaboration through design games. In: Proc. of PDC 2004, pp. 121–131. ACM (2004)
4. Brandt, E.: Designing exploratory design games: a framework for participation in Participatory Design? In: Proc. PDC 2006, pp. 57–66. ACM (2006)
5. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. Qualitative Research in Psychology 3(2), 77–101 (2006)
6. Bucciarelli, L.: Delta Design Game. MIT Press (1991)
7. Buur, J., Soendergaard, A.: Video card game: an augmented environment for user centred design discussions. In: Proc. of DARE 2000, pp. 63–69. ACM (2000)
8. Dalsgaard, P., Halskov, K.: Reflective Design Documentation. In: Proc. of DIS 2012, pp. 428–437. ACM (2012)
9. Ehn, P.: Work-oriented Design of Computer Artifacts. Arbetslivscentrum (1988)
10. Everitt, K., Klemmer, S., Lee, R., Landay, J.: Two Worlds Apart: Bridging the Gap Between Physical and Virtual Media for Distributed Design Collaboration. In: Proc. CHI 2003, pp. 553–560. ACM (2003)
11. Fallman, D.: Design-Oriented Human-Computer Interaction. In: Proc. of CHI 2003, pp. 225–232. ACM (2003)
12. Flick, U.: An introduction to Qualitative Research. Sage Publishing (2006)
13. Geyer, F., Reiterer, H.: Experiences from Employing Evernote as a Tool for Documenting Collaborative Design Processes. In: Proc. of the DIS 2012 Workshop on Supporting Reflection in and on Design Processes, Newcastle, UK (2012)
14. Geyer, F., Pfeil, U., Budzinski, J., Höchtl, A., Reiterer, H.: AffinityTable - A Hybrid Surface for Supporting Affinity Diagramming. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 477–484. Springer, Heidelberg (2011)
15. Haller, M., Brandl, P., Leithinger, D., Leitner, J., Seifried, T., Billinghurst, M.: Shared Design Space: Sketching Ideas Using Digital Pens and a Large Augmented Tabletop Setup. In: Pan, Z., Cheok, D.A.D., Haller, M., Lau, R., Saito, H., Liang, R. (eds.) ICAT 2006. LNCS, vol. 4282, pp. 185–196. Springer, Heidelberg (2006)
16. Habraken, J., Gross, M.: Concept design games. Design Studies 9(3), 150–158 (1988)
17. Hartmann, B., Ringel Morris, M., Benko, H., Wilson, A.: Pictionaire: Supporting Collaborative Design Work by Integrating Physical and Digital Artifacts. In: Proc. of CSCW 2010, pp. 421–424. ACM (2010)
18. Iversen, O., Buur, J.: Design is a Game: Developing Design Competence in a Game Setting. In: Proc. of PDC 2002, pp. 22–28. CPSR (2002)
19. Iversen, O., Buur, J.: User Centred Design through the Keyhole: Video Design Case. In: Proc. of INTERACT 2003, pp. 431–438. IOS Press (2003)

20. Kirk, D., Sellen, A., Taylor, S., Villar, N., Izadi, S.: Putting the Physical into the Digital: Issues in Designing Hybrid Interactive Surfaces. In: Proc. of BCS-HCI 2009, pp. 35–44. British Computer Society (2009)
21. Klemmer, S., Newman, M., Farrell, R., Bilezikjian, M., Landay, J.: The Designers' Outpost: A Tangible Interface for Collaborative Web Site Design. In: Proc. of UIST 2001, pp. 1–10. ACM (2001)
22. Löwgreen, J., Stolterman, E.: Thoughtful Interaction Design. MIT Press (2004)
23. Madsen, K., Dalsgaard, P.: Inspiration Card Workshops. In: Proc.of DIS 2006, pp. 2–11. ACM (2006)
24. Mendels, P., Frens, J., Overbeeke, K.: Freed: A System for Creating Multiple Views of a Digital Collection during the Design Process. In: Proc. of CHI 2011, pp. 1481–1490. ACM (2011)
25. Nakakoji, K., Yamamoto, Y., Matsubara, N., Shirai, Y.: Toward Unweaving Streams of Thought for Reflection in Professional Software Design. IEEE Software 29(1), 34–38 (2012)
26. Nakakoji, K., Yamamoto, Y., Takada, S., Reeves, B.: Two-Dimensional Spatial Positioning as a Means for Reflection in Design. In: Proc. of DIS 2000, pp. 145–154. ACM (2000)
27. Nelson, H., Stolterman, E.: The Design Way–Intentional Change in an Unpredictable World. Educational Technology (2003)
28. Rieber, L.P., Smith, L., Noah, D.: The value of serious play. Educational Technology 38(6), 29–37 (1998)
29. Rittel, H., Webber, M.: Dilemmas in a General Theory of Planning. Policy Sciences 4(2), 155–169 (1973)
30. Schön, D.: Designing as a Reflective Conversation with the Materials of a Design Situation. Research in Engineering Design 3, 131–147 (1992)
31. Schön, D.: Educating the Reflective Practitioner – Toward a New Design for Teaching and Learning in the Professions. Jossey-Bass Publishers (1987)
32. Schön, D.: The Reflective Practitioner. Basic Books (1983)
33. Sokoler, T., Edeholt, H.: Physically embodied video snippets supporting collaborative exploration of video material during design sessions. In: Proc. of NordiCHI 2002, pp. 139–148. ACM (2002)
34. Sokoler, T., Edeholt, H., Johansson, M.: VideoTable: a tangible interface for collaborative exploration of video material during design sessions. In: Ext. Abs. of CHI 2002, pp. 656–657. ACM (2002)
35. Swan, L., Tanase, D., Taylor, A.: Design's Processional Character. In: Proc. of DIS 2010, pp. 65–74. ACM (2010)
36. Törpel, B.: The design game in participatory design and design education: chances, risks and side effects. In: Proc. of PDC 2006, pp. 77–86. ACM (2006)
37. Turns, J.: Enhancing Design Education for Engineering Students: Supporting Planning, Tracking and Reflection. In: Proc. of ICLS 1996, pp. 593–594. International Society of the Learning Sciences (2006)
38. Vyas, D., Nijholt, A., Heylen, D., Kröner, A., van der Veer, G.: Remarkable: Objects: Supporting Collaboration in a Creative Environment. In: Proc. of UBICOMP 2010, pp. 37–40. ACM (2010)
39. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through Design as a Method for Interaction Design Research in HCI. In: Proc. of CHI 2007, pp. 493–502. ACM (2007)

Evaluating User Experience for Interactive Television: Towards the Development of a Domain-Specific User Experience Questionnaire

Regina Bernhaupt[1] and Michael Pirker[2]

[1] IRIT, Team ICS, 118, Route de Narbonne, 31062 Toulouse, France
[2] ruwido, Köstendorferstr. 8, 5202 Neumarkt, Austria
Regina.Bernhaupt@irit.fr, Michael.Pirker@ruwido.com

Abstract. This paper presents a questionnaire-based approach to evaluate the user experience (UX) while interacting with interactive Television (iTV) systems. Current contributions in the field of UX propose generic methods applicable to various application domains, whereas our contribution is dedicated to the specific domain of interactive TV systems. Based on a classification of UX dimensions from a literature review, the first version of the questionnaire is focusing on the dimension's aesthetics, emotion, stimulation and identification. A validation study with 106 participants was performed to assess the relations between the evaluated UX dimensions, as well as their fit to the underlying theoretical assumptions. Results showed that the UX dimensions aesthetics, emotion and stimulation are important for the domain of iTV, while identification was not confirmed. The study revealed significant correlations between the type of IPTV system used and the emotional and stimulation dimension. Additionally, a significant effect of the TV reception mode and the type of IPTV box owned on the emotion towards the system was observed. Beyond the contribution of the questionnaire that is directly applicable for any iTV system, the findings described in the paper demonstrate the need for user experience evaluation methods targeted at specific domains: the validation of the questionnaire shows that identification is not a central dimension of user experience when interacting with interactive TV.

Keywords: interactive TV, iTV, user experience, UX, questionnaire, Internet Protocol Television, IPTV, emotion, stimulation, identification.

1 Introduction

User experience (UX) is described as dynamic, time dependent [29] and beyond the instrumental [18]. The overall goal of UX is to understand the role of affect as an antecedent, a consequence and a mediator of technology [18]. The concept of UX focuses rather on positive emotions and emotional outcomes such as joy, fun and pride [18]. The development of a general definition of UX is still focus of scientific discourse [32], and despite the lack of a clear definition, the concept of UX has become an important design aspect of interactive systems.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 642–659, 2013.

There are a growing number of methods available to evaluate user experience in all stages of the development process. Surveys on these contributions are already available such as in [3] where Bargas-Avila and Hornbæk present an overview on UX and UX evaluation methods or in [55] where Vermeeren and colleagues have been collecting and classifying user experience methods. Beyond that work on generic methods, contributions have been proposed for specific application domains, e.g. for games and entertainment [4] or for the evaluation of mobile devices like mobile phones [49].

What is still not answered is whether generic user experience evaluation methods provide meaningful and sufficiently precise insights when applied to specific application domains. In the field of usability the answer to that question was already negative [12]. Due to the idiosyncratic nature of user experience that is situational, time dependent, and influenced by the technological context especially the type of system and functionality, we advocate that there is a need to adapt, customize and validate specific UX evaluation methods for the respective application areas.

The domain of interactive TV (iTV) is described as being specific due to a variety of contextual factors, including the spatial, temporal, social, personal and technological context [42]. Beyond the difficulties brought by these factors, the evaluation of user experience for iTV is complex, as the system to be evaluated consists of multiple components: the TV screen, (a set of) remote controls, (possibly) a set top box (i.e. the hardware device connecting the TV to the IP network) and potentially other devices in the living room (e.g. surround sound system). One of the characteristics of the iTV domain is also the multiple goals of the users aiming at being entertained and informed.

We have been applying various UX methods in the area of iTV [5, 44, 45] which otherwise has received only limited attention. These studies have been carried out in close cooperation with one of the industry leaders in hardware and software systems for iTV. However, even though the evaluation results were iteratively fed into the development process of the systems, customer feedback reported by our client were exhibiting users' frustration and dissatisfaction, demonstrating a user experience far below expectations. Based on these results we have been working on the definition of a user experience evaluation method specifically adapted for iTV systems.

This paper presents the development and the validation of the first version of a user experience evaluation questionnaire called iTV-UX. The goal of the questionnaire is to enable the evaluation of all types of iTV systems in terms of user experience. The iTV-UX is based on four UX dimensions: visual and aesthetic experience, emotion and affect, stimulation, and identification.

2 State of the Art

User Experience (UX) is a research topic within the HCI community that has gained a lot of attention within the past years [3]. The term UX still lacks a clear and agreed-on definition. The reasons therefore are rooted in the nature of UX, which is associated within a lot of different concepts and meanings within the community, or as Law et al. [34, p. 1] put it "User Experience (UX) is a catchy as well as tricky research topic, given its broad applications in a diversity of interactive systems and its deep root in various conceptual frameworks, for instance, psychological theories of emotion."

Despite several attempts to define and better explain the user's experience when interacting with an interactive system in the past, the HCI community still has no unified definition of what really makes up UX, how to measure or evaluate UX, and how to rate it. An ISO Standard defining UX exists, but leaves a lot of room for interpretation, and is said to be too imprecise [24]: "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service."

Additionally, the term UX is also influenced by several concepts from other areas, like fun, playability, or Csikszentmihalyi's concept of flow [8, 26]. Within this multitude of concepts, as pointed out by Law et al. [32], the inclusion and exclusion of particular variables seem arbitrary, depending on the author's background and interest. Further reasons stated by Law et al. include the differences in the unit of analysis, ranging from a single aspect of a single user's interaction with a specific application to multi-users' interactions with a company and its services from multiple disciplines; and finally the fragmented landscape of UX research with diverse theoretical models with different foci [32]. In more recent work, Law [28] argues that the current UX researchers and practitioners may also roughly be divided into two camps, a "design-based UX research camp" which focuses more on qualitative approaches and a "model-based UX research camp" with a focus on quantitative approaches.

Similar reasons have already been pointed out by Hassenzahl [17]. He stated that the lack of agreement in definition of key elements makes building up empirical knowledge about what constitutes UX difficult, and that future research must aim at unifying approaches to UX, with its major objectives being the selection of key constructs and a better understanding of their interplay. Hassenzahl and Tractinsky [18] categorize current UX approaches and concepts in three groups:

1. Beyond the instrumental: One of the criticisms of usability evaluation was that a focus only on tasks is too narrow. Beauty as a perceptual experience, on the other hand, goes beyond the instrumental, functional concept of usability. It satisfies a human need, and was also described as an important quality of technology usage [1]. Other needs that are described as being important and should be addressed by technology include needs such as surprise, diversion or intimacy [14]. Further concepts falling into this category can be summarized as hedonic aspects of technology usage, like stimulation and identification [17]. All these UX approaches link product attributes with needs and values, the common goal is to enrich current models of product quality with non-instrumental aspects.

2. Emotion and affect: There are two basic ways in dealing with emotions in UX concepts. One way can be summarized as research that stresses the importance of emotions as consequences of product use, focusing on emotion as a result of the interaction or usage of a product. The other way concentrates on emotions that are preceding product use and contributing to the formation of evaluative judgments [18]. Current UX research rather, focuses on positive emotions and emotional outcomes such as joy, fun and pride, and also deals with emotion as a design goal [11].

3. The experiental: The third perspective looks at temporal and situational influences, asking for the dynamics of an interaction, how unique, complex, temporary or situated an experience is. In this view, an experience is a unique combination of various elements, such as the product and internal user states (e.g. mood, expectations, active goals), which extend over time with a definitive beginning and end. The experiential assumes all these elements to be interrelated – to interact and modify each other. The outcome of this process is the actual experience [13].

Generally, UX is described as focusing on the interaction between a person and a product, and is likely to change over time and with an embedding context [18, 32].

2.1 User Experience Evaluation

A broad variety of UX evaluation methods is available today, which are described and discussed in detail by Vermeeren et al. [55, 57] regarding their methodological approach (e.g. field studies, lab studies), the development phase they can be applied (from concepts to products on the market), the studied period of experience, and the evaluator. Law and van Schaik [33] summarize that UX measurement should essentially be self-reported, trajectory-based and adaptive, in accordance with the common understanding of UX as subjective, dynamic and context-dependent [18]. Thus, traditional techniques such as questionnaire, interview, and think-aloud remain important for capturing self-reported data [33].

Following the classification of UX approaches and concepts by Hassenzahl and Tractinsky [18], approaches focusing on the evaluation of emotion and affect include questionnaires like EmoCards [11] that ask the user to indicate the emotional state based on graphical representations. Other ways to measure the emotional response include physiological measurements (e.g. heart rate, skin conductance) or the evaluation of valence and arousal, which are currently also applied in the games area [39]. To measure the user experience beyond the instrumental, task-based approach, Hassenzahl [17] introduced the AttrakDiff questionnaire. In order to evaluate situational or temporal experiences, some approaches in mobile UX exist, using conceptual-analytical research and data gathering techniques [49]. For prototypes, usability evaluation methods can be enhanced by including experimental aspects to the evaluations, e.g. in long-term field trials, where diaries, experience sampling, questionnaires, and focus groups can be used to collect additional data [49].

2.2 User Experience Evaluation of Interactive TV

In the TV and entertainment sector, UX has been evaluated using a broad variety of methods. Dimensions that are addressed include emotions [36, 43], social factors [15], values and requirements [5, 42, 44], the perception of the quality of the interaction or representation [53, 47], or service quality and content [16, 31].

These user experience evaluations have been applying standard UX methods, like the AttrakDiff questionnaire [45]. Other measurements include the SUXES evaluation method, as used by Turunen et al. [53] to evaluate UX with different modalities in a mobile phone controlled home entertainment system. Experiences in gaming and TV applications were evaluated using psycho-physiological measurements [36]. These measurements were also used to evaluate the users' experiences with multi-view 3D displays [47] or to classify emotional reactions to video content [43]. Obrist et al. [42] investigated users' requirement and experiences within an ethnographic study and identified patterns how iTV services can support people. Tsekleves et al. [54] investigated the TV experience and media use habits using ethnographic observations. Roibas et al. [48] used ethno-methodologies like cultural probing and collaborative design to uncover the UX in future scenarios of mobile and pervasive iTV.

None of the methods applied for evaluating the overall user experience of interactive TV systems has been adapted to fit the interactive TV context. Interactive TV is special as the interaction with the system typically takes place at home, involving a certain type of setting (physical context) or influenced by people who might join the TV experience (social context) (see [42] for a discussion of the different contextual factors). While for the area of social factors and social TV Geerts and De Groof [15] have been proposing sociability heuristics, there is no adapted and validated UX evaluation method focusing on a set of UX dimensions specific for interactive TV in general.

3 Research Goals and Approach

Goal of our research is to develop a user experience evaluation method that allows us to evaluate the iTV specific user experience. Our research goal is motivated by the demand from the industry to have only one questionnaire combining the most important UX dimensions in this specific domain. The problem for the construction of such a method lies in the specificities of the iTV domain: it is necessary to take into account iTV characteristics and the context of usage. The user experience is shaped by interacting (using an interaction technology, i.e. a remote control, or other means of interaction) from a distance to control an interactive TV system and navigate in its menu structure, services and features. The spatial and temporal context of usage also influences the perception of the system (likely at home, at leisure time), as well as possible social factors (shared usage, co-experience). The UX of the iTV system itself is of interest, not the UX related to entertainment content.

The goal is to cover an extensive and holistic collection of user experience dimensions for the domain of iTV. The questionnaire development proceeds in three major steps, and the following methodological approach was chosen to identify the user experience dimensions used in the questionnaire and its subsequent development and iteration:

1. Identification of UX dimensions from the literature and development of a set of attributes based on existing questionnaires (presented in this paper),
2. Development and validation of the questionnaire including a set of the most important UX dimensions to verify their applicability in the domain with currently available iTV systems (presented in this paper), and finally
3. Extension of the set of dimensions to address the context specifics by adding additional dimension, and revalidation (future work).

3.1 Overview on UX Dimensions

Based on a literature review, we identified publications related to UX and its dimensions, the evaluation of product experiences and UX evaluation methods from entertainment and games. Table 1 gives an overview of the diversity of UX dimensions that are referred to when describing user experience.

Table 1. Results of literature review identifying four major user experience dimensions (VM:Value/Meaning, SRC: Social/Relatedness/Co-Experience; C:Challenge; DST: Dependability/Security/Trust; SQ: Service Quality)

	Aesthetic	Emotion/ Affect	Stimulation	Identification	Others
Hekkert 2006 [21]	x	x			VM
Desmet and Hekkert 2007 [10]	x	x		x	VM
Alben 1996 [1]	x	x			
Hassenzahl 2004, 2008 [17, 19]	x		x	x	SRC
Karapanos et al 2010 [29]			x		
Jordan 2000 [28]	x		x	x	VM, SRC
Wright et al 2003 [56]	x	x			
Hassenzahl et al 2010 [20]		x	x	x	VM, SRC, DST
Jääskö and Mattelmäki 2003 [27]	x		x	x	VM
De Angeli et al 2006 [9]	x		x		DST
Steen et al. 2003 [52]		x			
Gaver and Martin 2001 [14]		x	x		SRC
Desmet et al 2001, Mandryk et al 2006; Mahlke 2005, Minge 2005, Norman 1994 [11, 36, 35, 38, 40]		x			
Sheldon et al 2001 [50]		x	x	x	VM
McCarthy/Wright 2004 [37]		x			VM
Hartmann et al 2008, [16]	x	x	x		SQ
Turunen et al 2009 [53]		x			VM
Sproll et al 2010 [51]		x	x		SRC, DST
Battarbee 2003, 2004 [2]					SRC
Pirker et al. [44]	x	x	x		VM, SRC
Lavie & Tractinsky 2004 [31]	x	x	x		SQ
Ijsselsteijn et al 08 [22]	x	x			Flow, C

The classification identified aesthetics and beauty, emotion and affect, stimulation, identification, meaning and value, social factors, flow, immersion, involvement and engagement; challenge; service quality and content; and dependability, trust and security. The *aesthetics* dimension describes how aesthetically pleasing or beautiful something is perceived. The visual/aesthetic experience deals with the pleasure gained from sensory perceptions [21]. It incorporates beauty [17], as well as classic aesthetics (e.g. clear, symmetric) described by Lavie and Tractinsky [31]. It follows

Alben's [1] statement that objects have to be aesthetically pleasing and sensually satisfying.

Emotion has been identified as a key factor of UX [18]. For Desmet and Hekkert [10], the emotional experience is one of the three main factors contributing to product experience, including feelings and emotions elicited. Also Alben [1] addressed the factor emotion in the form of the emotional response as an outcome of the interaction. Izard [25] described ten basic emotions, of which the three clearly positive emotions were chosen to be included in the questionnaire (interest, joy and surprise), as UX is described as focusing on positive experiences [18]. Additionally, the feeling of competence as a need fulfillment is covered within this dimension [50, 20, 39].

The *stimulation* dimension describes to what extent a product can support the human need for innovative and interesting functions, interactions and contents. Hassenzahl [19] describes stimulation as a hedonic attribute of a product, which can lead to new impressions, opportunities and insights. Hedonic experiences were subsumed by Karapanos et al [29] under the term innovativeness to describe hedonic experiences and the ability of a product to excite the user through its novelty. In the area of games, Jääskö and Mattelmäki [27] defined product novelty as one of the qualities of user experience.

The *identification* dimension indicates to what extent a certain product allows the user to identify with it. For Hassenzahl [17], the identification dimension addresses the human need to express one's self through objects. Thus, using or owning a specific product is a way to reach a desired self-presentation. Identification can be seen as self-expression through an object to communicate identity.

3.2 Developing the Questionnaire

For the questionnaire we chose the most prominent UX dimensions to verify their applicability in the domain with currently available iTV systems: aesthetics, emotion, stimulation and identification. The decision to not include dimensions like social connectedness, interaction and value in this first version of the questionnaire is based on the current market situation for TV and I(P)TV products.

At the moment, most IPTV systems do not support social communication or services; additionally, the interaction takes place using a standard infra-red remote control. The dimension of Value might not be addressed properly in a broad evaluation of different TV systems as it is heavily influenced by the assembly of IPTV and entertainment-oriented devices and might provide better results when used in a specific setting in the future.

The questionnaire is based on a set of word-pairs for each of the dimensions identified. Each word pair represents an item of the questionnaire and is based on a seven-point semantic differential rating scale. The bi-polar adjectives used in the semantic differential scale where placed at the extremes of the scale; inversion of items was used to avoid fill-in schemes. This kind of scale was chosen as semantic differentials are described as a good choice to evaluate positive affective responses [58], which is the case for the evaluated UX dimensions.

The dimension visual/aesthetic experience is evaluated using seven bi-polar adjectives addressing beauty, the composition (classic aesthetics) and the design of the IPTV system.

It includes items like beautiful vs. ugly, or appealing vs. unappealing. Items were based on the work of Hassenzahl [17], Lavie [31] and Desmet et al [10].

The emotional response as an outcome of the interaction and the emotions elicited are evaluated using 14 adjectives addressing the positive emotions joy, interest, surprise, and the need for diversion and competency. Seven of the items were focusing on the personal emotions (E-P) and the feeling of competence of the respondent (e.g. happy, proud, competent), and seven of the items were focusing on the emotional reaction towards the system (E-S) (e.g. pleasant, fascinating, fun). Items were based on work found in the literature [e.g. [10, 22, 25 39] and fit to the theoretical concepts.

To evaluate the stimulation dimension, the questionnaire uses six word-pairs including inventive vs. typical, or creative vs. standard amongst others. The identification dimension was evaluated using six word-pairs including premium vs. cheap, presentable vs. unpresentable, amateurish vs. professional amongst others. Item selection was mostly based on Hassenzahl's AttrakDiff [17] questionnaire and its dimensions hedonic quality - stimulation and hedonic quality - identification. To motivate respondents to respond on the basis of their concrete experiences, questions were asked in relation to their last TV usage that lasted at least 20 minutes, and also the repetition that the question is focusing on their TV system before each new UX scale in the questionnaire. The questionnaire consists of 33 items in total. It was developed in an English master version, which was then translated into [Language] for the administration of the questionnaire. The translation was checked using back-translation by native speakers against the master version, which is described as a common way of ensuring validity in cross-cultural research [6].

The questionnaire was piloted in two steps, first using a think-aloud test, followed by pre-tests where three native speakers completed the questionnaire. Modifications resulting from the pre-tests have been incorporated in the questionnaire before its administration. Additionally, to be able to investigate differences between certain types of TV reception and the different types of IPTV set-top boxes available at the market in France, questions of how the respondents receive their TV signal and which kind of IPTV set-top box they own were added in the demographic part of the questionnaire. To get a higher response rate, and due to the fact that because of the current change to Digital Terrestrial Broadcasting, which also needs a set-top box to be received, the questionnaire was not limited to IPTV households only.

Usability of the evaluated system was evaluated using the already existing and validated SUS questionnaire. The SUS was included for further analysis of possible influences on the UX dimensions.

4 Validation Study

4.1 Method

Following our methodological approach, the aim of the study was to validate the first version of the developed questionnaire and to verify the applicability of the identified

UX factors in the TV and IPTV domain. The questionnaire was presented as an online survey, allowing a uniform administration to a large number of respondents.

The dimensionality of the questionnaire items was analyzed using maximum likelihood factor analysis to investigate their fit to the underlying theoretical assumptions that UX in this setting can be evaluated using the UX dimensions aesthetics, emotion, stimulation and identification. This approach should subsequently inform the further development of the questionnaire, which aims to investigate domain-specific user experience dimensions in the TV and IPTV context.

4.2 Procedure and Participants

To validate the initial version of the iTV-UX questionnaire, the questionnaire was distributed online using the online survey tool surveymonkey.com. Participants were invited using e-mail, social networks, word of mouth and by personal invitation.

The iTV-UX questionnaire had a brief introductory section, including obligatory statements on data anonymity, followed by a series of demographic questions (age, gender) and questions regarding the TV system used (reception, type of set-top box, media usage frequency). The sections of the questionnaire – introduction, demographics and media usage, UX scales, and debriefing - were represented on individual (web) pages. The questionnaire parts addressing the UX scales additionally had a brief introductory text to highlight that participants should evaluate the most recent 20 or more minutes usage of their TV system. At the end of the questionnaire, a standard usability scale was added. Questionnaire items were randomly inverted to avoid fill-in patterns like "1111". No monetary incentives were given to questionnaire participants. Overall, the fulfillment of the whole questionnaire lasted 10 minutes on average.

Over a three week period, 106 complete datasets were retrieved. 65 participants were between 20 and 29 years, 21 participants were between 30 and 39 years, and 20 participants were older than 40 years. 59.6 % were male, 40.4 % female (2 participants did not name their gender). The number of persons living in the household of the person filling in the questionnaire ranged from one to six persons. 15.2 % were living in single households, 46.7% of the participants lived in 2-person-households, 24.8% in 3-persons-households, and 13.4% of participants lived in households with 4 or more persons (n=105).

The vast majority (91.5%) of the participants watch TV at least once a week, with 64.2% watching every day, 22.6% several times a week and 4.7% of participants watching TV at least once a week. More than half of the participants (58.7%) receive their TV signal via ADSL (i.e. IPTV), 28.8% via digital terrestrial broadcast (TNT), 9.6% via satellite – Pay-TV, and 2.9% via satellite without Pay-TV (n=104).

The duration of the ownership of their current I(P)TV decoder was for a cumulated 83.5% more than six months, with 40.8% owning the decoder for more than 3 years; 33.0% for 1 year or more and 9.7% between 6 and 11 months. 7.8% of participants owned their decoder between 2 and 5 months, 5.8% between one and two months, 1.9% between 2 and 4 weeks and 1% for just a week (n=103). The rather large amount of new boxes can be explained by the fact that within the last half year, two major IPTV providers introduced the newest version of their set-top-box with new features and introductory offers.

The last TV usage situation was in 64.4% a social situation where the persons watched TV together with other persons, whereas 35.6% of participants watched TV alone (n=104). The last TV usage relative to the questionnaire submission date was in 92.6% of the cases within the last week, and for 73.4% of the answers even on the same day or the day before submission (n=94).

4.3 Data Analysis

Before further analysis of the results, the inter-item reliability of questions was computed for each original scale of UX dimensions using Cronbach's Alpha [7], which indicates the extent to which questions correlate to each other. A scale is typically considered reliable if its value for alpha is above the threshold of .7. The five scales used in the questionnaire had alphas ranging from .773 to .869 (Visual aesthetics: ,861, number of items =7; Emotion System ,852, n=8; Emo Personal ,773, n=6, Stimulation ,869, n=6; Identification: ,783 n=6). This analysis confirms that the scales are reliable indicators of the dimensions of UX we chose to investigate.

4.4 Factor Analysis

The UX questionnaire was using 33 items that were evaluated on a 7 point semantic differential scale using bi-polar adjectives. The items were addressing the UX dimensions aesthetics, emotion, stimulation and identification that were identified in the literature. The dimensionality of the 33 items from the UX questionnaire was subsequently analyzed using a maximum likelihood factor analysis.

According to Kline [30], performing factor analysis requires several participants per item, where the rule of thumb lies between 4 and 10 respondents per item, with a necessary minimum of 100 participants. For our first evaluation study, 106 participants for 33 questionnaire items is slightly below the 4 person per item rule, which seems to be reasonably acceptable for a first validation study to inform the further development of the questionnaire.

Three criteria were used to determine the number of factors to rotate:

(1) the a priori hypothesis that the measures were not uni-dimensional,
(2) the initial statistics of the principal component analysis including the Scree plot, and
(3) the interpretability of the factor solutions in accordance to the theoretical UX factors framework.

The principal component analysis indicated that the initial hypothesis that the measures are not uni-dimensional was correct. The analysis yielded seven components with eigenvalues above 1, explaining 70 % of the variance. Based on inspection of the Scree plot (as the results for eigenvalues bigger than 1 was no supportive indicator), we decided to carry out a subsequent factor analysis with maximum likelihood extraction assuming 4 factors (according to the Scree plot, where the eigenvalues were flattening out after 4 factors and the fitting with the underlying theoretical assumptions), which were rotated using a varimax rotation procedure for the UX factors that were evaluated with the 7pt semantic differential.

The rotated solutions yielded four interpretable factors (see Table 2) for Visual Aesthetics (VA), Emotional reaction towards the System (E-S), Emotion Personal (E-P) and Stimulation (ST). The UX factors accounted for: 16% for Visual Aesthetics, 15% for Emotional - System, 13% for Stimulation, and 9% for Emotional Person of item variance. The factor labels from the theoretical background suited the extracted factors and were retained. A total of four items were eliminated because they did not contribute to a simple factor structure and failed to meet a minimum criteria of having a primary factor loading of .4 or above, and no cross-loading of .3 or above.

Overall, the Factorial analysis of the results of the first validation study showed that most of the evaluated UX factors were loading on the predefined factors and were fitting the assumed underlying theoretical concept. Nevertheless, the validation revealed some important insights for the further development of the questionnaire. The emotion dimension is as expected split into Emotion vs. the System and the Personal Emotion / Feeling of Competence of the respondent. The topic of identification as an UX dimension showed no clear own factor, and items that were assumed to fit into the identification dimension showed factor loadings on other dimensions in the factor analysis and were thus moved to these scales for further analysis and development of the questionnaire. The final set of items kept for the questionnaire are for Visual Aestethics (VA): beautiful-ugly; like/do not like design; appealing/unappealing, visually well-arranged/confusing; stylish/unstylish; premium/cheap; for emotional reaction towards system: fun/boring; entertaining/unamusing; pleasant/unpleasant; impressive/unimposing; exciting/lame, fascinating/uninteresting; for Emotion Personal: confident/unsure; competent/incompetent; happy/sad; proud/embarassed; for Stimulation: innovative/conservative; novel/commonplace; inventive/typical; creative/standard. All labels are translated from French.

4.5 Correlations between the UX Factor Scales

To investigate the interrelations between the different UX indicators, correlation coefficients (Spearman's roh) were computed among the four indicator scales. The results of the correlation analyses show that 8 out of the 12 correlations were highly significant ($p < 0.01$) and all were greater or equal to 0.403; In general, the results suggest that the user experience factors are highly interrelated, which goes in line with the opinion of most UX researchers in the HCI community that the elements of user experience are heavily interrelated and influencing each other mutually.

Moreover, these results could be interpreted as an indicator that these UX factors or dimensions cannot be viewed independently.

Table 2. Four Factors and their items, rev - reversed, * *item* indicates removed in final version

Ass. Dim.	Item	VA	E-S	E-P	ST
	Visual Aesthetics (VA)	VA	E-S	E-P	ST
VA	Beautiful - ugly	**,843**	,205	,027	,162
VA	like the design / don't like design	**,803**	,127	,018	,196
VA (rev)	appealing -unappealing	**,790**	,011	,047	,214
VA	visually well-arranged - confusing	**,650**	,281	,143	,193
I	stylish - unstylish	**,607**	,233	,135	,264
I	premium - cheap	**,604**	,313	,073	,374
VA (rev) *	*clear lines – irregular*	,566	,069	,038	,030
I (rev) *	*Presentable- unpresentable*	**,563**	,285	-,054	,218
VA (rev) *	*Flawless – imperfect*	**,514**	,071	,106	,197
VA *	*symmetric – asymmetric*	**,358**	-,004	-,061	-,026
	Emotional Reaction towards System (E-S)				
E-S	fun – boring	,113	**,843**	,167	,148
E-S (rev)	Entertaining – unamusing	,159	**,819**	,107	,095
E-S	Pleasant – unpleasant	,212	**,718**	,250	,113
E-S	Impressive – unimposing	,023	**,613**	-,009	,370
ST	Exciting – lame	,294	**,567**	,038	,321
E-S (rev)	Fascinating – uninteresting	,252	**,529**	,271	,152
I *	*it fits me – doesn't fit me*	,167	**,481**	,205	,073
E-P *	Sociable – solitary	,071	**,353**	,128	,043
	Emotion Personal (E-P)				
E-P	Confident – unsure	-,039	,182	**,826**	,112
E-P (rev)	Competent – incompetent	-,134	-,032	**,709**	,036
E-P	Happy – sad	,139	,443	**,668**	-,035
E-P	Proud – embarassed	,096	,361	**,528**	-,003
E-P (rev) *	Relaxed – stressed	,142	,096	**,450**	,014
	Stimulation (ST)				
ST (rev)	innovative - conservative	,289	,283	-,021	**,868**
ST (rev)	Novel – commonplace	,322	,257	,004	**,851**
ST	Inventive – typical	,193	,268	,065	**,763**
ST	Creative – standard	,307	,282	,184	**,635**
I (rev) *	*Professional – amateurish*	,328	,126	,078	**,558**
ST (rev) *	*Challenging – easy*	,032	-,166	-,455	**,508**
	Other eliminated items				
E-S (rev)	*good*	,386	,554	,367	,139
E-S	*asthonishing*	-,030	,416	,062	,312
E-S (rev)	*uncommon*	,066	,442	-,123	,351
I (rev)	*important for me*	,359	,319	,248	,375

4.6 Correlations between Usability and UX Scales

For further investigation of the interrelations between different indicators, correlation coefficients (Spearman's roh) were computed. As a reference for usability, the questionnaire included the items from the SUS questionnaire.

The SUS rating of the system showed a significant correlation with all UX factor scales (Visual Aestethic ,447 $p<0.001$, ES ,642 $p<0.001$; EP ,528 $p<0.01$; ST ,271 p<005). The results indicate that the usability of interactive systems remains an important issue in evaluating iTV systems and usability is influencing the overall user experience of the evaluated IPTV systems. It is important to investigate in more detail how usability influences the perception of user experience, or if the two concepts are independent.

During further investigation and correlation analysis of the SUS rating indicator, no significant correlations could be found regarding the independent variables "Type of TV Reception", "Type of IPTV Bow owned", "Period of ownership of decoder", "Age", "Sex", "Number of Persons in Household", or "TV usage frequency" and "Number of Persons in Household".

4.7 Other Insights

Regarding the independent variables "Age", "Sex" and "Number of Persons in Household", no significant correlation could be found for the UX indicators.

We observed no significant correlations between the UX scales and the type of TV signal reception (Terrestrial, ADSL, Satellite). Nevertheless, we discovered a significant correlation between the type of IPTV box owned and the Emotion-System scale (rho= 0.346, $p<0.01$), as well as for the Stimulation scale (rho= 0.267, $p<0.05$).

Overall, the UX ratings for the observed dimensions in our sample were rather low, but with a positive trend for the Visual Aesthetics and both Emotion vs. System and Emotion Personal dimensions. The negative trend in the results regarding the Stimulation dimension suggest that in general, TV systems are not perceived as stimulating devices. This shows that the user experience can be different for systems, even if the perceived usability is the same.

Using One-Way ANOVA on the means of the four UX scales extracted from the factor analysis, there is a significant effect of TV reception type on the Emotion-System scale at the $p<.05$ level for the four conditions [$F(3,99) = 2.88$, $p = .04$].

Post hoc comparisons using the Tukey HSD test indicated that the mean score for the satellite Pay-TV condition ($M = 1.11$, $SD = 0.80$) was significantly different than the ADSL-IPTV condition ($M = .17$, $SD = 1.02$) and the TNT condition ($M = .14$, $SD = .92$). However, the Satellite (without Pay-TV) condition ($M = .34$, $SD = .76$) did not significantly differ from the other conditions. Taken together, these results suggest that the Pay-TV offer via satellite does have a positive effect on the emotion versus the system.

Another One-Way ANOVA was conducted to compare the effect of the type of ADSL box on the on the means of the four User Experience scales extracted from the factor analysis.

There was a significant effect of the type of ADSL box, again on the Emotion-System Scale, at the $p<.05$ level for the six conditions [$F(5,52) = 2.79$, $p = .026$].

Post hoc comparisons using the Tukey HSD test indicated that the mean score for one of the recently introduced IPTV boxes ($M = 1.08$, $SD = .88$) was significantly different than the other IPTV boxes.

As the sample size was low (58, 3 entries for 2 types of IPTV boxes were excluded), and Type II Errors are frequently in small sample sizes, additional post hoc comparisons using LSD were carried out, which resulted in a significant difference of the same IPTV box compared to all the other boxes. Taken together, these results suggest that this newly introduced IPTV set-top box which is including e.g. a BluRay Player, HDTV, a remote control that includes some simple gesture interaction and has a very elaborate design approach does have a positive effect on the emotional reaction versus the system.

4.8 Discussion of Results

The work reported here is clearly the first validation of the questionnaire. At current stage we have to deal with the following limitations: (1) the composition of the current sample influences the results; (2) the systems used influence the results; (3) participants answered the questionnaire not directly after the interaction with their iTV system. The clear limitation of the rather small sample size will be addressed in the next iteration of the questionnaire, which will be evaluated using a larger sample to assure the validity of the factor analysis.

The strong correlations between the UX factors up to now only show that variables are related, but not the determining influence in a particular direction. Moreover, Spearman's Roh correlations also do not state which influences are dominating determinants on particular factors. Therefore, a regression analysis will be conducted in future to come up with a model for iTV-UX, which can also be used and applied by fellow researchers in the field. This statistical model approach will enable fellow researchers to use the complete questionnaire or to focus only on specific influences, which means only a part of the UX dimensions are used. This step will be taken within the analysis of the next iteration of the questionnaire to assure validity through a higher sample size and coverage of a higher number of UX dimensions.

The extensibility of the questionnaire is critical for the development of a successful iTV user experience evaluation method, as next generation iTV system will include for instance social functionalities, new ways of representation of content and information (e.g. 3D), and also more sophisticated interaction modalities like touch, gesture or motion interaction.

5 Summary and Future Work

In this paper, we presented the results of a study aiming at the evaluation of a set of literature-based user experience dimensions for TV and IPTV systems using a questionnaire-based approach. The assessed UX dimensions were aesthetics, emotion, stimulation and identification. Results from a performed factor analysis showed that except for the identification dimensions all evaluated UX dimensions were loading on the predefined assumed factors and were fitting the assumed underlying concept. Also the factor labels from the theoretical framework suited the extracted factors and were retained; the extracted factors were labeled Visual Aesthetics, Emotion vs. the System, Emotion Personal, and Stimulation. Correlation analysis showed that the UX dimensions are highly interrelated, which goes in line with the opinion of most UX researchers in the HCI community that the elements of User Experience are heavily interrelated and influencing each other mutually. Additional analysis revealed also strong correlations of the UX scales to the Usability rating gained from the SUS scale. The assumed UX dimension identification did not show a clear own factor, the items assumed to fit into this dimensions showed mostly loadings on other factors. This suggests that identification may not be a major UX dimension for the domain of IPTV. A detailed list of the results of the factor analysis and the items used in the evaluation can be found in Table 2. Overall UX ratings for the observed UX dimensions were rather low, although all dimensions except stimulation showed a positive trend, which suggests that in general, TV and IPTV systems are not perceived as stimulating devices.

During further analysis we observed significant correlations between the type of IPTV box owned and the emotional reaction towards the system and the stimulation dimension. We also observed a significant effect of the TV reception mode on the emotion towards the system, which was significantly better for Pay-TV Systems via satellite compared to ADSL/IPTV offers and TNT. Also the type of ADSL/IPTV box showed a significant effect on the emotion towards the system, where a newly introduced IPTV set-top box which is including e.g. a BluRay Player, HDTV and has a very elaborate design approach is evaluated significantly better than other boxes.

While the usability of the systems was perceived the same for all IPTV offers, the UX evaluation indicates differences in the perception, indicating that the two concepts are different.

For our research goal to develop a questionnaire that covers an extensive and holistic collection of UX dimensions for the iTV domain, the present evaluation study was the first step. We followed our methodological approach to start with a smaller set of literature-based UX dimensions (aesthetics, emotion, stimulation and identification) and validate their applicability for the IPTV domain. Results are promising and supported our underlying theoretical assumptions that these dimension are important in this setting and measurable.

A second phase of validation with a revised questionnaire will address the identified limitations by extending the sample, ensuring that the interaction with the iTV system occurs closely related to filling in the questionnaire and by enclosing additional UX factors. These will include the value of the system for the user, social factors (shared usage, co-experience), the interaction and interaction technology, and also the perception of the quality of service. This next version aims to cover the major dimensions that are influencing the user experience in this field. When looking at recent attempts on social interactive TV systems, social factors seem to become an important factor for the evaluation of future systems. The construct of relatedness will be addressed together with other social factors like shared usage and co-experience. Evaluating these social factors can provide more valuable insights if future TV or connected home systems support social applications like communication. Also the ongoing development of interaction technologies that go beyond a standard button remote, e.g. by integrating gyroscopes or using touch-screens, will be considered in future versions of the questionnaire, and the interaction will be addressed as a factor influencing the UX of an IPTV system. The next iteration of the questionnaire will also incorporate entertainment-oriented UX factors like flow, as on state of the art IPTV set top boxes, games and other services are introduced and iTV systems and gaming consoles are moving closer towards each other.

References

1. Alben, L.: Quality of experience: defining the criteria for eff. interaction design. Interactions 3, 11–15 (1996)
2. Battarbee, K.: Co-Experience the social UX. In: Proc. CHI 2003 EA, pp. 730–731. ACM (2003)
3. Bargas-Avila, J., Hornbæk, K.: Old wine in new bottles or novel challenges. In: Proc. CHI 2011, pp. 2689–2698 (2011)
4. Bernhaupt, R.: User Experience Evaluation in Games and Entertainment. Springer (2010)
5. Bernhaupt, R., Pirker, M., Weiss, A., Wilfinger, D., Tscheligi, M.: Security, Privacy, and Personalization: ACM C. Entertainment 6, 2 (2012)
6. Brislin, R.W.: Back-Translation for Cross-Cultural Research. Journ. Cross-Cultural Psych. 1(3), 185–216 (1970)
7. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika 16(3), 297–334 (1951)
8. Csikszentmihalyi, M.: Flow: The Psych. of Opt. Experience. Harper Perennial, N.Y (1991)
9. De Angeli, A., Sutcliffe, A., Hartmann, J.: Interaction, usability and aesthetics: what influences users' preferences? In: Proc. DIS 2006. ACM, New York (2006)
10. Desmet, P.M.A., Hekkert, P.: Framework of product experience. International Journal of Design 1(1), 57–66 (2007)
11. Desmet, P.M.A., Overbeeke, C.J., Tax, S.J.E.T.: Designing products with added emotional value. The Design Journal 4(1), 32–47 (2001)
12. Forbrig, P., Bernhaupt, R., Winckler, M., Wesson, J.: 5th Workshop on Software and Usability Eng. Cross-Pollination: Patterns, Usability and UX. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 718–719. Springer, Heidelberg (2011)
13. Forlizzi, J., Battarbee, K.: Understanding experience in interactive systems. In: Proc. DIS 2004, pp. 261–268. ACM, New York (2004)
14. Gaver, B., Martin, H.: Alternatives: exploring information appliances through conceptual design proposals. In: Proc. CHI 2000, pp. 209–216. ACM, New York (2000)
15. Geerts, D., De Grooff, D.: Supporting the social uses of television: sociability heuristics for social tv. In: Proc. CHI 2009, pp. 595–604. ACM, New York (2009)
16. Hartmann, J., De Angeli, A., Sutcliffe, A.: Framing the user experience: information biases on website quality judgement. In: Proc. CHI 2008. ACM, New York (2008)
17. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. J. of HCI 19(4), 319–349 (2004)

18. Hassenzahl, M., Tractinsky, N.: User Experience - a research agenda. Behavior & Information Technology 25(2), 91–97 (2006)
19. Hassenzahl, M.: Aesthetics in interactive products: Correlates and consequences of beauty, ch.11. Product Experience, pp. 287–302. Elsevier (2008)
20. Hassenzahl, M., Diefenbach, S., Göritz, A.: Needs, affect, and interactive products - Facets of user experience. Interacting with Computers 22(5), 353–362 (2010)
21. Hekkert, P.: Design aesthetics: Principles of pleasure in product design. Psychology Science 48(2), 157–172 (2006)
22. IJsselsteijn, W.A., Poels, K., de Kort, Y.A.W.: Measuring Player Experiences in Digital Games. Development of the Game Experience Questionnaire (GEQ)
23. ISO 9241, ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability. ISO, Geneva, Switzerland (1998)
24. ISO 9241-210, Ergonomics of Human–System Interaction – Part 210: Human-centred Design for Interactive Systems (2010)
25. Izard, L.: Die Emotionen des Menschen. Psychologie Verlags Union, Weinheim (1971)
26. Järvinen, A., Heliö, S., Mäyrä, F.: Communication and Community in Digital Entertainment Services. Online (2002)
27. Jääskö, V., Mattelmäki, T.: Observing and probing. In: Proc. DPPI 2003, pp. 126–131. ACM, New York (2003)
28. Jordan, P.: Designing Pleasurable Products. Taylor & Francis, London (2000)
29. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: Measuring the dynamics of remembered experience over time. IwC 22(5), 328–335 (2010)
30. Kline, P.: The handbook of psychological testing. Routledge, London (1993)
31. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. Int. Journ. of Human-Computer Studies 60(3), 269–298 (2004)
32. Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: Proc. CHI 2009, ACM (2009)
33. Law, E., Van Schaik, P.: Modelling user experience - An agenda for research and practice. Interacting with Computers 22(5), 313–322 (2010)
34. Law, E.: The measurability and predictability of user experience. In: Proc. EICS 2011, pp. 1–10. ACM, NY (2011)
35. Mahlke, S.: Studying affect and emotions as important parts of the UX. Position Paper HCI 2005 (2005)
36. Mandryk, R., Inkpen, K., Calvert, T.: Using psychophysiological techniques to measure UX with entertainment technologies. Beh. and Inf. Techn. 25(2) (2006)
37. McCarthy, J., Wright, P.: Technology as experience. Interactions 11(5), 42–43 (2004)
38. Minge, M.: 2005. Methoden zur Erhebung emotionaler Aspekte bei der Interaktion mit technischen Systemen. Master's thesis, Freie Universität Berlin (2005)
39. Nacke, L.E., Grimshaw, M.N., Craig, A., Lindley, C.A.: More than a feeling: Measurement of sonic UX and psychophysiology in a first-person shooter game. Interact. Comput. 22(5), 336–343 (2010)
40. Norman, D.A.: Emotional design: why we love (or hate) everyday things. Basic Books, New York (2004)
41. O'Brien, J., Rodden, T.: Interactive systems in domestic environments. In: Proc. DIS 1997. ACM, NY (1997)
42. Obrist, M., Bernhaupt, R., Tscheligi, M.: Interactive television for the home: An ethnographic study on users' requirements and experiences. Int. Journ. of HCI 24(2) (2008)

43. Oliveira, E., Benovoy, M., Ribeiro, N., Chambel, T.: Towards emotional interaction: Using movies to automatically learn users' emotional states. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 152–161. Springer, Heidelberg (2011)
44. Pirker, M., Bernhaupt, R.: Measuring User Experience in the Living Room: Results from an Ethnographically Oriented Field Study Indicating Major Evaluation Factors. In: Proc. EuroITV 2011, pp. 79–82 (2011)
45. Pirker, M., Bernhaupt, R., Mirlacher, T.: Investigating usability and user experience as possible entry barriers for touch interaction in the living room. In: Proc. EuroITV 2010, pp. 145–154. ACM, New York (2010)
46. Quinn, J.M., Tran, T.Q.: Attractive phones don't have to work better: independent effects of attractiveness, effectiveness, and efficiency on perceived usability. In: Proc. CHI 2010, pp. 353–362. ACM, NY (2010)
47. Rajae-Joordens, R.: Measuring experiences in gaming and TV applications. In: Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products, pp. 77–90 (2008)
48. Roibás, A., Johnson, S.: Unfolding the user experience in new scenarios of pervasive interactive TV. In: CHI 2006 Extended Abstracts. ACM (2006)
49. Roto, V., Ketola, P., Huotari, S.: User Experience Evaluation in Nokia. In: Now Let's Do It in Practice - User Experience Evaluation Methods in Product Development Workshop in CHI 2008, Florence, Italy (2008)
50. Sheldon, K.M., Elliot, A.J., Kim, Y., Kasser, T.: What Is Satisfying About Satisfying Events? Testing 10 Candidate Psychological Needs. Journal of Personality and Social Psychology 80(2), 325–339 (2001)
51. Sproll, S., Peissner, M., Sturm, C.: From product concept to user experience: exploring UX potentials at early product stages. In: Proc. NordiCHI 2010. ACM, NY (2010)
52. Steen, M., Koning, N.: d. and Hoyng, L. 2003. The 'wow' experience – Conceptual model and tools. In: COST269 Conference Helsinki (2003)
53. Turunen, M., et al.: User expectations and user experience with different modalities in a mobile phone controlled home entertainment system. In: Proc MobileHCI, ACM, NY (2009)
54. Tsekleves, E., Whitham, R., Kondo, K., Hill, A.: Bringing the television experience to other media in the home: an ethnographic study. In: Proc. EuroITV 2009. ACM (2009)
55. Vermeeren, A., Law, E., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K.: User experience evaluation methods. In: Proc. NordiCHI 2010. ACM, NY (2010)
56. Wright, P., McCarthy, J., Meekison, L.: Making Sense of Experience. In: Funology: From Usability to Enjoyment, Kluwer Academic Publ., Dordrecht (2003)
57. `http://www.allaboutux.org/` (accessed January 20, 2012)
58. Heise, D.R.: The Semantic Differential and Attitude Research. Attitude Measurement, pp. 235–253. Rand McNally, Chicago (1970)

Managing User Experience – Managing Change

Job Mashapa, Edna Chelule, Darelle Van Greunen, and Alida Veldsman

Institute for ICT Advancement, School of ICT, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
{Job.Mashapa,Edna.Chelule,Darelle.vanGreunen, Alida.Veldsman}@nmmu.ac.za

Abstract. Interactive products with innovative user interfaces are being designed while the user interfaces of existing products are being improved. The changes in user interfaces are being prompted by the need to design products that are useful, usable and appealing for an enchanting user experience to the people using the products. It is harmoniously agreed within the user experience domain that a change in the user interface of a product consequently affects the user experience of the people who use the product. Furthermore, user experience practitioners and academics acknowledge that user experience evolves over time. Paradoxically, there is lack of strategies for managing user experience as it evolves, or when a new user interface is introduced. Change in user experience is a process that needs to be managed for a positive user experience to be attained. Literature is awash with models aimed at guiding and managing change implementation. On the contrary, most of the change management models are aimed at managing change in organizations while neglecting managing the user experience of the people to which change is introduced. At the time of writing this paper, no evidence was found of an existing model aimed at managing user experience, both in theory and practice. Following the aforementioned premise, the purpose of this paper is to propose theoretical requirements for managing user experience of the people using interactive products. The paper commences with a conceptual background synthesis of related domain components required for managing user experience. Thereafter, the requirements for managing user experience are determined. After-which the requirements are presented in a User Experience Management Requirements (UXMR) framework. The paper culminates with proposed future work.

Keywords: User Experience, Usability, User Centred Design, Change Management, Requirements.

1 Introduction

Every product whether designed considering the requirements for user experience or not, stimulates some level of user experience to the users interacting with it [1]. Users create cognitive expectations on a product's ease of use, usefulness and level of satisfaction based on their first impression of it [2], [3]. When a product is designed to comprehend these expectations, users' mental models and capabilities, it pleases them

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 660–677, 2013.

and arouses their senses positively [4]. A product that meets or exceeds the expectations of the user provides a positive user experience. On the contrary, a product that fails to satisfy the expectations of the users result in a negative user experience [5]. Despite the publicity of designing products for positive user experience most applications on the market fail because of poor user experience [6], [7], [8]. Users are often resistant to change in user interfaces, and would prefer a poor, but familiar, user interface to new and improved ones [9]. With such, any changes in the user interface of a product should be designed and implemented in a manner that complements the needs of the users, enhances acceptance of the product and encourages a positive user experience. The development of user experience is not a once off thing, but a process involving various transitional phases over time spans of product use [10], [11].

There exist many models to guide the product development lifecycle as compared and analysed by Davis et al. [12]. However, at the time of writing of this paper, stipulated requirements for managing the user experience and subsequently promoting the development of positive user experience of the people using the products were lacking. The aim of this paper is to answer the research question that states:

> What are the requirements for managing and improving the user experience of people using interactive products?

The User Experience Management Requirements (UXMR) Framework was developed to answer the mentioned research question. The UXMR Framework serves as the basis for providing holistic requirements for managing user experience. The comprehensiveness of the Framework lies in its all-inclusiveness of planning for user experience improvement, managing user experience during implementation of product changes and providing guidance of how to sustain a positive user experience and making it last.

Section 2 presents the background to the paper. Literature of technology acceptance and usage are explored in section 2.1and the concept of user experience is introduced in section 2.2. Managing change requires an understanding of the change management processes as introduced in section 2.3. The paper proposes how change management techniques may be implemented as a component of the requirements for managing user experience. Section 3 presents the requirements for managing user experience through the User Management Experience Requirements Framework (UXMR). Section 4 provides some guidelines for the application of UXMR, whilst section 5 concludes with the significance of the UXMR Framework and future work to improve on the credibility and applicability of the Framework.

2 Background

2.1 Technology Acceptance and Usage

Although there are a number of benefits associated with the use of technology, the uptake of the services offered by the technologies is contingent upon the willingness of the people to accept and use the technologies [13]. It is important for product developers to gain an understanding of the factors that influence users to accept and use their products. The adoption of technology depends on a variety of factors in a given environment. A number of researchers have explored and developed theories and models that aim at explaining user behaviour in relation to their acceptance of technology. Examples of such include the Technology Acceptance Model (TAM) [2], and the Unified Theory of Acceptance and Use of Technology (UTAUT) [14].

TAM postulates that attitude to use a product results when the users perceive the product to be useful and easy to use. The UTAUT puts forward performance expectancy, effort expectancy, social influence and facilitating conditions and the determinants for technology acceptance and usage. Performance expectancy and perceived usefulness harmoniously concur that users will most likely accept and use a product that they believe to be useful in accomplishing their tasks. Effort expectancy is similar to perceived ease of use. When users find a product to be easy to interact with, they are more inclined to use that product quite often.

While these models play a vital role in determining what factors influence acceptance and usage of technology, they are inadequate in addressing user experience issues as a key determinant towards product use. McCarthy and Wright [15] argue that people do not just use technology, but technology becomes imbedded in their lives. The authors highlighted the importance of considering the emotional, intellectual and sensual aspects of the users' interactions with technology. Thus, products that fit into the lives of people are pleasant to use and do not disrupt the activities of the users and provide a positive engagement between the users and the product. It is therefore important to be able to analyse, understand and manage the felt experiences of users as they interact with products. The following section investigates user experience and the nature thereof.

2.2 User Experience

In this section, the concept of user experience (UX) is explored. UX is defined, followed by a discussion of the factors that influences UX. The section concludes with a discussion on the evolutionary nature of user experience and an introduction to the User Experience Development Lifecycle Chart (UXDLC).

Defining User Experience. The complexity and multifaceted nature of user experience makes it difficult to present a conventional definition of user experience as witnessed by the various user experience definitions [allboutux.org1]. The complexity in defining user experience indicates how difficult managing user experience can be. In this paper, user experience is defined as a subjective judgement and feeling of the quality of a user's interaction with a product, to complete a specific task in a specific context [16], [17]. The quality of a user's interaction with the product is judged based on the usefulness, usability and the level of satisfaction, whilst interacting with the product.

An analysis of literature on user experience identified that user experience is determined by the expectations of the users prior to interaction with the product.

[1] www.allaboutux.org

Furthermore, it results from long or short-term interaction with a product; and it is an evaluation of the user expectations against how they feel as a result of the interaction with the product [18], [19], [20]. Various authors have looked at user experience from different domains such as games [21] and mobile user experience in m-Learning [8]. Calvillo et al. [21] state that a positive user experience, when a user is playing a video game, is achieved when the user feels to be in control of the game. Ownership and enjoyment result from both the hedonic and pragmatic qualities of the video game device as well as the gamming application. In mobile interaction, a number of core elements exist that determine the level of user experience. These elements include the aesthetic and visceral aspects of the device and the applications on the device as well as the availability of external variables to offer services. An example of the external variable will be the network infrastructure [8].

User experience can be positive or negative, depending on how well a product satisfies the user's expectations [4], [5]. A negative user experience is when the user finds a product to be boring; difficult to interact with or does not fit the intended use. A positive user experience is developed when the user finds the product to be usable with pleasure and satisfaction [5]. While TAM and UTAUT lay fundamental components in understanding user behaviours towards technology acceptance and usage, the models left out important aspects of user experience as a core variable in product adoption. It is therefore vital to establish the factors that influence user experience. Understanding the factors that influence user experience, form the basis of determining requirements for managing user experience.

As depicted from the definition of user experience, it is not restricted to the period during product use but includes the expectations of the user before interacting with a product and the summation of quality of interaction over a long period of time. In addition, the definition of user experience reveals the elements of user experience. These elements are the hedonic and pragmatic qualities of a product, the psychological and physiological state of the user and the context in which the product is used [22], [23], [24], [25]. Understanding the time spans and elements of user experience is important in determining the factors that impact on user experience. One cannot manage what he / she does not know.

Factors Influencing the User Experience. A variety of factors may influence users' experience with a product. The factors can be determined based on the elements of user experience and the process of designing for user experience. Determining these factors is essential in order to establish a resultant user experience and why a user would have a particular user experience [18]. An understanding of the specific context in which a person has a particular user experience and the cause thereof is important in managing and improving user experience.

The user experience of a person depends on his or her previous experiences, motivation and the context of use. Future user experiences are influenced by the user's previous experiences and expectations of how things work. This corresponds with the user's mental model and is present even before the interaction starts. For instance when users are interacting with online mobile instant messaging applications, they expect to find where to type their message, an option for inserting emotions and the functionality for sending the message. If any of these functions are inaccessible, the users' expectations will be shattered, resulting in a negative user experience. The whole interaction between the user and system happens in a dynamic context (social, situational, locational etc.) that influences and is influenced by what the user does and experiences. The natural personal view explains how user experience is shaped over time, highly subjective and connected to the context. A user's current experience is dependent on the internal state of mind of the user, at that very moment. A user can interact with a system/product in many different situations with different states of mind. This can be positive, while in a good frame of mind and negative in another [5].

The factors that affect user experience are presented in Fig 1. The User Experience Factor Diagram (UXFD) (cf. Fig 1) maps the factors influencing user experience to a respective period of product usage and time span/phase of user experience. The periods of product usage refers to the user experience before a user interacts with a product, during interaction and after a long period of repeated usage. The type of user experience namely anticipatory, momentary, episodic, cumulative or reflective determines the time span phases.

Anticipatory user experience results from the expectation of the user before interacting with the product. Momentary and episodic user experience relate to the feelings of the user during interaction with the product. Cumulative and reflective user experience arises as the user continues to interact with the product, thus discover and compare the product with their expectations or experience with similar products.

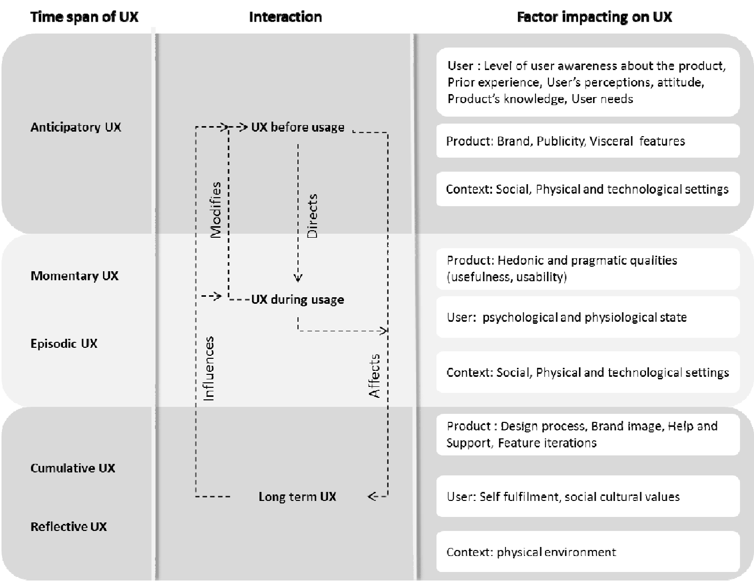


Fig. 1. User Experience Factor Diagram (Researcher's interpretation of literature)

Anticipatory user experience is directed by the factors that impact on user experience before the user interacts with the product. Such factors include the visceral aspects of the product.

Users are pleased with a product that is effective, efficient, safe and usable. If a user is using a video game and the system keeps freezing while the user is playing the game, then the user will experience a negative interaction with the product. User experience goes beyond the pragmatic and hedonic qualities of the product. It is also impacted by the physiological and psychological state of the user. Moreover, the social, physical and technological environment influences the users to enjoy their interaction with a product at one setting while they may have a negative user experience when the context varies.

Long-term user experience consists of cumulative user experiences and reflective user experiences. The factors affecting long-term user experience include the design process of the product. A product that is designed focusing on the needs, expectations and requirements of the users promote a bond of user loyalty to the product. Involving the users at every stage of product development and implementation creates a sense of belonging to the product thereby improving the user experience.

User experience, therefore, changes over time and is influenced by the internal state of mind of the user, the features of the product and the context in which the product is used. The development of user experience from anticipatory level to the reflective level is therefore a process. The next section describes evolutionary nature of user experience.

Evolutionary Nature of User Experience. The resultant relationship when a user interacts with a product is not static but composed of emotions subject to change over time as well as the context in which the product is used. Literature mentions the perceived pragmatic qualities of a product, user emotional reactions resulting from interacting with a product and perceived hedonic qualities of the product to influence the change in user experience and overall decision of the user about a product [10], [11], [26]. The visceral aspects of a product seem to have dominance in appealing to the sensory experience of the user only on purchasing a product. Such visceral aspects cease to be important to the user when the user interact with the product for a period of a month. Thereafter the touch and audition qualities of the product becomes a more important stimulus for the experiences. Fenko [27] and Mendoza [28] found out that the types of errors, error rate users make and the resultant user frustrations changes drastically over time. Thus, as users continue to use a product, certain aspects stimulating their user experience fade away and, as time progresses, other product attributes become more prominent determinants of user experience.

Abbasi et al. [11] developed a framework illustrating the evolution of user experience termed the User Experience Evolution Lifecycle (UXEL). The authors suggest that the user experience lifecycle consists of Phase1 – Designed UX, Phase 2- Perceived UX and Phase 3 – Actual UX. Their framework is aimed at guiding user experience requirements engineering and evaluating user experience at each of the evolutional stages. The UXEL contributed to the identification of actors involved in the evolution of user experience. It also helped in identifying the attributes of the actors at each phase of user experience development. However, the UXEL Framework is inadequate in that it does not show any aspects of change in user experience over time. Time span of user experience has been noted to be an important dimension in the development of user experience [18], [27], [29], [30]. The UXEL Framework only focuses on user experience during the user's interaction with the product in the specified context of use and invocation of the hedonic stimulations. This paper is unique in such a sense that, it proposes a User Experience Development Lifecycle Chart (UXDLC) to illustrate how user experience evolves over time. The UXDLC is based on the components of the User Experience Factor Diagram (cf. Fig 1). The UXDLC is presented in Fig 2.

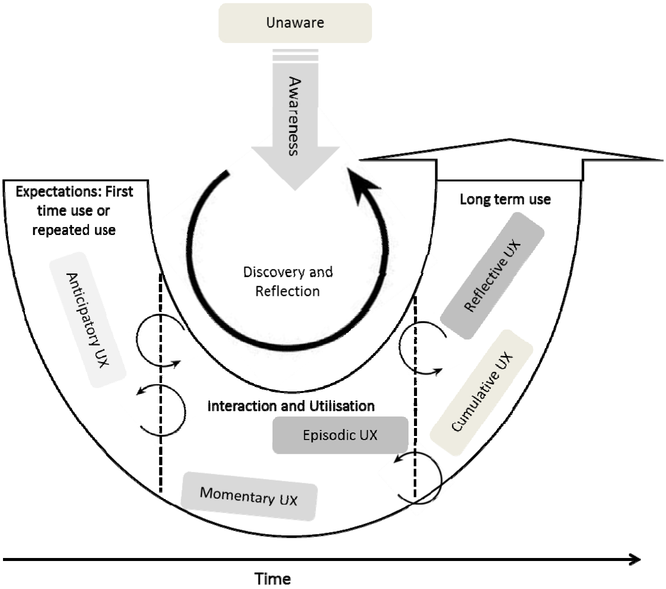


Fig. 2. User Experience Development Lifecycle Chart (Researcher's interpretation of literature)

User Experience Development Lifecycle Chart (UXDLC). The UXDLC begins with an unaware user. Unawareness refers to the user's lack of information either about the existence of a product, its instrumental qualities or hedonic attributes. The user is made aware of the product and its features. The UXDLC shows that awareness happens through the phases of the user experience development lifecycle. Awareness is created by product reviews, publicity about the product as well as the discovery of product features as users interact with it. As users continue to interact with the product, they discover new features and they reflect on them based on prior experience with similar products. A positive reflection is vital in building a lasting user – product relationship.

Awareness creates anticipatory user experience. Users interacting with the product for the first time, develop anticipatory user experience based on the mental models created from the reviews and publicity they receive about the product and the visceral aspects of the product. Whereas the anticipatory user experience of the users with prior interaction with the product is modified as they utilise the product.

Anticipatory user experience directs the expectations of the user as they interact and utilise the product. As the users use the product, they discover product qualities and reflect on them in momentary and episodic user experience. Accumulation of momentary user experience during a task session forms the episodic user experience. Momentary and episodic user experience is fluid and subject to change based on the context of use and the psychological and physiological aspects of the user at a particular moment. Episodic user experience accumulates into cumulative user experience as a user continues to interact with the product. As users continue to interact with the product, they discover product features resulting in the development of reflective user experience.

With continued usage and technological advances, users discover more about the product and their needs grow and change over time. The need to grow and adapt the product to meet these needs arise, hence the experience loops back to anticipatory user experience.

Thus, user experience evolves from an unaware user to anticipatory user experience. Anticipatory user experience is modified into momentary and episodic user experience as the users interact with a product and utilise it. Continual use of a product results in long-term user experience (cumulative and reflective user experiences). New features and product enhancements, which the users may not be aware of creates anticipatory user experience.

Product designers and developers who wish to have loyal users should aim at developing products appealing to long-term user experience. An effective long-term user experience is achieved by managing anticipatory, momentary and episodic user experience. It is therefore important to manage the change in user experience development. The techniques for managing user experience are inferred from change management. Change management concepts are discussed next.

2.3 Change Management

Lately there has been a notable paradigm shift in the focus of product design, from designing for functionality and usability to designing for user experience. Designing for user experiences involves adding or/and improving product features, changing the visceral look and feel aspects of the product, its navigation and interaction styles and the overall user interface. The resultant of such is change in user experience of the people using the improved products. It was discussed that the development of user experience is a transitional process for the users, involving changes in the subjective and objective factors. Like any other process, the process of change in user experience needs to be managed for a positive user experience to be achieved. Thus this section begins by defining change management and then analysing existing models aimed at managing change. A synthesis of the change management models contributes towards the requirements for managing the user experience.

Defining Change Management. Change is a complex process and often involves a revolution and transformation of the old and customary ways of doing things, with unfamiliar styles characterised by uncertainty, fear and risk to the lifestyles of the people involved [31 - 32]. Whenever a change is introduced into an organisation or society, it will ultimately affect the lifestyle of the people, organisational structure, tasks, job roles, processes and other related variables [33]. Creasey [33] differentiates change management from project management. He defines project management as methods, tools, skills and techniques to respond to change in project activities in order to meet project requirements while he defines change management as "the process, tools and techniques to manage the people-side of change to achieve the required business outcome". Thus, the processes, tools and techniques for managing the feelings of the people are examined through literature analysis.

Various models aimed at managing change and guiding implementation of change is available in literature [34 - 38]. Most of these models focus on managing change in organisations. This study aims at disseminating the requirements for managing change in user experience from the existing change management models. The uniqueness of requirements for managing user experience lies in the subjective nature of user experience and the need to adopt a user centred approach to managing change, contrary to the organisational oriented change management approaches.

Analysis of Change Management Models. An analysis of the various change management models shows that managing change consists of phases namely planning change, implementing change and managing change [35 - 36]. Preparation for change involves cultivating a favourable environment for change while change implementation refers to the execution of change activities. On the other hand, managing change deals with controlling the change process in order for the change to go according to plan. Creating a change-ready environment includes an examination of the environment to be changed, formulating the change strategy and creating the urgency for change [34, 38]. Activities for implementing change, include choosing change champions, communicating the change, imparting knowledge and the ability for people to change and creating a cultural fit for the change to last [37 - 38]. Managing change deals with the integration of lessons learnt during the change by measuring the progress of the change effort and reinforcing the change [35, 37 - 38]. These phases will be applied to propose the requirements for managing user experience.

3 User Experience Management Requirements Framework

The proposed User Experience Management Requirements (UXMR) Framework infers from change management models and factors affecting user experience. The UXMR Framework outlines the requirements for managing and improving the user experience of people using interactive products in a user centred approach. The user centred design philosophy has been the popularized approach aimed at designing products for positive user experience. Fig 3 presents UXMR Framework for managing user experience.

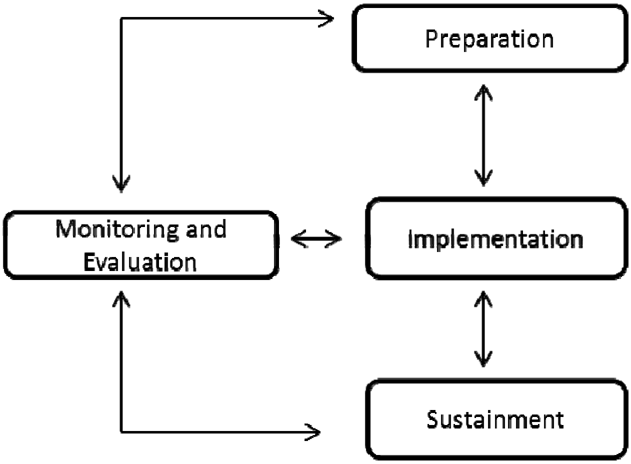


Fig. 3. User Experience Management Requirements Framework

The Framework is based on the three-phase change process and adopts a user-centred approach. The identified requirements are categorized into four categories namely: preparation (UXMR-I), implementation (UXMR-II), sustainment (UXMR-III), and monitoring and evaluation (UXMR-IV). It is important to note that the process of managing user experience is not as linear as presented in the diagram. Managing user experience is an iterative process. There is constant visiting and revisiting of each activity with the intention of perfecting the subsequent activities.

3.1 Preparation: (UXMR - I)

Managing user experience requires preparing the people for the change, managing the change in user experience during implementation and sustaining the new user experience. Preparation involves cultivating a favourable change environment, creating a change-ready audience and formulating the vision and strategy for the change. Table 1 presents the UXMR - I components and the related activities.

The Preparation phase requires an examination of the users and the context in which the product is used as well as the formulation of the change strategy and vision. Examining the users, involves determining their priorities that will lead to a favourable user experience, their user profiles and socio cultural background. It is important to assess the users' previous change history, as this will influence the present user experience. A successful user experience management exercise requires a 'test of the wetness of the waters'. Testing the wetness of the waters involves facilitating an informal awareness vibe about the new product initiative, stirring up change urgency by making the users to see and feel the need to change.

Table 1. User experience managment preparation requirements

UXMR – 1.1: Scanning the environment	
Task	Activities
Discovery of users and their tasks	User profiling, task analysis, assessing the level of user experience of the users
Assessing previous change history	Determining the resultant user experience of the users with respect to any previously introduced product. Analysing factors influencing user experience
Analysing the context of product use	Examining the physical, social, technological, social environment in which the product is used.
Testing the wetness of the waters	Facilitating an informal awareness vibe about the new product initiative, stirring up change urgency by making the users see and feel the need to change
Identifying the challenges and feature requirements of the current product	Determining what features of the product have to change, what new features have to be added based on the arising needs of the users.
UXMR- 1.2: Formulating the vision for change	
Task	Activities
Gathering a guiding team	Selection and defining roles for sponsors, change champions, user experience strategists, and user representatives to spearhead the process of managing user experience
Setting measurable user experience objectives	Stating the metrics for measuring the hedonic and pragmatic aspects of the product , defining effectiveness, satisfaction and efficiency
Defining the purpose and vision	Stating clearing what successful user experience should be
Identifying and recruiting stakeholders	Selection of cross sectional representatives of primary, secondary and tertiary users of the product so as to maintain a balance between user needs and the values and mission of the organisation
Crafting the change strategy and timeline	Defining the means of implementing change that suites the users and the environment to create a cultural fit promoting positive user experience
Brand strategy	Assessing and determining the brand and how the changes will potentially affect the brand perception

Further to scanning of the environment, user experience management preparation requires formulation of the implementation strategy and defining the vision. The user experience implementation strategy needs to align with the user and business requirements. This strategy needs to be supported by an agile change champions guiding team and cross-sectional representation of stakeholders. The guiding team is required to set the timeline and implementation plan for managing user experience. Measurable usability and user experience metrics for effectiveness, efficiency and enjoyment should be defined beforehand to ensure that the user and business requirements are meet. Clearly stipulated change management preparation requirements set the pace for a successful user experience implementation.

Thus, the output of the Preparation phase is a list of factors that may affect the user experience and a strategy on how to implement successful user experiences. The User Experience factor diagram presented in Figure 1 serves as a template guiding the preparation phase.

3.2 Implementation (UXRM – II)

The second set of requirements for managing the user experience involves managing the user experience during the implementation of the product changes. The components of requirements for managing the implementation of products for user experience are presented in Table 2.

Table 2. User experience management implementation requirements

UXMR – 2.1: Awareness	
Task	Activities
Formal communication of the change initiative	Formal communication on what aspects of the product are to be introduced or improved, how the change is going to affect the users, what aspects of the product will remain unchanged. Awareness of the potential benefits resulting from redesigning the product
UXMR – 2.2: Desire	
Task	Activities
Creating an affinity for people to embrace change with positive user experience	Employing persuasive techniques on the product so that the users will see and feel the need to change and embrace the new product with a positive user experience. Eliminating pockets of resistance hindering positive user experience and product acceptance
UXMR- 2.3: Knowledge	
Task	Activities
Passing on knowledge on how to use the product and the benefits of adopting the product	Training users that they will be able to discover the product features by themselves and be able to reflect on such
UXMR- 2.4: Ability	
Task	Activities
Giving the users a platform to demonstrate that they are able to transform knowledge into action	Promoting continuous and persuasive use of the product by observing the users as they interact with the product with the intention of improving their user experience

The requirements for managing user experience implementation are adopted from the ADKAR model [38]. Managing user experience requires making the user aware of the product, creating a desire for them to use the product, imparting knowledge on how to use the product and providing a platform for the users to demonstrate the ability to use the product.

Awareness. Managing user experience requires making people aware of the product by involving them from the onset of product development through the product lifecycle. A product with its features comprehending the expectations of the users, promotes a positive user experience. Putting users at the centre and making users aware of the changes are pivotal in a process of managing user experience for designs that are intended to be accepted and enjoyed. Without awareness, many users would rather prefer a bad, but familiar design to a better but unfamiliar one. Therefore, users need to be aware of what is changing, the potential benefits resulting from revamping the old product or introducing a new product. However, awareness on its own does not result in product acceptance. There is need to create a desire for the users to use the product.

Desire. User awareness does not result in making users develop a need to use the applications. There is a need to establish a sense of urgency and do away with pockets of resistance in using and re-using the product. There is a need to adopt emotional trust by persuasive design techniques to create an affinity for the users to embrace the changes with a positive user experience. Persuasive design aims at influencing change in the behaviour of the users. Conversely, the users may be aware of the changes and have a desire to embrace the new initiatives but if they do not have knowledge about how to use the product, then the whole change initiative is susceptible to the tyranny of legacy systems failing to provide a favourable user experience.

Knowledge. Managing user experience requires imparting the users with knowledge on how to use the product. Users need to know the associated benefits of using the products hence creating an emotional need to use the product. It is only when the users have the knowledge about the product, which they will be able to discover the product features by themselves and begin to reflect on their user experience. The more knowledgeable the users become about the product, the more they are inclined to use it productively and become loyal to it. Knowledge has to be applied in practice by providing the users with the platform to demonstrate that they are able to use the product.

Ability. The users should be given a platform to demonstrate their ability to use the product. Ability is achieved through continual usage of the product. Continual usage promotes long-term user experience, thus a product must have captivating features that promote its usage and adoption by the users.

3.3 Sustainment: (UXMR – III)

Change preparation and implementation aims at managing the behaviours and subjective emotions, resulting in positive user experience of the users interacting with the product. An effective approach to managing user experience during change preparation and implementation will result in a positive user experience. However, humans are creatures of habit and users are consequently subjected to revert to their traditional ways of doing, thus letting go of the aimed user experience. Therefore, the change in user experience must be sustained and improved upon. Table 3 presents the components of requirements for managing user experience sustainment.

Table 3. User experience sustainment requirements

UXMR – 3.1: Ensuring a lasting positive user experience	
Task	Activities
Celebrating short terms achievements	Recognising successful positive anticipatory, momentary and episodic user experience goals
Reinforcements	Promoting and ensuring an addictive point of no return positive user experience Employing persuasive techniques aimed at designing products for persuasion, emotion and trust.
Continuous user experience improvement	Continuous research to understand and address any emerging requirements, use of agile technologies to iteratively implement new features so as to satisfy an enchanting long term user experience.

Sustaining a lasting long-term user experience requires the celebration of short-term user experience achievements, making the user experience stick and allowing continuous improvement of the user experience. User experience development consists of anticipatory user experience, momentary user experience, episodic user experience and long-term user experience. Long-term user experience is a result of the accumulation of the overall user experience over time. It is therefore important to sustain long-term user experience. In the process of sustaining long-term user experience, successful anticipatory user experience, momentary user experience and episodic user experience goals have to be recognized and awarded. Once the users have shown to develop an affinity for the product, the positive user experience has to be reinforced to make it stick. Persuasive, emotional and trust research, and agile user experience design processes have to be implemented to keep the users captivated by the product they interact with. Such techniques aim at nurturing free will behavioral changes by appealing to the social influence factors of the people.

3.4 Monitoring and Evaluation: (UXMR – IV)

The process of monitoring and evaluation is central during managing user experience and has to be done during every stage. Monitoring and evaluation requires strategic communication, consolidation of lesson learnt as well as impact and outcome assessment of the process of managing user experience. Table 4 presents the requirement components of monitoring and evaluating user experience management.

Communicating the right message to correct people at the right time and contexts at every stage of implementation is an important requirement of managing user experience. Sending the wrong messages to the wrong people in the wrong context is a recipe for creating resistance and overall failure of managing user experience. Specific users have to be provided with contextually relevant information, only then will they positively accept the product being developed or introduced. The evaluation and measurement of the user experience is crucial to identify both hedonic and pragmatic user experience issues that may arise during product usage. Moreover, it is important to track lessons learnt in order to improve and manage user experience. This further helps to identify any loops missed and new requirements, which emerge from the users with an aim to unveil contingences of what has to be done for positive user experience to be achieved. Like any change management project, managing user experience requires an assessment of the impact and outcome process. The outcome is determined by comparing the business and user goals, metrics of user experience against the achieved user experience at a particular stage in managing user experience. Impact assessment thus entails evaluating the level of user experience of the product users at each stage of managing user experience.

The next section presents the users of the UXMR Framework and how the Framework works.

Table 4. User experience management monitoring and evaluation requirements

UXMR - 4.1 Tracking the user experience management process	
Task	Activities
Strategic communication	Sending the right message to the correct audience t the rightful context and time
Measuring user experience	Mapping transitional state to the desired outcome and evaluating the user experience level of the people at various phases during managing user experience. Iterative testing and implementation of recommendations found at each stage of the evaluation.
Consolidation of lessons learnt	Tracking any lessons learnt at each stage with the objective of improving on the process of managing user experience on next projects. Comparing what on hand with the initial plan to determine any variance so as to formulate what has to be done next to rectify the discrepancies

4 Application of the UXMR Framework

There exists a variety of guidelines and principles guiding the designing of products for user experience. However, the guidelines do not cater for managing change in user experience and the diverse factors that influence the user experience of the people. The target users of the proposed UXMR Framework are user experience practitioners and product developers. The Framework is aimed at guiding user experience practitioners on how to manage user experience by directing the designing processes of products for positive user experience. This is achieved through assessing the level of user experience of the target users, determining the factors that impact on user experience and providing interventions for improving the user experience. User experience practitioners will make use of the user experience factor diagram (c.f. Figure 1) to assess the factors that impact on user experience. A variety of user experience evaluation techniques may be used to determine the factors that influence user experience [www.allaboutux.org]. The findings are then used to develop a strategy for managing user experience. Practitioners will have to employ techniques for bringing awareness, creating desire, imparting knowledge and ability and sustaining the user experience. During the process of managing user experience, there is continuous monitoring and evaluation to ensure positive user experience. For product developers, the UXMR Framework serves to provide user centred design directions for designing products for positive user experience. The model integrates change management practices in the product development life cycle in order to design products that comprehend the expectations and skills of the target users.

5 Conclusion

This paper introduced user experience and the factors impacting on user experience by presenting the user experience factor diagram. The evolutionary nature of user experience was highlighted resulting in the development of the User Experience Development Lifecycle Chart (UXDLC). The evolutionary nature of user experience requires the need to manage change in user experience. Change management approaches were examined and inferences to such were made to determine the requirements for managing user experience. The User Experience Management Requirements (UXMR) Framework was developed based on the UX factor diagram and UXDLC. The target users of the proposed UXMR Framework are user experience practitioners and product owners, designers and an organisation's management as a whole. Using the UXMR Framework, user experience practitioners may influence the design of products for positive user experience. By assessing the factors that impact on user experience and providing interventions for improving the user experience, a positive user experience can be achieved. The requirements for managing user experience serve as design directions for user experience designers who wish to design their products for positive user experience.

The UXMR Framework was derived from a purely theoretical foundation. To improve on its applicability the Framework should be tested empirically. The plan is to validate the Framework through expert reviews. Experts from both the domains of change management and user experience will be involved to determine the appropriateness and relevance of the proposed construct requirements for managing user experience. The UXMR Framework will be also be applied in a case study in the design of a specific product to validate its usefulness and applicability based on the recommendations for improvement from the various experts.

References

1. Van Tyne, S.: Corporate User-Experience Maturity Model. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 635–639. Springer, Heidelberg (2009)
2. Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)
3. Zhang, P., Li, N.: The importance of affective quality. Communications of the ACM 48(9), 105–108 (2005)
4. Rogers, Y., Sharp, H., Preece, J.: Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, New York (2011)
5. Roto, V.: Web Browsing on Mobile Phones – Characteristics of User Experience. Finland (2007)
6. Bian, P., Zhang, N.: Gestalt Psychology Theory Based Design of User's Experience on Online Trade Websites. In: Xie, A., Huang, X. (eds.) Advances in Computer Science & Education. AISC, vol. 140, pp. 327–332. Springer, Heidelberg (2012)
7. Liedtke, C., Welfens, M.J., Rohn, H., Nordmann, J.: LIVING LAB: user-driven innovation for sustainability. International Journal of Sustainability in Higher Education, 106–118 (2012)
8. Botha, A., Herselman, M., van Greunen, D.: Mobile User Experience in Mlearning Environment. In: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists Bela Bela, South Africa, pp. 29–38. ACM (2010)
9. Constantine, L.: Beyond User-Centered Design and User Experience: Designing for User Performance. Cutter IT Journal 2, 1–12 (2004)
10. Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Karapanos, E., Sinnela, A.: UX Curve: A method for evaluating long-term user experience. Interacting with Computer 25(5), 1–11 (2011)
11. Abbasi, Q.M., Lew, P., Rafique, I., Li, Z.: User Experience Evolution Lifecycle Framework. International Journal of Social and Human Sciences 6, 39–44 (2012)
12. Davis, A.M., Bersoff, E.H., Comer, E.R.: A strategy for comparing alternative software development life cycle models. Software Engineering 14(10), 1453–1461 (1988)
13. Carter, L., Belanger, F.: Citizen Adoption of Electronic Government Initiatives. In: Proceedings of the 37th Hawaii International Conference on System Sciences (2004)
14. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27(3), 425–578 (2003)
15. McCarthy, J., Wright, J.: Technology as Experience. The Massachusetts Institute of Technology Press (2007)
16. Mashapa, J., van Greunen, D.: User Experience Evaluation Metrics for Usable Accounting Tools. In: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists Bela Bela, South Africa, pp. 170–181. ACM (2010)
17. Sutcliffe, A.: Designing for User Engagement: Aesthetic and Attractive User Interfaces. In: Carroll, J.M. (ed.) Synthesis Lectures on Human-Centered Informatics, pp. 1–55. Morgan and Claypool Publishers (2010)
18. Roto, V., Law, E., Vermeeren, A., Hoonhout, J.: User experience white paper: Bringing clarity to the concept of user experience (2011), `http://www.allaboutux.org/files/UXWhitePaper.pdf`

19. Obrist, M., Meschtscherjakov, A., Tscheligi, M.: User Experience Evaluation in the Mobile Context. In: Marcus, A., Sala, R., Roibás, C. (eds.) Mobile TV: Customizing Content and Experience, pp. 195–204. Springer, London (2010)
20. Hiltunen, M., Laukka, M., Luomala, J.: Mobile User Experience. IT Press, Finalnd (2002)
21. Calvillo-Gámez, E.H., Cairns, P., Cox, A.L.: Assessing the Core Elements of the Gaming Experience. In: Bernhaupt, R. (ed.) Evaluating User Experience in Games, pp. 47–71. Springer, London (2010)
22. Forlizzi, J., Ford, S.: The Building Blocks of Experience: An Early Framework for Interaction Designers. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 419–423. ACM, New York City (2000)
23. Arhippainen, L.: Studying user experience: issues and problems of mobile services case ADAMOS: User experience (im)possible to catch? Doctoral Dissertation University of Oulu. Finland (2009)
24. Mäkelä, A., Fulton Suri, J.: Supporting Users' Creativity: Design to Induce Pleasurable Experiences. In: Proceedings of the International Conference on Affective Human Factors Design, pp. 387–391 (2001)
25. Hess, W.: 10 Most Common Misconceptions About User Experience Design, `http://mashable.com/2009/01/09/user-experience-design/`
26. Mahlke, S., Thüring, M.: Studying Antecedents of Emotional Experiences in Interactive Contexts. In: CHI 2007 Proceedings, pp. 915–918 (2007)
27. Fenko, A., Schifferstein, H., Hekkert, P.: Shifts in sensory dominance between various stages of user-product interactions. Applied Ergonomics 40(1), 34–40 (2010)
28. Mendoza, V., Novick, D.: Usability over time. In: Proceedings of SIGDOC, Coventry, pp. 151–158 (2005)
29. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.: Measuring the Dynamics of Remembered Experience over Time. Interacting with Computers 22(5), 328–335 (2010)
30. von Wilamowitz-Moellendorff, M., Hassenzahl, M., Platz, A.: Dynamics of user experience: How the perceived quality of mobile phones changes over time. In: User Experience - Towards a Unified View, Workshop at the 4th Nordic Conference on Human-Computer Interaction (2006)
31. Vahs, D., Koch, V., Kielkopf, M.: Innovation Generating and Evaluation: The Impact of Change Management in Innovation and International Business Growth. Business Economics, 151–174 (2010)
32. Higgs, M.J., Rowland, D.: All changes great and small: Exploring approaches to change and its leadership. Journal of Change Management 5(2), 121–151 (2005)
33. Creasey, T.: Defining change management (2007), `http://www.change-management.com/prosci-defining-change-management.pdf`
34. Kotter, J., Cohen, D.: The heart of change: real-life stories of how people change their organizations. Harvard Business School Press, Boston (2002)
35. Jick, T.: Implementing Change. Note 9-191-114. Harvard Business School Press, Boston (1991)
36. Mento, A.J., Jones, R.J., Dirndorfer, W.: A change management process: Grounded in both theory and practice. Journal of Organizational Change Management 3(1), 45–59 (2002)
37. Price, A., Chahal, K.: A strategic framework for change management. Construction Management and Economics 24(3), 237–251 (2006)
38. Hiatt, J.: ADKAR: a Model for Change in Business, Government and our Community. Prosci Research, Loveland (2006)

What Do You See in the Cloud? Understanding the Cloud-Based User Experience through Practices

John C. Tang[1], Jed R. Brubaker[2,3], and Catherine C. Marshall[3]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98008
[2] Department of Informatics, University of California Irvine, Irvine, CA 92697
[3] Microsoft Research Silicon Valley, 1065 La Avenida, Mountain View, CA 94043
{johntang,cathymar}@microsoft.com, jed.brubaker@uci.edu

Abstract. End users have begun to incorporate cloud-based services into their collaborative practices. What spurs and constrains this adoption? Are the cloud services understood adequately and used effectively? How might we intervene to promote a better connection between user practices and cloud services? In this study, we focus on collaborative practices that surround the adoption, use, and understanding of two popular, but sometimes contrasting, cloud services for creating and sharing content: Dropbox and Google Docs. We conducted 22 in-depth interviews with people who used these services, including collaborators who used the services together, and people who had migrated from Google Docs to Google Drive. We found that users thought of the cloud in terms of the practices it helped them accomplish. Their understanding of the cloud was often shaped by the particular file storage and sharing technologies the cloud was replacing (remediation). Furthermore, collaborating with others through the cloud sometimes revealed different assumptions about how the cloud worked, leading users to develop socially negotiated practices around their use of the cloud. We use this analysis to identify some specific opportunities for designers to help users build more accurate conceptual models of the cloud and use its capabilities more fully: (1) when users are adopting the cloud to enact a practice; (2) when users are replacing an existing technology with the cloud; and (3) when users are encountering others' practices through collaboration.

Keywords: File synchronization, file sharing, online editors, collaboration, cloud user experience.

1 Introduction

Cloud-based file synchronizing (*syncing*) and sharing services have become a central element of everyday computing infrastructure. Ubiquitous access to storage, low-overhead file sharing, coordination between devices, and real-time collaboration facilities have prompted significant end-user adoption of these services. These services, like other cloud computing technologies, are realizing Weiser's vision of ubiquitous computing, in which computation is accessible everywhere, but seamless and quiet in its presentation [21]. However, this invisibility may also present some challenges to users of cloud-based syncing and sharing services. Despite the availability of these

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 678–695, 2013.

services, research indicates that many people still harbor significant misconceptions about how these services work and may not understand the features they offer [16].

To identify opportunities for improving the design and uptake of cloud-based services, we investigated how users understand and incorporate the services in their individual and collaborative practices. In particular, we were interested in three areas: (1) how people adopt cloud-based syncing and sharing services; (2) how people work together through these cloud services, and how their understandings of the services influence those of their collaborators; and (3) how designers can pinpoint appropriate opportunities to help users better connect their practices with functionality offered by cloud services.

To give ourselves a concrete starting place, we focused on peoples' experiences with Dropbox and Google Docs. Both are widely used services that provide ubiquitous access to files and support collaboration through shared storage in the cloud, but they go about it in ways that put varying conceptual demands on their users. Google Docs offers a browser-based environment for access to and co-creation of content directly in the cloud [11]; users need only conceive of cloud repositories as a distinct shared place to do their work (although they must learn to use the new editors). By contrast, Dropbox synchronizes local files among devices and people to provide ubiquitous access to content [9]; while interaction with the editors and local file system is familiar, fully understanding Dropbox requires grappling with a complex notion of file synchronization. The recent release of Google Drive [1], which marries a Dropbox-like syncing client with the remote editing capabilities of Google Docs, provided us with a window onto how users' cloud experiences evolved as they migrated from Docs to Drive.

Our analysis draws support from Bolter and Grusin's theory of *remediation* [2]. Remediation highlights the relationship between media forms and their predecessors, and hints at how the new media forms are understood in terms of the old. Bolter explains, "*designers of a new media form seek to borrow the cultural valence of one or more earlier forms. In most cases, however, they also want to define criteria by which the new form can surpass its predecessor, in order to give their audience a reason to adopt the new form.*" [3]

Remediation, which initially focused on media itself, has been extended to studies of media practice. According to Lanzara, "*[t]he appearance of a new medium in a domain of practice produces a perturbation in the complex ecology of agents and activities, objects and tools, uses and meanings that constitute the practice.*" [15] In this study, we examine how individuals make sense of cloud services and incorporate them into their practices by applying their understandings of the services' remediated predecessors. As with other forms of infrastructure, these new services are "*always built on an installed base*" [4], so that individuals can scaffold their understanding of these technologies using experiences with their predecessors.

We begin with a brief description of the study and its participants. Next we organize our results according to two central themes: service adoption and reconciling collaborators' varying conceptual understandings the services. In the discussion section, we synthesize these stories into key findings, focusing on the relationship between cloud technologies and existing and developing social practices. We conclude with implications of our findings, highlighting key challenges in both the design of cloud-based services and opportunities to educate users how to better incorporate the cloud in their practices.

2 Study Description

We conducted semi-structured, open-ended interviews with 22 people who use Drop-box and Google Docs (and in some cases, Google Drive) in individual and collaborative situations. Table 1 summarizes each participant's pseudonym, age, gender, background, and cross-references which collaborators we interviewed. Participants were primarily U.S. based, although one participant was from Canada and another was from New Zealand; they had used the services for between 3 months and 6 years.

Table 1. Summary of participant characteristics and pseudonyms. Participants were recruited through the authors' social media networks; direct contacts were excluded from participation.

Pseudonym	Gender	Age	Occupation	Collaborators
Dean	M	33	Artist and arts administrator	James
Mary	F	27	Graduate student	
Melissa	F	37	Graduate student	
Nathan	M	29	Graduate student	
Lance	M	31	Web developer	
Aaron	M	31	Research assistant/Grad student	Sao, Xu
Trisha	F	31	Program manager at a non-profit	
Jacob	M	29	Interaction designer,	
Sao	F	30	Graduate student	Aaron, Xu
Karen	F	23	Graduate student	Steve
Sarah	F	28	Graduate student	Sanjay
Xu	M	30	Graduate student	Sao, Aaron
Martha	F	45	Program manager	
Dillan	M	49	Archivist	
Sanjay	M	25	Graduate student	Sarah
Otis	M	37	Systems analyst for an oil refinery	
James	M	36	Online public relations	Dean
Gary	M	51	Urgent-care physician	
Andy	M	36	Field service technician	
Steve	M	25	Software engineer /IT	Karen
Jayden	M	33	Graduate Student	
Bruce	M	40	Project lead/organizer	

Interviews lasted from 30 to 120 minutes (60 minutes on average) and were conducted over video chat or phone. The interviews focused on participants' adoption and ongoing experiences with Google Docs and Dropbox, as well as how their use of these services changed over time. Because some participants had been using either or both services for a long time, we asked them to find their oldest files in each service's store to help them recall early experiences. At the end of the interview, they were asked to summarize each product and compare them.

Because we were interested in how collaborators influenced each other's understanding and use of the cloud services, we asked participants if we could contact the other people involved in the stories they told during their interviews. Five of the 22 participants were recruited this way. In one case, we interviewed three members of the same group; the others were dyads (who were sometimes part of larger groups).

Since our initial participants (17/22) did not have significant experiences with Google Drive's recently introduced local syncing client, we specifically recruited a final set of participants (5/22) who had migrated to it. Given the substantial change in the conceptual model from Google Docs to Drive, we thought it was an interesting opportunity to see how users reacted to the availability of a local sync client. These interviews were conducted the same way as the others, except that Drive was included in participants' comparisons of cloud services.

All interviews were transcribed and participants were assigned the pseudonyms that are used in this paper. The interviewer (the second author) briefed the other authors about participant narratives immediately after the interviews took place. The analytical coding was performed jointly among the three authors; additional review of the transcripts was performed as necessary so that everyone was familiar with the details of the interview data. We performed inductive analyses of the interviews using grounded methods. Using open coding and memoing practices (as described in [5]), we grouped individual participant narratives into preliminary categories and labeled them. Labels and categories were in part emergent and in part influenced by our a priori interests in the adoption and collaborative use of these services, as well as the impact they had on existing practices. These categories were then refined using a constant comparison method that "*combines inductive category coding with a simultaneous comparison of all social incidents observed*" [10]. We also examined the interview data on a per-participant and per-collaboration basis to reveal adoption and use storylines and specific instances of remediation.

3 Results

In this section, we describe how participants develop their understanding of cloud syncing and sharing services and the ramifications of these understandings. We begin by exploring how participants adopt the services, either in response to problems that arose in the course of existing work practices, or through changes in the constellation of devices they use. We then show how the understandings and practices established during adoption affect ongoing use. Finally, we demonstrate how previous experiences with remote storage technologies produced some subtle misconceptions for participants, and examine the implications of these misconceptions for individual and collaborative use.

3.1 Adoption as Problem Solving

Participants described first learning about these services via online blogs (*e.g.,* TechCrunch), from their peers, or in the context of project work with colleagues. Most were vaguely aware of these services prior to using them. For example, Dillan and Lance both reported that they were first introduced to Google Docs when Gmail opened attachments in it. For most participants, however, meaningful adoption (as opposed to brief engagement) was the result of "upgrading" their way of doing things to overcome specific limitations or cumbersome aspects of their current practice, which in turn influenced and constrained how they interpreted the technology.

For example, Dean adopted Dropbox to overcome email restrictions. As a graphic designer, he often needs to send remote collaborators Photoshop and Illustrator files that are too large for email attachments: *"It was just an easy way to share files, rather than having to use YouSendIt, or WhaleMail, or any of those other sort of temporary big file e-mail systems."*

For some, the limitations are latent. Sanjay, for example, was aware of Google Docs' centralized approach and facilities for real time collaboration. But it was only after he started co-authoring a report with a colleague in another country that he found the motivation to adopt the service. Limitations may prompt the adoption of more than one cloud service. Sanjay later adopted Dropbox too so he could automatically share reports saved to his local file system.

As Sanjay's scenarios demonstrate, adoption was often spurred by collaborative project work. The initial collaborations reported were often small, *ad hoc*, project-specific, and rarely resulted in artifacts that were seen as archival. Even in more traditional professional environments, the collaborations remained lightweight (although adoption can be management-mandated). Trisha explained how her boss, who traveled, introduced her team to Dropbox, and how casual use became commonplace. Through use, Dropbox evolved from a literal dropbox to a shared repository.

> *We really use [Dropbox] a lot for… PowerPoint presentations that we can upload there so that when he's on the road … he has access to our latest figures… And it started out from just like "Oh, we're just going to use it for a couple communication things or a couple staff documents," and now we're basically constantly sending everything into Dropbox. It's become our de facto shared drive.*

Regardless of how individuals were introduced to the services, the motivation for their initial use shaped their subsequent understanding and expectations. Google Docs most frequently remediates editing suites like Microsoft Office, while Dropbox remediates storage and file transfer between devices or people, similar to uses identified by Dearman and Pierce [6]. As a result, subsequent use was often constrained to a narrow (albeit critical) set of cases.

3.2 Adoption in Evolving Device Ecologies

Not only did participants adopt cloud services to support existing practices when the old methods failed them; they also found themselves in situations in which evolving technological ecologies—including the introduction of new devices or the loss of old ones—prompted discovery and adoption of these services.

The introduction of new mobile devices seems tied to cloud service adoption. This observation is well-aligned with the results of Sohn et al.'s study [20]. Participant Martha purchased a tablet to support her graduate studies; this in turn prompted her to adopt Dropbox when she found she was unable to use a thumb drive to transfer files to the device. She had learned about Dropbox from a blog post returned by a search for "How to get files onto iPad." Adding a device created a problem that was solved by adopting a cloud service.

New devices can also spur new practices, which in turn provide an opportunity for incorporating cloud services. This adoption path was common for our Google Drive participants. Bruce learned about Google Drive after finding it preinstalled on his new Android phone; to his delight, he discovered that he was able to pull up an agenda on his phone and conduct a meeting without a laptop. Similarly, Andy, a service technician, explained how important Google Drive had become for on-site access to notes, manuals, and other documentation:

> *At first it was just something that I installed just to see what it was all about, and then when I realized I could … look at the same documentation on my laptop as I could on my phone… the phone has become this extremely useful tool.*

Device loss was another change that prompted adoption. In Nathan's case, the loss of his USB drive forced him to reconceptualize the function of Dropbox. He had already adopted Dropbox for backup, but he still moved files between school and home using a thumb drive. This function only became apparent, however, when *"I misplaced the USB at one point, and that's when I realized 'Oh, I'll just put it all on Dropbox.'"* Dropbox functionality did not change, but Nathan's practices (and thus his use) did.

New technology, as well as failures in existing technology, can promote adoption of cloud-based services and cause participants to reflect on what these services do. Hardware or new apps can present solutions to problems participants did not know that they had. Across our participants, we saw that understandings of these services most commonly broadened when participants' everyday practices were disrupted, and their focus was on the service itself, rather than the activity the service supports. By contrast, stand-alone information campaigns, such as those accompanying the introduction of Google Drive, did not seem to promote adoption or influence an individual's understanding of what Google's new service offered. Participants ascribed the migration to branding ("*the logo changed*") or found it inscrutable ("*something about it is slightly different and I can't put my finger on it whatever that is*"—Sarah)

3.3 Conditional Adoption

When participants described how they used the cloud, many saw the cloud through the lens of the limitations the services imposed and how they had changed the conditions of their work. This perception, in turn, either restricted the scope of their adoption (as when Google Docs replaced Word) or caused participants to compensate for disruptions in their normal practice (possibly by moving in and out of the service).

Limited by the Basics. Because Google Docs remediated desktop editing suites (and not a shared document repository), many participants lamented the service's limited functionality and polish, and did not use it beyond simple documents like brainstorming output, meeting notes, and roommate expenses. For a few like Martha, who owned a small business, Google Docs was not viable for anything beyond notes: *"[If] we're writing a proposal,"* she explained, *"it has to look professional."*

For those drawn into Google Docs for its synchronous collaboration functionality, a tension was created between it and the full-featured editor it was remediating. According to Karen, "*The main issue is formatting… I have to copy and paste it all into Word. When you're getting towards the final revision, [Docs] just isn't feasible*." Thus users adopted a cut-off point at which they would migrate from Docs back to a full-featured word processor to format the final document. However, this late-stage editing is often when the collaborative functionality of Google Docs is most needed. Some participants compensated for this by using Google Docs and shared Dropbox folders in tandem. Participants like Jacob described copying content from a Google Docs document into a Word file saved in Dropbox. Yet, while this approach retains some collaborative functionality, as both Martha and Melissa explained, using Dropbox this way limits concurrent access to the document.

Google Docs' inability to address the demands of pre-existing editing practices limited the extent to which participants were willing to commit to the service, and limited their willingness to acclimate to (and see the value of) Docs' novel functionality. Instead, participants were left weighing the benefits of Google Docs (often, the predicted contributions of collaborators) against its inability to format their content. The role of the remediated technology was clear: *"Real time collaboration in Word would be the ideal world. I can't think of anything more fantastic than that."* (Karen)

Acclimating to the Cloud. Participants discussed a variety of new practices they adopted to collaborate using cloud services. First, as other researchers have reported in their studies of shared repositories and workspaces [8, 18], the shared storage must be kept intelligible and consistent (so collaborators can find what they are looking for and negotiate practices to coordinate access so changes are not overwritten). Second, as we would expect from other studies of collaboration (e.g., [7]), collaborators must accommodate to a new level of visibility of their work and actions. This effect may become more pronounced when the collaborators are not peers (e.g., professors and their students, managers and their reports). The two effects are often intermingled—the document in its incoherent state is rendered abruptly available and visible by the new cloud services, either synchronously in Google Docs, or as an incomplete version when Dropbox syncs.

We observed multiple instances of how the remediated technology (in this example, Word) created an expectation of (and perhaps a genuine need for) change coordination facilities. Sarah described how her team's adoption of Google Docs resulted in breakdowns with one collaborator: "*She always was worried that her work wasn't there or that she was being overwritten and she wasn't sure who was writing what*."

Some participants coped with the perceived difference between their writing practices and those supported by the service by copying files in and out of the cloud. Sao, a graduate student, copied an important grant proposal out of the cloud and onto her local storage so her edits would not interfere with those of other collaborators. She explained that she prefers to do her initial drafts in Word, and then transfer her content into a shared Google Doc when she is ready to respond to feedback from her colleagues and make minor changes. Major edits, however, are always done outside of Google Docs:

> *Sometimes you will change a lot of things. And sometimes you want to take a long time to think about what you're going to write… So I first write on my own Word document and then later copy and paste to Google Docs.*

Many participants commented on the authoring styles of their collaborators, which were abruptly rendered visible by the real-time collaboration in Docs, as well as the changes it caused in their own. Sanjay, while a strong advocate Google Docs' collaboration functionality, explained that under some circumstances he felt like he was under surveillance: *"Let's say the deadline is tomorrow, the advisor is doing his part, and there is still things to do in my part. I can't sign off… it might look bad."*

3.4 Troubled Conceptions

If participants' understandings of cloud syncing and sharing services develop in the context of the problems they solve or the practices and technologies they replace, and collaborative adoption proceeds in a hand-to-hand, viral way, then we might wonder about the effects of misconceptions: Do they hinder adoption? Do they limit successful use? Are they transmitted with the services? Do they lead to asymmetries in adoption, thus interfering with the overall benefit of services [12]? In this section, we explore the sources and effects of misconceptions.

"Cloud" Just Means Remote, Right? Thinking that files synced via Dropbox or Google Drive were not stored locally was a surprisingly common misunderstanding if participants had previous experience with remote or network storage. Dean, a graphic designer, harbored a misconception that stemmed from his prior use of an FTP server that he had set up to share files with his collaborators. Initially he described Dropbox as a portal between his computers, but later in the interview, he revealed that he uses Dropbox to *"keep as much as possible off of my local hard drive"* to prevent his primary computer, an aging laptop, from getting too bogged down, and that he assumed files that appeared to be local in the Dropbox folder actually lived in the cloud:

> *I assume that all of that shit lives on Dropbox's servers in the cloud and those [file icons] are basically links to the files... And if I just double click this text file I'm opening my text application and it's pulling the file from Dropbox's server. But the file is not living locally on my machine.*

How do these misconceptions that an individual brings from experience with a remediated technology play into subsequent collaborations? Are these misconceptions corrected through interaction with collaborators, or do they remain in the background? We examine data from collaborations to find out.

Conceptions in Collaboration. In individual use, incorrect or incomplete understandings of cloud services appeared to have limited implications. However, we had anticipated that these misconceptions would come to light during collaborations where users might encounter and be held accountable to the conceptual models maintained by their collaborators.

While we suspect that these conversations occur, the relative number of successes our participants reported in their collaborations struck us. One such surprise came when we interviewed Dean's primary collaborator, James. While James also had a somewhat incomplete understanding of Dropbox, his practices around the tool avoided any conflict with Dean. James copied files in and out of their shared Dropbox folder, and only edited them after copying them to a location outside of Dropbox. This practice enabled him to have a master and a renamed version with his changes:

> *If I'm editing a poster that Dean and I are working on together, if he puts [the file]… that he just touched, in [Dropbox], I'll copy it to my desktop and then open it in Photoshop. I want to preserve his changes, [but] I want to be able to delete stuff, or just have [a] backup, or make my changes. So I'll leave it [the original file], so it's like a primitive version control.*

James subsequently complained that *"Dropbox is… a little too much like a drawer, where something is either in or out of the drawer."* Because James never edited a file in place in the shared Dropbox folder, he effectively avoided the very kind of version conflicts that might have caused Dean to reflect on his theory that he was working on a separate downloaded version of a remote file. Although these compatible misconceptions remained invisible in Dean and James' collaborations, the proliferation of files in their shared folder effectively exacerbated any problems Dean was experiencing as a result of a full hard drive.

By contrast, another collaboration (an academic research group that included participants Karen and Steve) negotiated a social workaround only after incompatible conceptions and misconceptions surfaced in a series of breakdowns. The group had reached a situation in which co-authors were working with out-of-date versions and overwriting one another's changes. These conflicts grew out of incompatible conceptual models. For Karen, Dropbox remediated a practice of emailing files to herself to ensure she had up-to-date versions on all of her computers:

> *I have a computer in my lab at my university and I was using a laptop and then my personal desktop. ...Usually I was e-mailing things to myself. And if I forgot to e-mail it to myself or I e-mailed the wrong version then I was constantly redoing work I had already done because it was on the wrong machine.*

On the other hand, Steve (an undergrad who eventually administered the group's Dropbox folders) modeled his understanding of Dropbox on version control systems such as Subversion[1]. Subversion helped Steve make sense of how Dropbox syncs local and remote files, but also left him with the mistaken belief that Dropbox sometimes merged changed versions of files:

[1] `http://subversion.apache.org/`

The only thing… is [Dropbox] didn't really merge conflicts very well… if you and I are working on the same document at the same time and we both save it and it can't merge it together, it will save your copy as [AUTHOR NAME]'s conflicted copy and mine as Steve's conflicted copy.

Steve went on to provide a detailed (albeit mistaken) account of how Dropbox attempts to automatically merge changes and only produces conflict files when it is unsuccessful. This is similar to how code repositories like Subversion work: when changes are committed to the repository, any changes that can be merged automatically are. When they cannot be, users are asked to resolve conflicts before committing a particular file.

While Karen, who was also familiar with version control systems realized, "*there's no real version control to merge things back together*," Steve believed that the others in the group harbored misconceptions about Dropbox's ability to merge changes:

One of the professors…that [Karen] and I both worked under…didn't seem to understand that Dropbox was capable of merging. And so she would open up a document and resave it with her initials appended to it. And so over time you'd have eight documents with her initials on it and then somebody has to go through and merge her stuff into the original document when all she really had to do was open the original document. So she's a good example of not using it properly.

This was a complicated problem that involved both social components (the co-authors were not coordinating changes with one another) and technical ones (group members were syncing infrequently, in part because they had varying understandings of how to take better advantage of Dropbox). A breakdown over an important document (a grant proposal) led to the adoption of an explicitly social solution that did not resolve *any* (mis)conceptions, as Karen explains:

We have to call and say I'm working on this portion. This is how I'm going to save my folder or the document and what the name will be. … So we actually include initials and time stamps and file names so that we have a way of tracking [who did what when].

Steve explained the resolution to the version conflict problem: in exchange for additional personal space on Dropbox, he became "*the official Dropbox guy*" who "*had to merge crap all of the time. So once a week I'd go check it and say oh, you know, there's eight conflicts… and I would get rid of all of those.*" Just as in James and Dean's case, it was easier for participants to develop practice-based workarounds than it was to revise their mutual understanding of the cloud service. But in so doing, they missed an opportunity to learn how the cloud could better support their collaboration.

4 Discussion

What stands out in our results is the primacy of practices. They not only spur adoption—people adopt the cloud to solve problems—but they also suggest that people may use the cloud without really understanding how it works. In some ways, this is unsurprising: naturally people adopt practices, not features. And often they are using the cloud to remediate or maintain a practice, not a technology. But what we see is the consequential nature of *what the cloud is replacing*. If cloud syncing and sharing services are replacing FTP servers, then users are far more likely to understand the new service in terms of the old one. Their expectations are thus set, sometimes in ways that make them resistant to evolving understandings or extended use of these services. In this section, we reflect further on the relationship between user practices and cloud services, and examine more closely what this says about user education, addressing misconceptions, and the conditions for successful adoption and continuing use.

4.1 Practices Overshadow the Cloud

Across the board, participants reported that using Dropbox and Google Docs was straightforward and that the services were easily understood. However, participants often failed to see opportunities to expand their use of these services. This is in part because they focused on their work instead of on the tools they were using to do the work. Trisha provided an example of this tension when she explained that Dropbox would enable her team to maintain a single versioned master of a document "*…instead of having 17,000 different versions of a single file as you update it, in theory there's one, although we actually have not been using it that way.*" Despite her awareness of this capability, her group continued to name versions: "*it'll say like 'edits by' and then initials at the end, which is kind of silly… because then it's like you're using way more space than you need, but once people start doing something one way and get used to it, it's really hard to make them stop.*"

Google continues to add novel functionality to make its collaborative environment more robust. Yet, participants reported their dismay at the lack of support for the practices they already have – in the case of Google Docs, authoring content in a sufficiently robust client – more than their excitement for new modes of working. Our findings suggest that the inability of Google Docs to meet the standards of the technology it is remediating (i.e., the desktop office suite) limits most participants' desire to replace their existing tools with Google Docs, and for many, their ability to benefit from the collaborative features Docs provides.

Although these services have novel functionality that individuals' stories suggest they would find beneficial, they were largely unaware of this functionality. Instead, users focused on the essentials as defined by their tasks: Formatting and speed in Google Docs, and sharing files and managing storage quotas in Dropbox.

Understanding via Remediation. Bolter and Grusin assert, "*Each new medium is justified because it fills a lack or repairs a fault in its predecessor, because it fulfills the unkept promise of an older medium.*" [2] Participants turned to remediated practices and technologies to make sense of the cloud. Projecting their understanding of how prior technologies enabled them to accomplish their practices onto how the cloud operates shapes the way they can understand the novel features that cloud-based services offer. The users' focus on their particular practice of interest narrows their view of what cloud services can do, which can in turn contribute to an incomplete or inaccurate conceptual model of how it works.

When to Educate Users. The ideal time to educate users about new functionality is when their focus is on practices associated with the tools and tasks in which they are engaged. For example, practices around using the tool, such as setup or maintaining the total storage within the tool's free quota, are when the user's attention is on the tool and not trying to accomplish a productivity task. Mary illustrates this point in describing how the Dropbox scavenger hunt helped her understand how to use it:

> *I told him to do the scavenger hunt, because not only do you get like the free storage, which was great, but then you also get to learn about like how to use it in like kind of a fun, interesting way where you're not just watching like a video, no matter how exciting they try and make it.*

This Dropbox technique not only catches users when they are focused on managing the tool, but also gets users to enact practices they have learned (rather than just watching a video of them) to earn more storage space to ease their management work.

By contrast, Google's in-browser notifications about the transition from Docs to Drive, for example, were largely dismissed when they were encountered in the process of going to Google Docs to work on a file for a task. Furthermore, the wording of the transition notification ("Google Drive is the new home for Google Docs") suggests that no new practice is needed to migrate from Docs to Drive. Consequently, we found most users who had migrated to Drive were still using it in much the same way as they used Docs, without taking advantage of the new features that Drive offered.

Users were more likely to learn about the new features in Google Drive when they installed the mobile app (or found the app preinstalled) on their Android phone. At this point, they were focused on integrating a new mobile device or feature into their practices, making it an opportune time to learn more about how the cloud could help them. Taken together, these stories identify different types of teachable moments as users go about their work or as they manage cloud-based services.

4.2 The Social Effects of Clouds

Our previous study explored models [16]. For example, one collaborator would rely on a synced folder for archival storage, and the other would do housekeeping and delete a portion of this implicit archive. In this study we sought multiple perspectives on the same collaboration to breakdowns in participants' collaborations that could be attributed to the collaborators' conflicting conceptual see how they worked through these breakdowns.

While we continued to find examples of collaborators' conceptual mismatches, in this study, we were struck by how the participants worked around or managed their misconceptions. Generally, we found that many of the workarounds were social, and did not necessarily involve reconciling conflicting models. For example, Steve was designated as "*the Dropbox guy*" to resolve version conflicts and merge changes. It is probable that neither Steve, Karen, nor the professor revised their conceptual models.

We expected (or hoped) that encountering these kinds of problems as a group might provide teachable moments during which someone would invest the effort to understand how things are supposed to work and educate the rest of the group. Instead, we largely found that groups reverted to familiar practices that relied less on the cloud's ability to sync and manage changes and more on manually or socially managing updates, which they understood how to control. Thus, instead of refining or updating their collective conceptual model of cloud functionality, groups would construct social explanations and practices that avoid the features that they did not understand.

Compatible Misconceptions. In scenarios like Dean and James's, we see how two collaborators operate with conceptual misunderstandings that, nonetheless, do not result in breakdowns. While compatible misconceptions may be innocuous, during some interviews, it was clear that these misconceptions would likely lead to a breakdown in the future. For example, groups whose members believe that Dropbox supports simultaneous editing will eventually encounter conflicts where changes will need to be merged by hand.

Ad Hoc Collaborations. The majority of participants' cloud-based collaborations would best be described as "ad hoc." Since the cloud easily transcends organizational boundaries and firewalls, it is well-poised to facilitate ad hoc collaborations; in fact, the ability to cross firewalls is often a precondition for the adoption of syncing services [17]. As a result, two issues are worth noting: First, while Dropbox and Google Docs are often most productively used by groups, users may initially adopt these services individually.

Second, while these cloud services are often used outside traditional enterprise structures, they rely on collaborators having accounts with the services' providers. Participants often talked about the ease of using Dropbox and Google Docs for a quick project because *"everyone had an account"* (Aaron), but this technological ease can violate some implicit seams that participants use to productively distinguish one use venue from another. Trisha, for example, noted concerns about tools that spanned across work and home:

> *I didn't want to necessarily use it for my own personal stuff and accidentally save something in the work folder. Because when we first signed-up for Dropbox [at work], I was using it with my personal address because I already had an account, and I just didn't want to mix work and personal life more so than I'd already been doing.*

These large-scale services appear to be enabling a broad set of new collaborative practices, but also present challenges for traditional enterprise work situations in which these services might not be viable. Perhaps to the dismay of network administrators, many participants in traditional enterprise settings spoke of using these tools to circumvent organizational policies to, as Lance said, *"get things done."*

5 Implications

The adoption of Google Docs and Dropbox and the conceptual understandings individuals develop of the services highlight issues around when and how systems make features and processes visible, as well as how to connect these with users' practices. The drive towards seamlessness can leave users ill-equipped to adequately understand the nuanced behavior of their tools or make appropriate decisions when they encounter problems [19]. We make two arguments on this front: First, designers of cloud-based services can improve user experiences by exposing technical functionality and processes in relation to people's practices. Second, service providers can facilitate the development of cloud-based practices by taking advantage of teachable moments to educate users about their functionality.

5.1 Seeing the Cloud through Practices

Cloud services need to aid users in connecting their practices with what these cloud services are actually doing. As previous research has noted [16], the lack of process transparency may prevent individuals from broadly adopting cloud-based services. Our findings suggest that simply exposing the functional behavior of these services may be insufficient. Instead, designers should provide more robust user feedback on processes but in relationship to user practices and their working context and not in terms of system processes or features.

The Google Docs text editor provides an excellent example of how system processes can be exposed in a way that aligns with user practices. The addition of co-authors' cursors in Doc's word processor, for example, and synchronous appearance of a coauthor's text provide little room for ambiguity about the collaborative editor's functionality: co-authors are changing the same document. When and how Google Docs saves content is less clear. When connected to the Internet, saving is unimportant: Google automatically saves content as it is added. In fact, Google has removed "Save" buttons and menu options from their editors. Saving becomes ambiguous, however, when individuals proactively want to save their content – for example, when the user's Internet connection has been lost. In this case, Docs notifies the user that it is trying to reconnect to Google's servers, but does not explain how changes made in the meantime will be merged, or what options individuals have to preserve their work if the interruption in network connectivity persists.

Turning to Dropbox, we see a good example of how the process of creating a shared folder connects with user practices. When users share a folder (either directly on Dropbox's website or via their operating system's context menu), the user is presented with a webpage on which they can provide the email addresses of those they would like to invite to the folder. An invitee, upon receiving the email invitation, can click a link to accept the invitation, at which point the invitee is notified about the addition of the new shared folder both in a web browser and the local syncing client. When sharing a folder, the user's practice (e.g., email invitation, files appearing in a folder) and the system's functionality are highly aligned.

However, once a folder is shared, functionality is far less clear. A number of participants described confusion about the relationship between their Dropbox folder and the rest of their hard drive. Dragging a file from Dropbox to another folder moves the file rather than copying it. James' comment about Dropbox acting "*like a drawer*" stems from this interaction. What is not immediately clear to users is that moving a file outside of Dropbox affects all collaborators – it effectively removes the file from everybody's Dropbox. Moreover, this breakdown was compounded when participants found social explanations for technical breakdowns. Dropbox could address this confusion by simply alerting users that moving files out of shared folders will remove access to those files by other collaborators.

Our implication that processes should be exposed relative to user practices presents some significant challenges for services that operate at either the system or the interface level. Dropbox, for example, intentionally limits the focus of its syncing client to the file system and remains agnostic to programs and user practices outside of how they read and write to the hard drive. Yet, many of the collaborative practices around using Dropbox involve editing those files using productivity tools. Since Dropbox's implementation does not afford providing any process transparency in the context of those editors, users can get confused about how Dropbox manages concurrent or conflicting changes while editing those files. Conversely, the web-based editors in Google Docs made it transparently obvious when concurrent editing occurred. Yet Google Docs removed commands for file management (no Save command), making it unclear how to save work when an Internet connection is interrupted or lost. Cloud service designers need to consider the full scope of user practices to help users understand how to fully integrate the cloud's capabilities into their work.

5.2 Learning How to Look at Clouds

Given the scope of changes to cloud syncing and sharing services over the last year, we expected to see user practices and understandings change in response. Instead, we found that participants' conceptualizations of these systems remained relatively unchanged. Since users' practices have been established and largely remained stable, changes in the cloud services by themselves were not enough to engage users in the work of learning about them.

Revisiting an earlier example, it was noticeable how most of our participants – active users of cloud services – were unaware of the particulars around the transition from Google Docs to Google Drive, despite prominent notifications. Indeed, given the lack of immediate impact on their practices, many thought that the change to Drive was simply a rebranding of the service.

If our aim is to enable broader use of the cloud, our findings demonstrate that improving the design of cloud services alone is insufficient. Service providers must also educate users about how their practices can take greater advantage of the cloud. People may not notice new features even when they are prominently announced. Moreover, the rate at which the cloud services are changing challenges users' ability to keep up with those changes or even identify "expert users" to help work through conceptual breakdowns, as they have in the past [14]. Our study identified four different kinds of teachable moments when users were more amenable to learning about new features or capabilities of the cloud. We go through each type in turn.

Teachable Moments Around Adoption. Adoption presents a clear period during which initial understandings are being formed. Critical windows for forming understandings are not limited to service adoption. The addition of new devices also caused participants to reassess their use of cloud-based tools. Indeed, the most successful instances of Google Drive adoption reported by our participants were the result of discovering the Drive application pre-installed on a new mobile device.

Teachable Moments around Remediation. Services should expect and account for the fact that new users bring existing practices with them. By understanding which prior technologies and practices a cloud service remediates, providers can anticipate user expectations that can inhibit deep adoption; in this way, they can also prevent misconceptions. Perhaps providing different on-ramps tailored for the common tools that are being remediated by the cloud would help connect users' practices with cloud features of most interest to them, and would help anticipate and circumvent many standard misconceptions.

Teachable Moments around Cloud Management. User understandings were pliable when practices were focused on managing the services themselves. Dropbox's disk quota, for example, presents an overarching user problem that requires that heavy users "maintain" their storage from time to time. Dropbox rewards users with additional storage for completing a variety of tasks, most of which introduce users to Dropbox functionality or prompt them to tell others about Dropbox. Numerous participants described learning about Dropbox from its "Tour", "Simple tasks", and "Scavenger hunts." Dropbox takes advantage of a limit that they have imposed on their users to encourage them to learn more about the service and teach others about it.

Teachable Moments around Collaboration. Collaborating with others naturally exposes other people to the consequences of actions in the cloud. Updating or deleting files in the cloud will affect what others see. These collaboration consequences could be used to help users understand how the cloud is mediating their collaboration. Besides the visibility of concurrent editing in Google Docs, there are other ways in which the cloud could create teachable moments during collaboration. File deletion presents an opportunity to show users the effect of this action on their collaborators (i.e. they will no longer have access to it). Alerting users as to when and why conflict resolution files are created would help users understand the limits of synchronization. A "sync inspector" that more transparently indicates who last edited the file (and on which device) would help users have a better sense of how changes from various people and devices are being coordinated through the cloud (Cimetric [17] is an early attempt at providing such a sync inspector). Browsing sync history could also help users diagnose why they do not have access to the expected file version (for example, the one they saved on another device), since they might be able to identify when a sync was missed, which might help them narrow down possible causes for a sync failure. While collaboration naturally exercises many of the cloud features needed to keep files in sync, users need more help to observe and understand these processes to appreciate how the cloud is working to support collaboration.

6 Conclusion

Our study demonstrates the ways in which individuals adopt cloud-based syncing and collaboration services into their work practices. We found that practices from remediated technology fundamentally shaped the ways participants understood and used cloud functionality. Individuals adopt cloud services to solve problems or when incorporating new technology into their routines. However, when cloud services do not fully support users' existing practices, they may only partially adopt cloud functionality. Moreover, given the limited transparency of cloud processes, experience from remediated technologies or collaborating with others who have mismatched conceptions often resulted in misconceptions that also limited their use of the cloud.

We identify opportunities to improve cloud experiences, both by redesigning aspects of the services and by appropriately educating users about how their practices might benefit from the cloud. Designers of cloud-based services need to strike a balance between visibility and seamlessness. Decisions around visibility should be made based on users' practices and not the system processes. Perhaps the more salient opportunity stems from identifying specific teachable moments for educating users how to more fully incorporate the cloud into their practices. We identified these teachable moments as they occur during adoption, remediation, management, and collaboration. As cloud services and users' cloud practices continue to develop, we expect that these implications will help to improve current and future cloud services.

References

1. Biggs, J.: Google Drive Is Live With 5GB Of Free Storage And Google Docs Hoo-kup. TechCrunch (2012), `http://techcrunch.com/2012/04/24/google-drive-is-live/`
2. Bolter, J., Grusin, R.: Remediation: Understanding New Media. MIT Press, Cambridge (1999)
3. Bolter, J.D.: Formal Analysis and Cultural Critique in Digital Media Theory. Convergence: The International Journal of Research into New Media Technologies 8(4), 77–88 (2002)
4. Bowker, G.C., Star, S.L.: Sorting Things Out. MIT Press, Cambridge (1999)
5. Corbin, J., Strauss, A.L.: Basics of qualitative research. Sage, Newbury Park (2008)
6. Dearman, D., Pierce, J.S.: "It's on my other computer!": Computing with Multiple Devices. In: Proc. CHI 2008, pp. 1144–1153 (2008)
7. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: Proc. CSCW 1992, pp. 107–114 (1992)
8. Dourish, P., Lamping, J., Rodden, T.: Building bridges: Customization and mutual intelligibility in shared category management. Group 1999, 11–20 (1999)
9. Dropbox. What is Dropbox (2012), `https://www.dropbox.com/tour/1`

10. Goetz, J.P., LeCompte, M.D.: Ethnographic Research and the Problem of Data Reduction. Anthropology & Education Quarterly 12(1), 51–70 (1981)
11. Google. An overview of Google Docs (2012), http://support.google.com/docs/bin/answer.py?hl=en&answer=49008
12. Grudin, J.: Groupware and social dynamics: Eight challenges for developers. CACM 37(1), 92–105 (1994)
13. Hamlen, K.W., Liu, P., Kantarcioglu, M., Thuaisingham, B., Yu, T.: Identity management for cloud computing: Developments and directions. In: Workshop on Cyber Security and Information Intelligence Research (2011)
14. Kellogg, W.A., Breen, T.J.: Evaluating user and system models: Applying scaling techniques to problems in human-computer interaction. In: CHI 1987, pp. 303–308. ACM Press (1987)
15. Lanzara, G.F.: Remediation of practices: How new media change the ways we see and do things in practical domains. First Monday 15(6-7) (June 2010)
16. Marshall, C., Tang, J.C.: That syncing feeling. In: Proceedings of Designing Inter-active Systems Conference (DIS 2012), pp. 544–553. ACM Press (2012)
17. Marshall, C., Wobber, T., Ramasubramanian, R., Terry, D.: Supporting Research Collaboration Through Bi-Level File Synchronization. In: Proc. Group 2012, pp. 165–174. ACM Press (2012)
18. Rader, Emilee, Yours, Mine, and (Not) Ours: Social Influences on Group Information Repositories. In: Proc. CHI 2009, pp. 2095-2098 (2009)
19. Ratto, M.: The Ethics of Seamlessness: Resources and Future Directions. International Review of Information Ethics 8(12), 20–27 (2007)
20. Sohn, T., Li, K.A., Griswold, W.G., Hollan, J.D.: A Diary Study of Mobile Information Needs. In: Proc. CHI 2008, pp. 433–442 (2008)
21. Weiser, M.: The computer for the 21st century. Scientific American 265(30), 94–104 (1991)

Awareness, Transience and Temporality: Design Opportunities from Rah Island

Pedro Ferreira[1], Pedro Sanches[2], and Alexandra Weilenmann[3]

[1] Kungliga Tekniska Högskolan, The School of Computer Science and Communication, 100 44 Stockholm, Sweden
[2] Swedish Institute of Computer Science, Isafjordsgatan 22/Kistagången 16, Box 1263 164 40 Kista, Sweden
[3] University of Gothenburg 106 91 Göteborg, Sweden
pedro@mobilelifecentre.org, sanches@sics.se, alexandra.weilenmann@gu.se

Abstract. This paper deals with the implications of the socialness of private communication. Drawing upon ethnographic observations of first time mobile phone users in Rah, an island in Vanuatu, we revisit the debate on how the mobile phone reconfigures private and personal communication. Our observations show how the advent of the mobile phone disrupts and challenges existing practices around how private communication is managed on the island. These observations are used to open up a design space where we explore the socialness of personal, private communication. Drawing on the analysis, we discuss three directions for future thinking of mobile interaction design: (1) designing for spatial awareness; (2) designing for transience and (3) designing with temporality. We expand on these to discuss the notion of digital patina, which we argue, is an exciting topic to explore for the design of personal, social communication.

Keywords: Privacy, personal communication, social communication, transience, temporality, translucence, awareness, design.

1 Introduction

Mobile telephony has opened up for different forms of socio-technical reconfigurations wherever it has sprung up. We have seen a variety of reports and publications that account for the ways old forms of socialization and power are challenged, opening up for radically new forms of engagement, which are still developing and taking different shapes. Studies from places as distant as Sweden [1, 2], the Philippines [3], Japan [4], Jamaica [5] or Israel [6] among others, despite providing us with rich and diverse pictures, also point to similar, emerging behaviors, some of which we will follow up on in this work. One such aspect that has been suggested early on was that mobile phones would allow for more private communication, since it was thought of as a personal device, rather than, for instance, a device going into people's homes or workplaces [7, 8].

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 696–713, 2013.

In this paper, we wish to revisit this discussion, and in doing so we draw upon a set of observations from an ethnographic study conducted in Rah Island, Vanuatu, a nation in the South Pacific. Particularly, we observed how the mobile phone reconfigured private communication on the island, thereby disrupting and challenging existing practices for how people communicated. The observations from the field allow us to rethink the way that privacy is considered in relation to mobile phone use, and enable us to open up a design space dealing with the socialness of private communication. We use this argument as an inspiration to consider design implications on how to explore the socialness of this type of communication.

We will analyze two situations in which mobile communication was at odds with expectations on privacy: the persistent nature of digital data and the layering of different spaces in communication. Using three concepts from Social Translucence: *visibility*, *awareness* and *accountability* [9], as a lens, we argue for three design domains worth exploring in HCI: spatial awareness, temporality and transience. We conclude by introducing the concept of digital patina, as a possible consequence of the exploring the domains introduced, and an exciting one to unpack in HCI.

2 On the Study and Approach

The data used as a starting point for our discussion in this paper are part of a larger ethnographic study conducted in Rah Island, in Northern Vanuatu, in February 2010. The data was analyzed with the help of colleagues, through different sessions (or workshops) where we analyzed bits of data trying to map out what the main insights for HCI could come from this fieldwork. Other papers have been published on this fieldwork, focusing on aspects of bodily orientations around mobiles [10] and playfulness and mobiles [11]. The data and insights presented in this paper are however mostly undocumented in our previous publications. Before we move on, we will provide a framing for the study conducted, both in terms of the location, the source of the data gathered as well as our particular approach to dealing with ethnographic data in an interaction design context.

2.1 Studying Mobile Phone Use on Rah Island

Rah Island is a small island of around 189 inhabitants [12] in Northern Vanuatu, a Melanesian country in the South Pacific. A cellphone tower had just begun working in the area, the very same day the first author, who conducted the fieldwork, arrived in the island, The tower was placed in Motalava, a larger island almost adjacent to Rah, where one could walk across during low tide, as can be seen in Fig. 1. This meant that the almost three weeks the researchers spent in the island were the first weeks of GSM connectivity on these islands. Perhaps worth noticing is the fact that many mobile phones were already present in the Island, at this time, given that they were sold by the network operator months before, as part of a network launching promotional package. When the tower began operating, many phones were already present, awaiting connectivity.

Fig. 1. Rah Island, as seen from the shore in Motalava

Rah and Motalava are two islands where, at the time of the study, there was no centralized means of electricity production and where people lived mainly of their own agricultural practice with only a small minority disposing of regular financial income [12]. The first author spent a total of around three weeks in the Island, during which he collected observation data, including around a couple of hours of video, almost two hundred photographs and around 30 pages of fieldnotes. He also conducted four semi-structured interviews, though there was some difficulty in attracting people's time and attention for long amounts of time. The quotes used in this paper are thus mostly from notes taken during informal conversations, often in groups of more than one person. Some of the data comes from different local meetings, usually held at the request of the local community, either for us to explain the purpose of our stay in Rah, or as a way for people to express different concerns and requests. These meetings were first held under the assumption, which we immediately corrected, that we were coming there on the behalf of the telecommunications company. This entailed that they were particularly keen on expressing concerns and requests.

In this work we focus on the potential contributions that this particular ethnographic fieldwork might bring to interaction design practice and research. The use of ethnography in HCI has been the object of much controversy [e.g 12, 13]. We take a particular stance, which is very much design-oriented rather than ethnographic-centric, concerned more with advancing the field of mobile interaction design, rather than theoretical advances in ethnography or advancing the knowledge of a particular cultural setting. Thus, we focus on particular bits of data, which inspired our design thinking and informed our discussion rather than presenting a comprehensive ethnographic account or "thick description" [15] of the culture and practices in Rah. We focus on drawing broader inspiration for design [14], in someway pursuing "ethnographically inspired design", instead of following a more "ethnographically informed design" approach in the strict sense, as if ethnographic work would serve as a sort of "requirements gathering".

In fact, the informants that will appear in this work are not even seen here as the ultimate benefiters (or users) of any potential system, or specific design contributions. Rather, in this paper, the fieldwork is drawn upon in order to inspire thinking about design in HCI, by unpacking mundane aspects of mobile phone usage in relation to everyday privacy concerns and management of secrecy, expressed and illustrated by observations in Rah. It is more a matter of uncovering somewhat hidden or overlooked opportunities within design oriented HCI, which can add value to designing for a wide variety of settings, practices and systems. In making this connection, we have relied upon *visibility*, *awareness* and *accountability*, all key concepts within HCI and CSCW, as formulated by Erickson and Kellog in their *social translucence* [9] approach. This will be expanded upon in the next section.

2.2 Social Translucence as a Design Approach: Visibility, Awareness and Accountability

Social translucence [9] is a systems design approach developed over a decade ago, as a method to design collaborative systems for large groups [9]. Three concepts are introduced which we used to take our observations and move into a design reasoning process: *visibility*, *awareness* and *accountability*. Ericksson and Kellogg illustrate social translucence with a metaphor drawn from the physical word: a *glass window* on a door, as a way of mediating the social interaction between the outside and outside users of this same door. In the case described by the authors, the glass window comes as a solution to the problem of knowing whether someone is standing outside the door before one attempts to open it, potentially bumping the door against the other person. By having the glass window, they argue, one is able to see on the other side and thus be able avoid this inconvenient incident.

Visibility, the first of the concepts, refers to the socially significant information, which is now available to people on both sides of the door, that is, the person on the other side. That is a property of the glass window within the context in which it was installed. Visibility, the authors argue, allows for the emergence of *awareness* and *accountability*, which will be responsible for the upkeep of the social conventions.

The *awareness* that someone is on the other side of the door, helps avoid the inconvenient situation, as the authors put it: “we have been raised in a culture in which slamming doors into other people is not sanctioned” [9]. Thus, by knowing there is a person on the other side, one is made aware of a potential awkward incident, and so could be persuaded to avoid it.

This brings us to the third concept, that of *accountability*. Accountability means that each person is aware that the other is aware. Or, in the authors’ words: “I know that you know that I know you’re there and therefore I will be held accountable for my actions” [9]. So accountability refers to the fact that, regardless of whether each of the people involved in that situation (of coming face to that door and ultimately deicing to open it), are not only already *aware*, but they are aware of the other’s awareness, and so deciding to open the door at that precise moment, and ignoring the others’ presence, is not only breaking a social norm, but is doing so, knowing the other might hold them accountable.

We found these principles useful for looking at situations regarding privacy and the violation of secrecy around mobile phones. Particularly, we were inspired by their questions: "What might it mean to have social translucence in a digital system? How might making social information more visible actually change the way digital systems are used? Why might this be a desirable thing?" These were the main questions, which drove our process of harnessing Social Translucence as a generative concept for HCI, when looking at situations which are not apparently, or strikingly social.

3 Privacy, Secrecy and Mobile Communication

Privacy is a long-standing concern within the field of human computer interaction. As argued in the introduction, we wish to draw particular attention to the implications of the social character of private communication. It has been argued, as the mobile phone was introduced in different settings, that it allowed for more private communication [e.g. 4, 7, 8]. In contrast to this body of work, more complex pictures have been documented, for instance when it comes to sharing of mobile phones. Of particular importance for our discussion here is the body of work, which challenges this strict notion of mobile phones as an individual technology [2] noted how young people had various ways of sharing their personal devices. By sharing information on their phones, by passing them around or reading aloud, these young people had found interactional strategies to integrate their phones into their social activities, in ways that challenge the personal nature of this device. This behavior was such an integrated part of their ways of using their phones that those sending messages oriented to the fact that others, than the owner of the phone to which the message was being sent, might potentially read that message. In a similar vein, there is a body of work describing text messaging among this younger age group as a form of gift giving [16], providing further evidence to the social character of the mobile phone. Other studies show how communication on the phone is not just a social situation involving the person you are communicating with, it is also a social situation in other ways. When we are talking on the phone we sometimes withdraw from the current situation, but it is important to recognize that also that action is a social action [17–19].

Obviously, these types of social and sharing behaviors around personal devices and information call for more different interactional mechanisms if one is worried about preserving a certain degree of privacy, or secrecy, in relation to one’s data and communication. We will return to this point when discussing our material from the field.

Finally, it should be added that while privacy is a complex and loaded concept within HCI, we use the term here in its most commonsensical and mundane sense. “Privacy” and “secrecy” are used interchangeably referring to the same phenomenon. Along these lines we will also refer to terms such as “eavesdropping” in the commonly understood sense of the term, as defined by Black’s Law Dictionary: “the act of secretly listening to the private conversation of others without their consent”. We are not referring to advanced technological eavesdropping techniques or concepts also used in HCI or Computer Science.

4 Dealing with New Means of Private Communication on Rah

At the time mobile connectivity was launched in Rah, there was already one landline post existing in Motalava, situated fairly close to the shore connecting to Rah. The landline had been in use for several years, allowing for phone communication to and from outside the island. The arrival of mobile phone connectivity was something that extended those possibilities by quite a bit. It was now possible to perform phone calls from within the islands as well as calls to and outside the islands from different locations. These possibilities, in particular of intra-island communication was seen with some wit some concern, such as Fajo who wondered: "*who knows what the younger ones are conspiring? […] It is a very worrying situation*". So even if one could see many manifestations of joy over the arrival of mobile telephony, it was also not rare to hear these types of concerns. As in any society, these new technologies, and modes of sociality between people defied the current state of things and prompted both happiness as well as concern.

These concerns are supported by studies in other contexts, showing for instance, how mobiles have been used by teenagers to create private/secret arenas for communication outside of parental control [3, 7]. This increase of secrecy, as a side effect from the introduction of mobile communications was not the whole story, as we will illustrate in the following examples. We will focus on two situations, which inspired us to explore this domain: the layering of the physical space, with the virtual space of mobile communication and the challenges presented by the persistent nature of digital information.

4.1 Layering of Spaces

As we arrived in Rah, we were conducted to a town meeting so that we could openly explain what our study was about, listen to people's concerns and suggestions, as well as answer several questions people had regarding mobile telephony, given its very recent introduction. The participation was very broad, and lasted for quite a long time. Once people were seemingly satisfied with our answers and justifications we headed back to our bungalows to sleep. At this moment, one of the locals came up to us, in a very secretive fashion and speaking in a very low tone, as if concerned someone could overhear us. This person expressed concerns over the lack of privacy he had experienced with mobile phones: "*Before, with TVL [the landline infrastructure was also provided by TVL] you can talk, and no one hears […] now, with the mobile phone, everyone can hear your conversation*". After that he claimed he had given up his mobile phone and reverted to the landline phone as his only means of phone communication. This came to us a surprise at first, since we intuitively felt that the possibility of moving around in space with a personal device, apparently independent from fixed physical locations, would represent an increase in privacy. This was also the concern expressed to us by Fajo previously, as well as something documented in previous research [8] where the ability to move around with the phone in the physical space as well as the ability to conceal oneself are seen as increase in the potential of secrecy, rather than a threat to it.

One can speculate that the usage of the landline, in terms of trust, has been one that has been negotiated for a longer time than that of the mobile phones. This may be one of the reasons that the landline presented itself as an intuitively better choice. Later, Jorege, one of our main informants, complemented this suspicion, by expressing similar concerns: "*everyone can hear when talking on the mobile*", Jorege complained. When we witnessed Jorege performing a phone call on the mobile, we began to understand what the problem was. As Jorege stood up to answer the phone, he walked back and forth along the seashore, repeating louder and louder "*Hello! Hello?*" He then hung up, failing to get a clear communication line, telling us that this was precisely the problem. The low strength of the network connectivity in certain parts of the island and at certain times had certain unintended consequences. This observation echoes studies in Western environments, showing how the mobile phone makes people behave differently in public places, sometimes in ways that provoke strong reactions (e.g. [20]).

First and foremost the network issues induced people into concentrating in areas of the island where the signal was stronger. Generally speaking this was by the water, facing Motalava, where the phone tower was installed. This resulted in large gatherings around those areas, of people making and receiving phone calls, as well as many just sitting around waiting for potential calls. This immediately made those areas populated with many different potential people overhearing conversations, possibly unintentionally. Secondly, and within that area and others, it got people moving back and forth while talking on their phones, seeking a better quality of communication, in those moments, presumably, the focus was not so much on where they were walking or who could be overhearing the conversation, but mainly on having a reliable and audible communication channel. Finally, it made people speak louder and louder on their mobiles and repeating the same sentences several times, in order to get their messages across. What was interesting, in any case was that, rather than simply allowing for a private, intimate sphere of communication, it resulted in a distancing from the physical situation, causing a loss of awareness of the surroundings, as well as one's own tone of voice and so on. Although it is also the case that the speaker is still attending to the local environment in certain ways, this can be seen as a form of "absent presence" [21],

This example from the field illustrates a situation in which the intersection of different spaces, such as the space where people are physically present and the space where there is network connectivity, intersecting to generate an awkwardness in regards to privacy, very different from the challenges presented by the landline, geographically fixed, phone. While we have discussed in previous work how this different layering of landscapes impacts the way bodies orient themselves in space [10], here we examine the implications for the intersection between private and public.

Drawing on social translucence, one can see how the *visibility* of others could be immediately present, yet, given this "absence presence", an extra effort is required to keep track of both the communication on the mobile and keeping a constant *awareness* of others in the surrounding. We will draw on this in the discussion later by extending the concept of *visibility*, not just to other people but the network itself, treating the network coverage as relevant social information for this particular interaction.

4.2 On the Persistent Nature of Digital Data

Given the prevalence and enthusiasm around mobile telephony, it was clear that there were many desirable aspects to them in comparison to the landline. One such aspect, highlighted by Jorege, addressing the previous concerns around lack secrecy and intermittent communication, was the SMS. In fact the SMS had two advantages from Jorege's perspective. One was that, given the often-poor quality of the network signal, the SMS represented a single packaged communication unit that would go through completely or not, allowing for a potentially less ambiguous form of communication. The other advantage was that it assured Jorege that "*no one listens to what I [Jorege] say*", and thus allowing for a more private mode of communication.

As we will now illustrate, that was not the whole story when it came to SMS communication. To illustrate this issue we relate an encounter from the field, of a young couple, Lisa and Robert who, for about eight months, had been in a secret, and otherwise forbidden relationship with each other. We do not intend to unpack all the cultural complexities of the forbidden nature of their relationship, nor is it of added value for the purpose of this paper. Suffice is to say that, prior to the arrival of the mobile phone, as Lisa told us, communication and arrangement of meetings between the young couple was made possible by using close friends as couriers. Despite their cautionary measures, one night, Lisa's sneaking out from her parents' house to visit Robert did not pass unnoticed, and her relationship was exposed to the whole community. At that time, and in order to avoid further problems they assured their parents and the community that they would stop seeing each other.

When mobile phones arrived to the island, at the time of our visit to Rah, the romance was still active. The couple was taking additional measures now, to ensure secrecy, and they saw the mobile phone as an opportunity to do so. Without the need to engage others in their secrets, Lisa and Robert were now using SMS to communicate: "*I text him every morning to wish him a good day*", Lisa told us. Although they were now able to communicate directly with each other, without the need of frequent physical proximity, and despite no one else being involved in their secret anymore, the usage of SMS turned out to have flaws of its own. Perhaps underestimating her father's technological skills, she left her mobile in the house one day, and while she was away her father went through her phone's inbox, and found out that Lisa and Robert were still involved in a relationship.

Perhaps as a consequence of this episode, as well as other concerns, and despite the relative novelty of mobile communications, the concern over phones as facilitating this type of forbidden relationships were echoed throughout our interviews in Rah: "*boys are meeting girls in secret*", said Fajo, "*[mobiles provide] easy access for boys to call girls*", echoed Fali. Other situations around married couples were also reported: "*mobiles have been causing problems inside the family […] already five mobile phones went to the toilet for causing conjugal disturbances*", concerns echoed by Fajo: "*it is recent [the arrival of the mobile] but already causing disturbances […] maybe someday will cause divorce […] married people are calling other married people and arranging meetings like that*". Communication embodies the primary way in which individuals engage in extra-marital, or infidelity affairs [22], with studies reporting how SMS plays a role in starting romantic relationships [3, 23].

These situations put the spotlight, on the gap that exists between expectations of secrecy, which could be expected from a mobile phone, particularly if used as a personal device, and the harsh reality of the fragility of that secrecy once it is violated and exposed. The persistent nature of the data in the mobile, whether SMS, call logs, or others, accumulates as a kind of *lifelog*, with a value and fragility, of which we are not always aware. A study in the Philippines reported cases of infidelity, where similar fragilities are reported [3].

These types of observations show how personal communication, and even the more secret and private type of personal communication, involves social elements. Receiving and storing an SMS in this context, is not just a social situation involving the person who was the intended recipient of that message, but by looking at eavesdropping, it becomes a social situation with the eavesdropper as well. We draw on social translucence to consider this less explored facet of sociality in mobiles. In terms of *visibility*, the act of eavesdropping may be conducted in a way which would make it impossible for the "victim" to see, such as in the case of Lisa's father. Unless she had caught him red-handed, in the eventuality that he wouldn't have found any sensitive material in her phone, his actions would have gone unnoticed. The SMS however is a perfectly *visible* and *shareable* [2] bit of social information. What happens is that the owner of the message can be easily deprived of any *awareness* of the eavesdropping, while the eavesdropper may easily remain *unaccountable* for his or her actions.

5 Designing for Awareness, Transience and Temporality

At this stage we would like to acknowledge the legitimate tensions that occur around secrecy regarding personal information on the mobile, in Rah, as well as in most other places. Issues of privacy, or in this case, a very specific situation of secrecy, eavesdropping, social/cultural norms and parental/community control, are of concern to most societies where the mobile made its way. In this work, we will stay away from the grand issues that are debated in much finer detail and greater complexity in anthropological, sociological and privacy studies, among others. Suffice it to say that we are not arguing for a disenfranchising of the parent's role in education, and some level of control of their children's education, as avoiding extreme misconduct, for instance, is essential to preserve the child's need for agency and autonomy, required for a healthy development, as documented in psychological as well as HCI literature [24, 25]. Our aim here is rather to elicit interesting domains to explore in design.

We reflect on these episodes and discuss three directions, which we find relevant for enriching the current state of interaction with mobile devices: spatial awareness, transience and temporality.

5.1 Spatial Awareness

When users communicate with mobile phones, they inhabit several places during the same clockwise time: the physical space and the conversational space, and are able to shift between them, almost as if occupying both spaces at the same time [26]. However, we saw how that flow was disrupted in Jorege's case. During his phone conversation, he tried to position his body in places with better network coverage. He also raised his voice, as this is an action that makes sense in the physical world whenever one has difficulties making oneself heard. The lack of coverage was a disrupting factor in the conversational space that he was orienting to as part of his mobile phone call. As he did this, however, he was also moving in a physical space, occupied by others. His body movement forced him to enter public areas where others might overhear part of his conversation. The inability to master the invisible coverage space and smoothly occupy all these spaces simultaneously could help explain why some wanted to drop their mobile phones and go back to using the landline, which is located in a very specific physical place.

With some notable exceptions [2], the interface of one mobile phone is the mobile phone is typically used and designed for personal use. However, the mobile phone is but one piece of the infrastructure that enables communication. When looking at the act of mobile phone mediated communication through social translucence, we can see social dynamics in the physical space that would benefit from the properties of visibility, awareness and accountability. Some authors have commented on the impact of mobile communication in institutions, their role in condensing or otherwise restructuring concepts of time and space, and how they have changed global markets or mediated experiences of remote places [27–29]. Our focus is rather on the ways that mobile communications shape individual actions. Mobile phones enable people to take part in two places at once, the physical space, where the body is located physically, is sometimes set aside in favor of the conversational space where communication occurs: "People talking on mobile phones seem wholly or partially unaware of their surroundings" [30], in a state of "absent presence" [21]. Due to this, there might be sometimes a 'friction between mobile users and co-present others' [31]. This friction has been widely noted in a number of cross-cultural studies [26, 32, 33], where authors explain how mobile phones temporarily disrupt public spaces for both users and bystanders. They also describe the different strategies that people use to create their own private spaces, through body orientations, refraining to talk about certain issues, or negotiation of contactability [33]. These studies highlight the complex relationships between physical space and the collective and social practices of the people who occupy it. Because of this, we argue for a contextual awareness, one that brings elements of the physical world into mobile phones (i.e. mixed reality [34]). Early work on interweaving of physical and virtual spaces has tried to design for awareness of the physical, using it as a resource in the interaction [35, 36].

More interestingly, mobile phone networks occupy a space that reaches beyond the personal device. Yet most of its components such as the network cell towers and the electromagnetic field they create remain *invisible*. The landline, on the other hand – at least the end terminals where communication takes place – has a known location, and is exposed, which makes the user and the possible observers *visible*. The network coverage space is essential to characterize the mobile phone conversational space because, if nothing else, it delimits it. Network coverage can be explicitly manipulated as in the case of deploying cell signal dampeners in public places, such as churches and libraries [29]. This is an example of how a social norm can be enforced through technology. The example we have observed in Rah, however, shows us a different side: the network coverage was limited due to the landscape of the archipelago and the reduced number of cell towers. Rather than enforcing a social norm, the limited coverage required people to navigate blindly, trying to find a place with better reception, especially as few of such well-known places were invariably already occupied. In Rah, the notion of "seamlessness" in communication breaks down in a very visible way. It is also apparent that the existing mobile phones were unable to provide help in defining boundaries and socially acceptable intersections of these different spaces. This indicates opportunities for designs that expose the seams of the technology – in this case the network coverage space – that mobile telephony rests on [37].

5.2 Designing for Transience

One of the aspects that struck us in the encounters described, was the persistent nature of the messaging communication, the call log, the SMS inbox and so on. These ultimately served as evidence against the youngsters. Of course this is a simplification of the complex social norms at stake here, not less significantly the myriad of perspectives on sexuality, infidelity and forbidden love of different sorts [22]. However, what inspired us was the mismatch between the amount of information stored in mobile devices, its rigid structuring and modes of visualization, with the everyday experiences and memories of people. To frame it within the concept of social translucence, in that particular social situation, the information of the phone may be openly exposed to whoever happens to get hold of that phone (*visible*). *Awareness* may be a challenge, in this case, given the often-long time ranges involved, and amount of information stored, making it hard for one to immediately grasp all the implications of disclosure. Finally, there is, in general no *accountability*, as visualization of that data by a third party is not registered in general (an exception would be in case someone is caught red-handed or would leave the phone in a different state than the user expected it to be, by leaving the messages open for instance). We will now discuss how these aspects can be harnessed, not to address directly the complex issues of infidelity and privacy, which require more than a technological intervention to be addressed, but rather how we can enrich our current HCI practice using these insights.

The overwhelming amount of digital information that we accumulate is not a new concern, with studies on email overload dating from nearly two decades ago [38]. Today, one could look back at years of email or SMS messaging history and get a very concrete outline and detailed information on different situations. At the same time, the "owner" of that communication history may only remember very fragmented bits of that history, and sometimes none at all. This property of digital information, which allows it to be stored in its intact, original form, is undoubtedly one of its most desirable properties. With memory space becoming cheaper and, in particular for SMS, the information occupying a relatively reduced amount, it may be tempting to store information on indefinitely [39]. What we can observe, on a fundamental level is that there are obvious differences between human memory and digital device memory [40]. This mismatch provides us with opportunities for rethinking how to design those systems with inspiration from both angles, particularly in regards to specifically designing, not only for persistence, but also for transience.

Though it has been argued that our digital possessions, like our physical possessions, help shape who we are, and how we present ourselves to others [41, 42], there are plenty of valid reasons for disposing of some those possessions. For instance, during romantic breakups [43]. Storage capacity, though as mentioned before seems less of a concern today, could still be an issue, with some modern Android phones discarding, by default, older messages, once the default cap of 200 is reached. This rather arbitrary way of deciding between persistence and transience is one which could be replaced, or complemented, with more meaningful ways that take into account other facets of those bits of communication to determine transience/persistence. We argue that there are opportunities to rethink this aspect, in relation to HCI in general, not only to address the specific issues around privacy, but rather to enrich our everyday interaction by, for instance helping us make immediate sense of our digital possessions, such as SMS messages [38, 44]. Given that, like physical possessions, not all digital possessions embody the same meaning or importance to their owners [45], there are two elements which we would like to explore, and see more explored in designing personal communication systems such as mobile phones. One is to look at how often those possessions (or SMS, in the particular case presented in this paper) are accessed, or looked at. This expression of the past activity in the present visualizations and representations is already existing in how systems such as Google, prioritize their search results. Another possible, albeit potentially more complex, approach would be to mine the data for meaning. By doing this in simple ways for instance, one could differentiate between long messages charged with emotional terms, for instance, from a shopping list or a mundane communication bit such as "call you right back!" Both of these are rather simple examples of domains we hope to see further explored in HCI.

5.3 Temporality

The previous discussion on transience invites us to think about *visibility, awareness* and *accountability*, in the context of privacy, by suggesting the direction of increased personalization and meaning of persistent data in personal communication systems. We would like now to extend that discussion to the broader concept of temporality. In particular, we consider the often-similar way in which information is presented in digital devices, where timestamps are usually used as the main structuring scaffold, whether for logs, display of messages and other interactions within digital technologies. These ways of structuring interaction are highly impersonal and universal. These universal modes of structuring are part of what results in such an open *visibility*, by making all the data comprehensible, while not drawing on the concepts of *awareness* and *accountability.*

Time and space are inseparable, synoptic concepts that constitute the fabrics of our perception [46, 47]. According to Merleau-Ponty, it makes no sense to talk about one without the other [46]. Time is generated by movement, or changes in space and inversely; space exists and changes with time. This phenomenological account of time is different than the clockwise account of time, which is most prevalent in current mobile systems. The latter is intended to be independent from our perception and activity, it is the time measured by clocks and calendars and the one we mostly use to communicate and understand other people. The universal synchronization of clock time has its roots in managing railroad traffic, perhaps one situation in which it would make immediate sense.

At times it seems as if we draw on these past constructions of time, and do not question them in the context of the designs that will be built into. As design guru John Chris Jones puts it: "The *time dimension*, if we may call it that, is left to *take care* of *itself*" [48]. In the following, we discuss a few ways that time can be used as a resource in the design process.

5.4 Time as a Resource in Design

The second example from our fieldwork shows how a series of communication exchanges between two lovers, happening in a private space over a long period of time, can be disrupted. One is able to gaze into the past and observe all the conversation since the moment of its inception, as if time had collapsed. Although being personal in this context, the mobile phone has no mechanisms for protecting the privacy of the space where the conversation existed. The way the mobile phone stores the conversation, linearly, makes it easily readable by anyone. Mobile phones, in their design, tend to adopt the linearity of clocks and calendars and represent content based on a concept of clockwise time, which unlike phenomenological embodied time is space and context independent. Harper et al. draw on a similar distinction by Bergson, on *temps* and *durée*. The former more connected with the linear, objective, clock time, and the latter, reflecting a more experiential aspect [49]. As seen in studies on the Facebook timeline, linear modes of representing time can also cause awkwardness on an experiential level, for instance when photos are posted a long time after the event they refer to [49].

It is precisely on these experiential aspects we would like to draw on for design. We can find inspiration in already known modes of experiential time, relevant to our everyday experiences, such as cyclic modes of experiencing time, day/night cycles, yearly cycles and circadian rhythms, to name a few. In HCI work has been done, for instance on how to design for reflection on personal experience [50, 51]. The Affective Health system [51], where biodata is represented in a spiral, in an analogy to cyclic modes of representing time and drawing on memory recall and reflection on past events, is an example of attempts to make what is now digital data into representations which are more meaningful to users. The scheduling system named Kairoscope, also builds on this experiential mode of time to build a scheduling system which only assigns specific times to nearby events and uses a pie-chart-like interface to represent one's scheduled events [52].

The focus of the works mentioned is not specifically aimed at addressing issues of privacy and secrecy. However, creating and exploring these personal representations around time, has been thought of as designing for empowerment [53], rendering representations which are directly meaningful, only to the "owner" of the data. Privacy comes as a side effect of thinking about designing more personalized experiences of time. In future work we will explore how to create representations around SMS messages, which are not represented around a linear representation of the time at which the messages were received, but rather on other aspects, such as "time spent" looking at each message, and how often they are looked at, the last time they were accessed and so on. It is these directions which we find interesting to explore in HCI, since they can create representations which connect better to the users' experience and memory, as has been discussed in HCI [54].

5.5 Beyond Time

Extending on the previous discussion, one could imagine dropping time in design altogether. It seems like, in many cases, time is more an inherited attribute, which receives no attention in organizing data and representations in several of these systems. If not properly justified, one could wonder "why include time altogether?" Rather, one can look at more meaningful ways to visualize and organize the same information. Perhaps the frequency of looking at a certain message, briefly discussed above, and altering the ways in which messages are visualized for instance, could be a simple example of how we can think about logs without necessarily designing around clock time, whilst providing potentially more value and meaning to the user.

Worth to notice is that also in this case, one would be implicitly adding elements of *accountability.* That is, if someone were to look on another person's messages and that activity would be reflected on posterior visualizations of those same messages (as is the case with organizing messages around the frequency of visualization), then the element of *awareness* would be present, since by the mere act of eavesdropping would be posteriorly visible to the owner of the mobile, in the case of the SMS inbox. The eavesdropper, if knowledgeable on the workings of the system, would also know that their activity would be perceived later on by the owner of the mobile (who made the owner *aware* of this activity), thus adding an element of *accountability*. That action then becomes formed by the decision of whether the potential gains in doing so are higher than the potential dissatisfaction from the other party. Again, we are by no means suggesting that this is a good solution, to address the situation exemplified by the case in Rah. One could in fact speculate that it probably would not have changed the course of events, since the community would have confronted the youngsters with the incident in question once possessing that knowledge. However, one may wonder whether the person looking through the phone's messages would have avoided it so with fear of not finding anything relevant and be subjected to the potential anger of the other party whose privacy was just violated.

6 Towards an Understanding of Digital Patina

In this paper, we have dealt with issues of spatial awareness, transience and temporality. We have seen how these can be used as a way of thinking about mobile interaction design in a generative way. This discussion is not a silver bullet in addressing these issues; in fact they are not intended to be solutions at all, since we do not define those encounters as "problems" to be solved. Rather, they illustrate interesting opportunities for research in HCI and thinking about design for mobile devices. As a final remark we would like to discuss some of the potential contributions that drawing on temporality and transience can contribute to HCI and digital interaction design, specifically, the notion of *digital patina*.

Patina is a term referring to decay of physical artifacts, which has been known to add charm, and sometimes increased attachment or cherish for those artifacts [44, 55]. In interaction design we have seen inspiration from this, particularly in the form of bringing aesthetic, often visual, elements from real world materials, often thought of as developing patina, such as metals or leather, into the designing of visual elements for mobiles or computers at large. Some have referred to this type of inspiration as digital patina [44], since it is about bringing the elements of real world patina into the design of digital artifacts. However, we argue that there could be something much more specific to digital patina, than this skeuomorphic take on digital patina. Perhaps the temptation of adopting this approach is to draw on analogies, or metaphors, from the real world, which may remind us of different artifacts, and thus potentially evoke certain emotions [44], such as a leathery casing of an old agenda. However, as new generations, or other cultural contexts have not been exposed to those artifacts, these appropriated forms of patina may not maintain the same relevance.

Explorations around temporality and transience, for instance, and particularly forms of visualizing those dimensions in digital systems, may offer some insights into this domain. If we rethink simple aspects such as how logs and lists of messages are organized, how messages are kept and deleted, we may start getting particular forms of treating digital information which may "age" with the user and their interaction. Playing with the ageing of these structures, as opposed to perfectly conserved linear representations around universal, clock-based, timestamps, might be one way to being seriously considering what patina means for digital information. The unlimited ability to perfectly replicate data, a characteristic pertaining to digital data, makes the exploration of patina, and more generally of digital data, a most exciting direction to pursue in HCI research.

References

1. Arminen, I., Weilenmann, A.: Mobile presence and intimacy—Reshaping social actions in mobile contextual configuration. Journal of Pragmatics 41, 1905–1923 (2009)
2. Weilenmann, A., Larsson, C.: Local Use and Sharing of Mobile Phones. In: Brown, B., Green, N., Harper, R. (eds.) Wireless World: Social and Interactional Aspects of the Mobile Age, pp. 99–115. Springer (2001)
3. Ellwood-Clayton, B.: All we need is love—and a mobile phone: Texting in the Philippines. In: The International Conference on Cultural Space and the Public Sphere in Asia, Seoul, Korea (2006)
4. Ito, M., Okabe, D.: Intimate connections: Contextualizing Japanese youth and mobile messaging. In: Harper, R., Palen, L., Taylor, A. (eds.) Inside the Text Social Perspectives on SMS in the Mobile Age, pp. 127–145. Springer (2005)

5. Horst, H.A.: The blessings and burdens of communication: cell phones in Jamaican transnational social fields. Global Networks 6, 143–159 (2006)
6. Hijazi-Omari, H., Ribak, R.: Playing with fire: On the domestication of the mobile phone among Palestinian teenage girls in Israel. Information Communication Society 11, 149–166 (2008)
7. Ito, M.: Mobile Phones, Japanese Youth, and the Re-Placement of Social Contact. In: Ling, R., Pedersen, P. (eds.) Mobile Communications Renegotiation of the Social Sphere, pp. 1–15. Springer (2001)
8. Ling, R.S.: The mobile connection: the cell phone's impact on society. Morgan Kaufmann (2004)
9. Erickson, T., Kellogg, W.A.: Social translucence: an approach to designing systems that support social processes. ACM Trans. Comput.-Hum. Interact. 7, 59–83 (2000)
10. Ferreira, P., Höök, K.: Bodily orientations around mobiles. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, p. 277. ACM Press, New York (2011)
11. Ferreira, P., Höök, K.: Appreciating plei-plei around mobiles. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI 2012, p. 2015. ACM Press, New York (2012)
12. 2009 National Census of Population and Housing, Port Vila, Vanuatu (2009)
13. Crabtree, A., Rodden, T., Tolmie, P., Button, G.: Ethnography considered harmful. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009. Segunda Ed (2009)
14. Dourish, P.: Implications for design. Presented at the (2006)
15. Geertz, C.: The Interpretation of Cultures: Selected Essays. Basic Books (1973)
16. Taylor, A.S., Harper, R.: The Gift of Gab? A Design Oriented Sociology of Young People's Use of Mobiles. Computer Supported Cooperative Work 12, 267–296 (2003)
17. Licoppe, C., Heurtin, J.-P.: The cellular phone as a tool for managing risk and trust in social bonds: re-locating mobile phone users and re-synchronizing unexpected contexts in cellular phone interactions. In: Katz, J., Aakhus, M. (eds.) Perpetual Contact: Mobile Communication, Private Talk, Public Performance, pp. 94–109. Cambridge University Press, Cambridge (2002)
18. Licoppe, C.: "Connected" presence: the emergence of a new repertoire for managing social relationships in a changing communication technoscape. Environment and Planning D: Society and Space 22, 135–156 (2004)
19. Green, N.: Who's watching whom? Monitoring and accountability in mobile relations. In: Brown, B., Green, N., Harper, R. (eds.) WirelessWorld: Social and Interactional Aspects of the Mobile Age. Springer (2001)
20. Monk, A., Carroll, J., Parker, S., Blythe, M.: Why are mobile phones annoying? Behaviour & Information Technology 23, 33–41 (2004)
21. Gergen, K.: The challenge of the absent presence. In: Katz, J.E., Aakhus, M.A. (eds.) Perpetual Contact Mobile Communication Private Talk Public Performance, pp. 227–241. Cambridge University Press (2002)
22. Mattingly, B.A., Clark, E.M., Weidler, D.J., Bullock, M., Hackathorn, J., Blankmeyer, K.: Sociosexual Orientation, Commitment, and Infidelity: A Mediation Analysis. The Journal of Social Psychology 151, 222–226 (2011)
23. Byrne, R., Findlay, B.: Preference for SMS versus Telephone Calls in Initiating Romantic Relationships. Society 2, 48–61 (2004)
24. Maccoby, E.: The role of parents in the socialization of children: An historical overview. Developmental Psychology 28, 1006–1017 (1992)

25. Yardi, S., Bruckman, A.: Social and technical challenges in parenting teens' social media use. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, p. 3237. ACM Press, New York (2011)
26. Höflich, J.R., Hartmann, M.: Mobile communication in everyday life: ethnographic views, observations and reflections. Presented at the (2006)
27. Castells, M.: The Rise of the Network Society, vol. I. Wiley-Blackwell (2010)
28. Tsatsou, P.: Reconceptualising "Time" and "Space" in the Era of Electronic Media and Communications. Journal of Media and Communication 1, 11–32 (2009)
29. Dourish, P., Bell, G.: The infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space. Environment and Planning B Planning and Design 34, 414–430 (2007)
30. Persson, A.: Intimacy Among Strangers: On mobile telephone calls in public places. Journal of Mundane Behavior (2001)
31. Cooper, G.: The mutable mobile: social theory in the wireless world. Springer Computer Supported Cooperated Work Series, pp. 19–31 (2001)
32. Ling, R.: One can talk about common manners!': the use of mobile telephones in in appropriate situations. Telektronikk 94, 65–78 (1998)
33. Haddon, L.: The Social Consequences of Mobile Telephony: Framing Questions. Sosiale Konsekvenser av Mobiltelefoni. Telenor, Oslo (2000)
34. Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information Systems E77-D (1994)
35. Benford, S., Greenhalgh, C., Reynard, G., Brown, C., Koleva, B.: Understanding and constructing shared spaces with mixed-reality boundaries. ACM Transactions on Computer-Human Interaction 5, 185–223 (1998)
36. Dix, A., Rodden, T., Davies, N., Friday, A., Palfreyman, K.: Exploiting Space and Location as a Design Framework for Interactive Mobile Systems. ACM Transactions on Computer-Human Interaction 7, 285–321 (2000)
37. Chalmers, M., MacColl, I.: Seamful and seamless design in ubiquitous computing. Presented at the (2003)
38. Whittaker, S., Sidner, C.: Email overload: exploring personal information management of email. In: Kiesler, S. (ed.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Common Ground, pp. 276–283. ACM (1996)
39. Bell, G., Gemmell, J.: Total Recall: How the E-Memory Revolution Will Change Everything. Dutton Adult (2009)
40. Van House, N., Churchill, E.F.: Technologies of memory: Key issues and critical perspectives. Memory Studies 1, 295–310 (2008)
41. Kirk, D., Sellen, A.: On Human Remains: Excavating the Home Archive. Microsoft Research Cambridge UK. MRS–TR–2008–8 (2008)
42. Odom, W., Zimmerman, J., Forlizzi, J.: Teenagers and their virtual possessions. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, CHI 2011, p. 1491. ACM Press, New York (2011)
43. Sas, C., Whittaker, S.: Design for Forgetting: Disposing of Digital Possessions Afer a Breakup. In: Proceedings of the Forthcoming SIGCHI Conference on Human Factors in Computing Systems. ACM, New York (2013)
44. Odom, W., Banks, R., Kirk, D., Harper, R., Lindley, S., Sellen, A.: Technology heirlooms?: considerations for passing down and inheriting digital materials. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI 2012, pp. 337–346. ACM (2012)

45. Odom, W., Pierce, J., Stolterman, E., Blevis, E.: Understanding why we preserve some things and discard others in the context of interaction design. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, p. 1053. ACM Press, New York (2009)
46. Lanigan, R.: Speaking and semiology: Maurice Merleau-Ponty's phenomenological theory of existential communication. Walter de Gruyter (1991)
47. Lash, S.: Another modernity, a different rationality. Wiley-Blackwell (1999)
48. Jones, J.C.: Design Methods. Wiley (1992)
49. Harper, R., Whitworth, E., Page, R.: Identity, Time and Durée on Facebook. Selected Papers of Internet Research (2012)
50. Lindley, S.E., Glancy, M., Harper, R., Randall, D., Smyth, N.: Oh and how things just don't change, the more things stay the same": Reflections on SenseCam images 18 months after capture. International Journal of Human-Computer Studies 69, 311–323 (2011)
51. Sanches, P., Höök, K., Vaara, E., Weymann, C., Bylund, M., Ferreira, P., Peira, N., Sjölinder, M.: Mind the body!: designing a mobile stress management application encouraging personal reflection. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 47–56 (2010)
52. Martin, R., Holtzman, H.: Kairoscope: managing time perception and scheduling through social event coordination. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1969–1978. ACM, New York (2011)
53. Höök, K., Ståhl, A.: Interactional empowerment. In: Proceedings of the Twenty- ..., pp. 647–656. ACM, New York (2008)
54. Harper, R., Randall, D., Smyth, N., Evans, C., Heledd, L., Moore, R.: The Past is a Different Place: They Do Things Differently There. Sociology The Journal Of The British Sociological Association, 271–280 (2008)
55. DeSilvey, C.: Observed Decay: Telling Stories with Mutable Things. Journal of Material Culture 11, 318–338 (2006)

Comparison of Phone-Based Distal Pointing Techniques for Point-Select Tasks

Mohit Jain[1,3,*], Andy Cockburn[2], and Sriganesh Madhvanath[3]

[1] IBM Research, Bangalore, India
mohitjain@in.ibm.com
[2] University of Canterbury, Christchurch, New-Zealand
[3] HP Labs, Bangalore, India

Abstract. Many different mobile phone-based distal pointing techniques have been proposed and evaluated. Interaction with distant large-screen displays, including interactive TV, requires active pointing and selection of target items. In this paper, we empirically compare four different phone-based distal pointing techniques for point-select tasks. Results show that participants prefer a discrete pointing technique using the phone's touchscreen as a touchpad. This method also achieved the highest accuracy among the techniques studied, with comparable speed. We discuss the implications of our findings for distal interaction.

Keywords: Interactive TV, iTV, Phone, Distal Pointing, Evaluation.

1 Introduction

Interactive TV (iTV) shifts the nature of interaction with distant displays from being predominantly passive consumption of traditional TV to more active control of media and content. Accompanying this shift is an increase in the range of interactive menus displayed on screen and the need for rich input to perform complex tasks including point-and-select. Conventional remote controls consisting of an array of buttons are not ideal for these tasks [11]. This has led to the development of novel distal interaction techniques using phones [4], [9], game controllers [6], hand gestures [17], and novel hardware [5], [10], as input devices. In particular, phones offer several advantages: they are computationally powerful [3], have various built-in sensors, and support wireless communication.

Pointing at targets is a fundamental component of interacting with distant displays. Previous research has successfully used phone's built-in sensors such as *touchscreen, camera, accelerometer, gyroscope,* etc. to sense user's pointing input. *Touch* input, sensed using the touchscreen have been found to work well for distal interaction [7], [12], [13] due to the users' familiarity with laptop touchpads. Applications like Air Mouse [1] allow in-air movements of the phone for interaction. It uses a combination of *accelerometer and gyroscope* data to enable direct pointing, similar to a laser-pointer. The accelerometer measures the acceleration applied to the device, whereas the gyroscope provides the angular orientation of the device in space. The data

[*] Part of this work was done while the first author was an intern at HP Labs India.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 714–721, 2013.

obtained from these sensors is accordingly mapped to a pointer position on the display. (Techniques tracking phone's camera and flashlight [16] have also been proposed to sense motion inputs, however due to low accuracy and extra hardware requirements, we are not considering them.) Touch and motion are the two most commonly used input techniques, and an understanding of their impact on distal interaction is crucial.

On the distal display, there are two ways to provide visual feedback while selecting an item, irrespective of the input action – (a) *Continuous cursor*, in which a cursor moves continuously over the display (*e.g.*, Gmote [9]), and (b) *Discrete selection-block*, in which a selection-block moves discretely between selectable items (*e.g.*, Apple TV [2]). Note that continuous cursor is target agnostic, while discrete selection-block is target dependent. Continuous and discrete methods converge if targets are a single pixel in size with zero pixels of separation between them. To date, both types of feedback have been widely used, but there has been little work on how on-screen feedback affects distal interaction performance for point-select tasks.

In this paper, we study the relative merits of *continuous* versus *discrete* feedback for pointing interaction, when input is provided through *touch* sensed using the phone's touchscreen and through *movement in the air* sensed using the combination of the phone's gyroscope and accelerometer. We study four distal pointing techniques that satisfy our iTV design goals of single-handed usage with minimal physical and mental load: a) Continuous Touchpad, b) Continuous inAir, c) Discrete Touchpad, and d) Discrete inAir. We conducted a study varying the size and separation of target items to find the relationship between techniques and point-select task difficulty. As these phone-based distal interaction techniques have all been described in prior literature, the contribution of this paper is the empirical comparison of these techniques for point-select tasks. Our study showed that the participants prefer Discrete Touchpad over the other techniques. It also achieved the highest accuracy with comparable speed. We conclude with a discussion of the design implications for distal interaction.

2 Design Goals and Techniques

For phone-based interaction with iTV, we have identified a set of design goals important for end-user adoption: (a) *Single-handed Interaction*: remote controls are usually used with one hand [5]; (b) *Minimal Visual Attention Switching*: the controller should not require direct attention, in order to allow users to focus on the distal display, because attention switching has been found to result in higher error rates [15]; (c) *Minimal Learning Curve*: use of a controller should not require expertise (many pointing techniques described in the literature [4], [5], [13] were found to have a strong learning curve); (d) *Minimal Physical Load*: the controller should use small wrist or thumb movements for interaction to minimize physical load, because techniques requiring arm movement increase fatigue due to the *gorilla-arm-effect* [5].

We studied the following four distal interaction techniques that satisfied our design goals. A *single-tap* on the phone's touchscreen is used for selection.

Continuous Touchpad (CT): In CT, the *touchscreen* of the phone acts as a laptop's touchpad (similar to Gmote [9]), allowing the user to move a cursor over the display

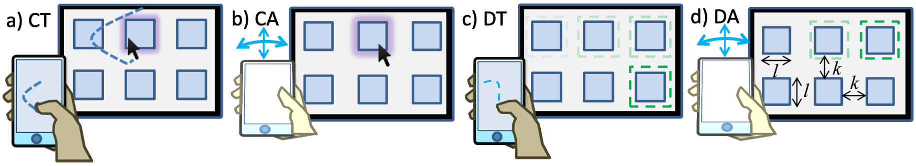


Fig. 1. a-d shows the four interaction techniques. d also shows the definition of item size ('l') and gap size ('k').

(Fig. 1a). For the mapping between display and touchpad, we used standard Windows XP's cursor acceleration curve, such that a single rapid swipe across the touchscreen allows the cursor to traverse the entire width of the distal display.

Continuous inAir (CA): CA uses the built-in motion sensors of the phone to determine angular orientation and acceleration of the phone and thus control cursor location (similar to Air Mouse [1]). The user moves the phone in air with small wrist-motions to move the cursor (Fig. 1b) to the required target item, followed by a single tap on the phone's touchscreen to trigger selection. The mapping is such that a 30-degree device rotation along the horizontal x-axis, allows the cursor to traverse the width of the display. This mapping value, also known as Control Display gain (CD gain), was chosen using a pilot study with 5 participants. We tested four values of CD gain (20, 25, 30 and 35 degrees) and found maximum accuracy with 30-degrees.

Discrete Touchpad (DT): DT is the same as CT, except that a *discrete selection-block* replaces the cursor as the feedback mechanism, and hence it supports clutching (*i.e.,* unlike a cursor, the selection-block method can never be on a non-selectable or dead space) (Fig. 1c). For example, to move to the selectable item to the immediate right, the user is required to make a small swipe on the touchscreen towards the right. The user can perform a longer swipe to move rapidly through the selectable items.

Discrete inAir (DA): DA is the same as CA, except that a *discrete selection-block* replaces the cursor (Fig. 1d). The user moves the phone in the air with small wrist-motions to move the selection-block. DA has the same CD gain as that of CA.

3 Experiment

3.1 Participants and Apparatus

Twelve male employees of an IT organization participated in the study (mean age 25.6 years, sd 3.4). All but one were right-handed; all reported using computers for 8-10 hours a day; and all but one had previously used a touchscreen phone on a regular basis for 6 months or longer. None had previously interacted with a distant display using a phone. Participants were given a ~$10 voucher as reward.

An iPod Touch 4th generation (display resolution: 960×640) was used for the four interaction techniques. A 42" LCD display (1360×765) connected to a computer as an external monitor was used to simulate the iTV. The display was positioned at eye level and participants sat comfortably on a sofa (with hand-rest) placed 10 feet from the display. The phone and computer interacted over a local wireless network.

3.2 Task, Factors, and Design

Most previous research [4], [11], [14] has used the traditional Fitts' Law point-select task [8] for evaluation. However, the standard Fitts' Law experimental manipulation of controlling Index of Difficulty (the log of the ratio of amplitude to target width) is likely to misrepresent the difficulty of discrete targeting. Thus we performed a variation of the Fitts' Law task. The display was divided into a grid of square-shaped regions of side length ($l + k$) each of which had a square target of side l at its center. k is the gap between two consecutive items (Fig. 1d). The l and k values were varied, which accordingly varied the sizes of the items and the distances between them, respectively. The different values of l and k were selected by studying state-of-the-art iTV interfaces such as Apple TV, and Roku. Nothing was displayed on the phone's screen.

A within-subject design was used for the study. The independent variables were: *Feedback* (2: Continuous, Discrete), *Input* (2: Touchpad, inAir), *Item size l* (3: 100, 60 and 40 pixels), and *Gap between items k* (3: 80, 40 and 20 pixels). The 9 layouts formed by combining the different item size and gap values had different numbers of targets: 28 (# of rows: 4, # of columns: 7; item size: 100, gap: 80), 45 (5, 9; 100, 40), 66 (6, 11; 100, 20), 45 (5, 9; 60, 80), 91 (7, 13; 60, 40), 153 (9, 17; 60, 20), 66 (6, 11; 40, 80), 153 (9, 17; 40, 40), and 264 (12, 22; 40, 20). No item touched the display border, permitting overshooting for continuous techniques, even for the border items.

3.3 Procedure

All the tasks were performed while sitting, holding the phone in the dominant hand (in portrait mode) (Fig. 1). The facilitator demonstrated each of the techniques. Following a practice set of 30 trials with varying target density, participants performed a set of 30 trials for each technique for each of the nine layouts, in a random order. The ordering of the techniques used a Latin Square to mitigate learning and fatigue effects.

In each trial, one of the on-screen items was randomly selected as the target and shown in green. The participant was required to select this item, as accurately and as quickly as possible. When the cursor or selection-block hovered over an item, the item color changed to yellow, providing visual feedback. A correct response is received when the participant triggered selection using a single-tap on phone's touchscreen, while the cursor or selection-block was hovering over the correct target item. Selection of the wrong item was shown by turning it red. Participants had to keep trying until a correct response was received, in order to measure error rate. After each technique, participants were asked to rate the technique in terms of perceived speed, perceived accuracy, and measures from NASA TLX, on a 5-point Likert-scale (for instance, for *fast* 1was very slow and 5 was very fast).

4 Results

The total number of trials = 4 techniques × 12 participants × 3 item sizes × 3 gap sizes × 30 trials = 12960. Participants took ~70 minutes to complete the study, including breaks of 2-5 minutes between techniques.

a)

Tech.	**Accuracy**	**Sel. Time**	**Errors**
CT	93% (7%)	2047 (332)	49% (6%)
CA	81% (7%)	1491 (176)	140% (70%)
DT	96% (2%)	1894 (196)	4% (2%)
DA	86% (5%)	1516 (200)	16% (7%)

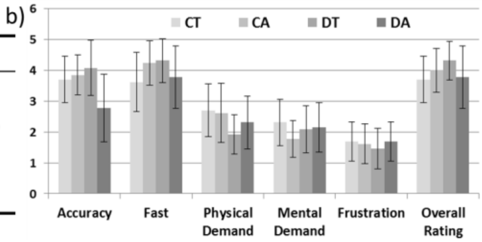


Fig. 2. a) Performance measure means with standard deviation in brackets. b) Likert-scale rating for the techniques (with error bars showing standard deviation)

Our analysis uses a 2×2×3×3 RM-ANOVA for factors *Feedback* (2) × *Input* (2) × *Item size* (3) × *Gap size* (3), and a 4×3 RM-ANOVA for Technique (4) × Distance (small, medium, large). Here distance is the number of items between the starting item and the target item, which can vary from 1 to 20. For analysis, we divided this into three distance levels – small (1-5 items), medium (6-13) and large (14-20).

4.1 Accuracy

Accuracy is defined as the proportion of times the target item was selected correctly on the first attempt. Participants achieved the highest accuracy of 96% with DT (Fig. 2a). The 4×3 ANOVA showed CA was least accurate, with DT more accurate than DA and CA: $F_{3,33}=23.9$, $p<0.001$. The 2×2×3×3 analysis showed a significant effect for input ($F_{1,11}=48.5$, $p<0.001$), with touchpad (m=94%, sd=5%) more accurate than inAir techniques (m=83%, sd=7%). For feedback, discrete (m=91%, sd=7%) outperformed continuous (m=87%, sd=9%) techniques, with $F_{1,11}=8.5$, $p<0.05$.

As expected, there was a significant main effect for item size ($F_{2,22}=9.2$, $p<0.01$), with larger items being more accurately selected (Fig. 3a). Importantly, there was a significant *feedback* × *item size* ($F_{2,10}=4.9$, $p<0.05$), and *feedback* × *gap size* ($F_{2,10}=21.4$, $p<0.001$) interactions. For *continuous* methods, the accuracy decreased with decreasing item size (Fig. 3a), and increased with decreasing gap size (Fig. 3c). This shows that continuous methods are better suited for large items with small gaps between them. The 4×3 analysis found a significant effect for distance ($F_{2,22}=13.7$, $p<0.001$), and a strong interaction between technique and distance ($F_{6,66}=7.3$, $p<0.01$). Post-hoc analysis revealed that selecting targets at small distances was significantly more accurate than medium and large distances, with $p<0.001$ (Fig. 3e). Higher inaccuracy with larger distances may be because of overshooting effects, as found in [4]. This effect was more noticeable for the inAir techniques (Fig. 3e), where the participants reported difficulty in controlling the cursor over longer distances. On analyzing overshooting, we found 67.3% (sd=32.6) of the erroneous selections for DT and DA were due to erroneously selecting items situated a hop away from the target item.

For each selection task, multiple attempts were allowed. Note that this allows error rates to be higher than 100%, as in the case of CA (m=140%). Participants took the fewest attempts with DT, with an error rate of 4% (Fig. 2a). Error rate analysis showed an effect similar to that of accuracy, the details are omitted for brevity.

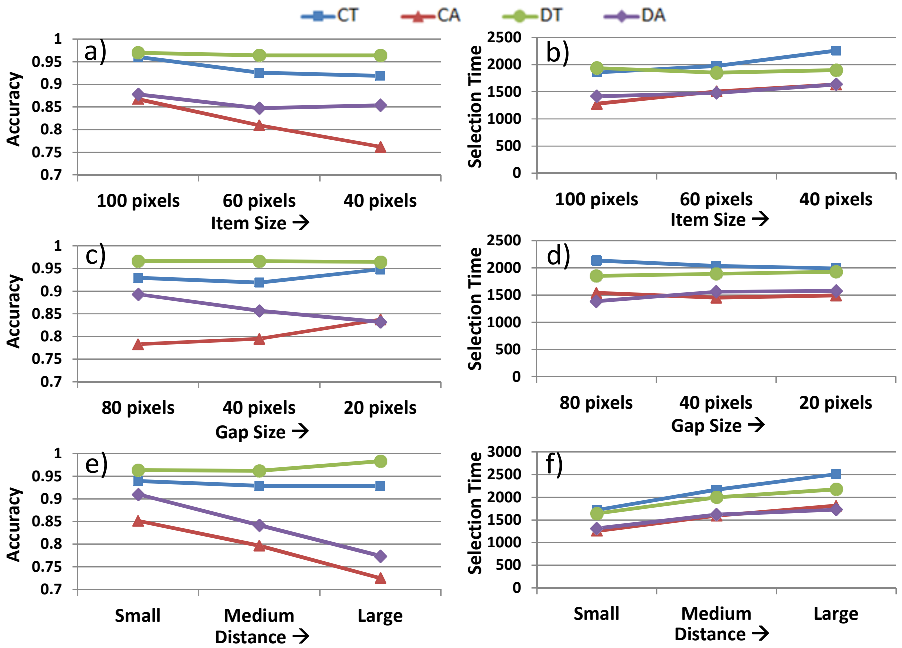


Fig. 3. a) Accuracy and b) Selection Time respectively for different item sizes (a,b), gap sizes (c,d) and distances (e,f)

4.2 Selection Time

Selection Time is the error-free target selection time (trials containing errors were removed). Although CA was least accurate, it was the fastest with an average selection time of 1491 ms showing the speed-accuracy tradeoff. The 4×3 ANOVA showed CA, DA, and DT to be significantly faster than CT ($F_{3,33}$=28.1, p<0.001) (Fig. 2a). The 2×2×3×3 ANOVA showed significant main effects for *input* ($F_{1,11}$=36.3, p<0.001), with *inAir* techniques (m=1504, sd=185) being significantly faster than *touchpad* (m=1971, sd=273). No significant difference was observed between the two feedback methods for selection time.

There was a significant main effect for *item size* ($F_{2,22}$=31.7, p<0.001) with large items being the fastest to select (Fig. 3b). Similarly, large gaps resulted in smaller selection time (Fig. 3d). There were significant *feedback* × *item size* ($F_{2,22}$=16.9, p<0.01) and *feedback* × *gap size* ($F_{2,22}$=17.1, p<0.001) interactions (Fig. 3b, Fig. 3d). For *continuous* methods, the selection time increased with decreasing item size (Fig. 3b), while it remained unaffected by gap size (Fig. 3d). This shows that continuous methods are better suited for large items. For *discrete* methods the selection time remained unaffected by item size, and increased with decreasing gap size; this may be because of the increased density of items. The 4×3 ANOVA found a significant effect for distance ($F_{2,22}$=178.4, p<0.001), and a strong *technique* × *distance* interaction ($F_{6,66}$=6.7, p<0.01). All techniques were significantly faster at selecting targets at small distances, compared to medium and large distances.

4.3 Subjective Assessment and Feedback

Participants' responses to the 5-point Likert scale questions showed that DT received the highest overall rating. DT was also assessed as the fastest, most accurate, involving the least physical effort, and least frustrating (Fig. 2b). All the participants stated that as none of the techniques required looking at the phone screen, it helped them to be more effective. CT was rated the worst in terms of physical and mental load, and for inAir techniques, participants noted some learning curve. Five participants had difficulty with the continuous cursor techniques, citing problems with initially locating the cursor (*e.g.*, *"hard to find the cursor when it has gone to the border of the screen"*) and with making precise selections (*e.g.*, *"selecting smaller targets is tough"*). This may be due to the overshooting effect, resulting in erroneous selections performed while the cursor was in the dead space between target items. On the other hand, participants praised the effectiveness of discrete methods for selecting smaller targets.

Participants liked that the touchpad was *"highly sensitive"*. Three participants complained that moving diagonally towards the upper-left of the touchpad was difficult using the thumb. For inAir techniques, four participants praised the speed offered, while two participants asked to reduce the sensitivity, as it was causing errors while selecting nearby targets. Using CA with small target size, participants complained that *"tapping for selection resulted in slight phone movement, leading to erroneous selection"*. This error was exacerbated due to cursor jitter resulting from hand unsteadiness, as reported in [14]. Hence, for using CA, the application design should have targets of medium to large item size and considerable gap between consecutive items. However all participants *"enjoyed"* using the inAir techniques, which may be due to the novelty factor.

5 Discussion and Design Implications

Overall, the study revealed that touchpad-based techniques were significantly more accurate for point-select tasks. The inAir techniques were perceived as enjoyable, as also reported in [15]. However, participants preferred using touch over motion input, potentially due to accuracy. In terms of feedback, the discrete methods were more accurate than their continuous counterparts, with comparable (or faster) selection times. This is interesting because most current distal pointing systems [9], [10] use continuous feedback. Our results suggest that discrete methods may be preferable unless there is a strong need for a continuous modality, *e.g.,* to support drawing and annotation on screen. Our results also lead us to specific design recommendations for such interactions:

Layout Considerations: For touchpad-based techniques, moving diagonally towards the upper-left of the touchpad was reported to be difficult; hence less-used items should be placed at the upper-left corner region of the display. For continuous feedback, we found overshooting effects, similar to [4]. As a design decision, continuous techniques should always be implemented with gaps between consecutive targets, and performing the selection action while in the dead space should not be an error.

Finger-up Selection: The high error rates with inAir techniques resulted from difficulty in controlling the cursor, specifically from phone movement while tapping for selection. A design solution could be to initiate the inAir movements when a finger touches the phone's screen, with finger-up used for triggering selection, as that might minimize phone movement for selecting items. Finger-up for selection may be counter-intuitive to the current mental mapping and hence require further research.

Walk-up-and-use System: Interactive systems installed at public locations are walk-up-and-use system, and hence require maximal accuracy with minimal learning curve. For such systems, CT seems like a wise choice, as CA has the highest error rate, while DT and DA requires learning to minimize the overshooting errors.

Given the generic nature of the point-select tasks, our findings and design recommendations are applicable beyond iTV to distal interactions in general. The present study results are based on a limited number of male right-handed participants, with a brief continuous interaction in a lab setting. In the future, we propose a longitudinal study with users from different genders and handedness, and diverse background to explore and evaluate real-world performance of different distal interaction techniques.

References

1. Air Mouse by Movea, `https://itunes.apple.com/us/app/air-mouse/id402762122?mt=8`
2. Apple Remote app, `http://itunes.apple.com/us/app/remote/id284417350?mt=8`
3. Ballagas, R., et al.: The Smart Phone: A Ubiquitous Input Device. In: IEEE Pervasive Computing, vol. 5, pp. 70–77. IEEE (2006)
4. Boring, S., Jurmu, M., Butz, A.: Scroll, Tilt or Move It: Using Mobile Phones to Continuous Control Pointers on Large Public Displays. In: OzCHI, pp. 161–168. ACM (2009)
5. Choi, S., et al.: RemoteTouch: Touch-Screen-like Interaction in the TV Viewing Environment. In: CHI, pp. 393–402. ACM (2011)
6. Cox, D., Wolford, J., Jensen, C., Beardsley, D.: An Evaluation of Game Controllers and Tablets Controllers for Interactive TV Applications. In: ICMI, pp. 181–188. ACM (2012)
7. Enns, N.R.N., MacKenzie, I.S.: Touchpad-Based Remote Control Devices. In: CHI, pp. 229–230. ACM (1998)
8. Fitts, P.M.: The information capacity of the human motor system in controlling amplitude of movement. Journal of Experimental Psychology 47, 381–391 (1954)
9. Gmote, `http://www.gmote.org/`
10. Loop Pointer, `http://www.hillcrestlabs.com/loop/`
11. MacKenzie, I.S., Jusoh, S.: An Evaluation of Two Input Devices for Remote Pointing. In: Nigay, L., Little, M.R. (eds.) EHCI 2001. LNCS, vol. 2254, pp. 235–249. Springer, Heidelberg (2001)
12. MacKenzie, I.S., Oniszczak, A.: A Comparison of Three Selection Techniques for Touchpads. In: CHI, pp. 336–343. ACM (1998)
13. McCallum, D.C., Irani, P.: ARC-Pad: Absolute+Relative Cursor Positioning for Large Displays with a Mobile Touchscreen. In: UIST, pp. 153–156. ACM (2009)
14. Myers, B.A., et al.: Interacting at a Distance: Measuring the Performance of Laser Pointers and Other Devices. In: CHI, pp. 33–40. ACM (2002)
15. Rashid, U., Kauko, J., Haakila, J., Quigley, A.: Proximal and Distal Selection of Widgets: Designing Distributed UI for Mobile Interaction with Large Display. In: MobileHCI, pp. 495–498. ACM (2011)
16. Shirazi, A.S., Winkler, C., Schmidt, A.: Flashlight Interaction: A study on mobile phone interaction techniques with large displays. In: MobileHCI, pp. 1–2. ACM (2009)
17. Vatavu, R.D.: User-Defined Gestures for Free-Hand TV Control. In: EuroITV, pp. 45–48. ACM (2012)

Extending Mobile Interfaces with External Screens

Julian Seifert[1], Dennis Schneider[2], and Enrico Rukzio[1]

[1] Ulm University, Ulm, Germany
{julian.seifert,enrico.rukzio}@uni-ulm.de
[2] University of Duisburg-Essen, Essen, Germany
dennis.schneider_1@uni-due.de

Abstract. Mobile phones allow for the use of all kinds of applications, and their mobile applications often provide similar functionalities as desktop applications. However, they are constrained by the limited screen size of the mobile device. Accordingly, designs of mobile user interfaces require optimization for small screens. As a consequence, users are provided with less context and often have to switch views or resize content such as maps or pictures. We present MobIES[1], a novel approach for *extending mobile user interfaces by using external screens* (e.g., the mobile phone and a large screen). Users can utilize more space and can thus overview a larger information context. We present a novel interaction and application concept and describe how user interfaces can be spanned across displays. Further, we contribute an original approach for using Near Field Communication to detect the devices' spatial relation. We report on a user study which compared MobIES with standard mobile settings. Results from the system usability scale show that interaction with MobIES is subjectively more usable. Furthermore, it provides higher perceived information clarity and supports faster sharing of information to others.

Keywords: Mobile phones, distributed user interfaces, interaction.

1 Introduction

Today's mobile phones enable users to perform a large variety of tasks in mobile contexts. Given the increased computing power, battery capacity, and data connectivity, users can perform the same tasks as by using traditional personal computers (e.g., browsing the web, viewing and editing photos). One of the limiting factors is the screen size of the mobile devices [2]. The screen size affects users mainly in two ways: First, only a limited amount of information can be displayed on the screen at once. Hence users often have to change the view (i.e., zooming in or out, switching between different screens). Second, collaboration with co-located persons is inherently limited, as only a certain amount of people can comfortably view the information.

[1] A demo video is available at `http://youtu.be/dZaCNV64ltk`

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 722–729, 2013.

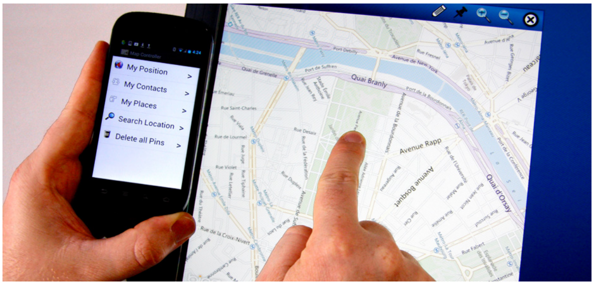


Fig. 1. Spanning a mobile user interface across the mobile phone and an external display, here showing a map application

In this paper, we contribute MobIES, a system that allows users to extend their mobile applications through temporarily spanning the user interface across multiple screens. In short, the technique requires users to touch the border of an available screen (e.g., a public display, TV, or desktop screen) with their phone during the interaction. The system detects this event and initiates the distribution of the user interface across the mobile and the external screen (see Fig. 1). Subsequently, users benefit from the extended screen space which facilitates tasks such as viewing a map, browsing the web, or showing and exchanging pictures to and with other users. When the phone is removed from the border of the external screen, the user interface returns to the original mobile mode. That is, users can take advantage of existing screens in their environments without the need to carry additional hardware.

The contribution of this work is twofold: in this paper, we present the concept and prototype implementation of MobIES and further, we present findings of a user study investigating the system by comparing it with the mobile condition.

2 Related Work

Early work on seamlessly connecting devices of different classes investigated how users can share information from their PDAs with others on a large shared device to support collaboration [3]. Integration of personal mobile devices with pre-installed devices in the environment has also been explored [12]. Ullmer et al.'s *mediaBlocks* showed how data attached to mobile tokens can be transferred to external devices [15]. Hinckley et al. demonstrated how multiple devices with touch screens allow users to drag-and-drop items from one device to another using the stitching technique [6]. Connecting large screens to mobile phones has been investigated [10] while other work focused on creating larger logical screens by combining several devices such as tablet computers [9] and considering the spatial relation of devices and users to each other [8]. Near field communication (NFC) has been used to detect the relative position of mobile devices to larger displays (e.g., [5,11]). Yet no work considered placing NFC tags around an external display which allows a novel way of interaction by using the displays of both devices together. Baur et al. present virtual projection which enables users to transfer data (e.g., pictures) from their phone to a large screen and display it thereon [1]. This approach allows users to take advantage of existing displays in their environment. However, the user's interaction is limited to the mobile device. Another approach is to distribute application interfaces on different devices and associated displays [4]. For instance, using mobile devices and large shared displays at which the phone is used as tool by touching the shared display in order to execute actions [13]. Our approach enables users to interact simultaneously with the phone and the extending display. In contrast to the discussed work, MobIES focuses on mobile situations in which the users have the need for more screen space to perform a specific task. The distribution of the user interface of the mobile application onto both devices - both allowing for interaction - increases the user's capabilities.

3 Concept

The concept of MobIES is based on users temporarily creating a physical and spatial connection between their mobile device and an external screen to create a larger logical display that consists of the *mobile interface* and an *extended interface* on the external screen. We assume that displays in the users' environments can temporarily be used (e.g., public displays, kiosk terminals, TV sets, interactive surfaces, and even screens in cars or airplane seats). User interfaces of mobile applications can display only a limited amount of information (Fig. 2 (left)). By connecting the phone with an external display more screen space is available, thereby allowing for the distribution of the user interface on two screens (Fig. 2 (middle)). Existing work that investigated connecting mobile phones and external screens did not consider the potential of using the mobile and the external screen simultaneously for displaying information. The event of connecting the phone with the display can be sensed, for instance, by using NFC tags that are placed around the external display which is a novel way to use NFC tags for device location detection.

Fig. 2. (left) The mobile user interface allows for the display of a limited amount of information. (middle) Connecting the mobile and an external display extends the available screen space. (right) Items can be shared with others via drag-and-drop to another connected mobile device.

While the phone is connected with an external display, sharing and exchanging data such as pictures, documents, or contact cards can be performed in a straightforward way. Given that the external display supports touch-based interaction, users can simply drag-and-drop items from the external part of their mobile application to the public space. For instance, this can be used in order to leave a message on a bulletin board. In addition, two users can exchange data by both connecting their devices to the same display and drag-and-dropping items from one phone to another (Fig. 2 (right)).

4 Implementation

Our prototype of MobIES consists of two main components. First, a server application running on a PC connected to a host application that is displayed on the stationary touch screen (Dell ST2220T, 22” screen (1920×1080 px)). Second, a mobile client (for Android) running on the user’s phone (Nexus S; 4” screen (800×480 px)). The server and the client manage the communication (via TCP over a wireless network) between the distributed application parts. Each application (e.g., a photo album) consists of a mobile component implemented as an Android application and a matching remote part implemented using the Microsoft Surface Toolkit. Depending on which application is active on the mobile phone when the phone touches the rim of the large display, the server launches a matching instance of the remote part of the application in the host application.

Fig. 3. Display border with NFC tags (left), covered with tape (right)

NFC tags are used to detect when a phone is placed on the border of the large display. NFC is supported by large number of different mobile devices (e.g., Samsung Nexus and Nokia devices). Every 50 millimeters, an NFC tag is placed on the display rim (see Fig. 3). When a phone equipped with an NFC reader is placed on the rim, it reads the tag content. This includes the position on the border, the display server’s IP, and the name of the wireless network. If the phone is not connected to the server application, the phone client establishes the connection with the wireless network and connects to the server. Finally, the phone client sends back the tag position and the ID or the currently active mobile application to the server which then launches the remote part of the application.

Using NFC tags allows for the extension of any existing screen to support MobIES interactions. This includes non-touch-enabled displays (e.g., public displays), as users can perform input on the phone while the external display extends the screen space.

5 Evaluation

We conducted a comparative user study to investigate to what extent MobIES supports users in performing typical mobile tasks. In particular, we were interested in gaining insights concerning usability and how participants perceive this extension of the user interface through holding the phone next to the extending screen compared to the familiar practice of using only mobile phones.

For the experiment, we implemented based on Schneider et al. [14] three applications that allow users to experience the MobIES concept. These include a *photo album*, a *map*, and a *web browser* application. All applications could be used with an additional external display or as a *stand-alone* mobile application using only a mobile phone. Using only the mobile phone without the extension of the user interface on an external display was used as a comparative condition for the practical tasks (in the following referred to as the *mobile-only* or *MO* option). The features of the applications cover standard functionalities inspired by existing Android applications.

Fig. 4. The photo sharing application: (left) extended overview; (middle) focus on a single image; (right) sharing images with another user by dragging an image from one extended interface to another

In the mobile mode, the photo album application enables users to organize photos taken with the phone in different albums. After selecting an album, contained items are displayed as small thumbnails. Touching a thumbnail activates the full screen mode. When the user launches the extended interface by holding the mobile phone next to the display border, the phone displays the album list and the extended interface shows an overview of picture tiles (Fig. 4 (left)). Which album is displayed can be selected using the list on the mobile interface. Selecting an item in the overview magnifies the picture to fill the application window on the extended interface (Fig. 4 (middle)). For the transfer of pictures from one mobile phone to another, users drag-and-drop items from one extended interface to another (Fig. 4 (right)).

The web browser application provides a history overview and supports tabbed browsing and bookmark management (in both modes). As the user connects the phone to the external display, the phone shows a menu containing options (e.g., History, Open Tabs) and the extended interface shows the corresponding content such as the list of bookmarks (see Fig. 5). For typing in text, the user can use a virtual software keyboard either on the phone or on the external display.

The map application enables users to display addresses of contacts on a map, as well as the selection of points of interests from a list, and searching for places.

Participants were asked to perform a number of tasks via MobIES and the comparative MO option while using a preconfigured mobile phone on which all required data (e.g., pictures or contacts) were available. With the photo album application, participants performed the following tasks: 1) Show the investigator pictures showing people from three different albums; 2) Search for the picture showing the {Eiffel Tower, rocks} in the albums and delete it. With the *map* application, participants performed the following tasks: 1) Find the Eiffel Tower / Tower Bridge on the map and show it to the investigator; 2) Show the investigator the addresses of two contacts from the

Fig. 5. The web browser application: (left) extended web page view; (middle) selecting bookmarks; (right) browser tab overview

address book as a pin on a map. For the third block of tasks, participants used the browser application: 1) Open the test webpage and look up the contact information of the author; 2) Add a test webpage to the bookmarks and check if the URL was added.

The investigator introduced MobIES and the MO option and participants practiced using them. Then the participants performed a series of tasks, once using MobIES and once as a comparative approach using mobile phones only (MO). The order of systems was counterbalanced and the task order was randomized. Participants filled in a questionnaire regarding usability, including the computer system usability questionnaire [7], after performing the tasks with each system.

We recruited 16 participants (5 females), aged between 20-33 (M=26). All participants were students with diverse fields of studies. All participants used smartphones with a touch screen and 14 reported having experience with multi-touch displays. They received 10 Euro after the study session which lasted an average of 45 minutes.

6 Evaluation Results

On average, each system condition was used for 20 minutes. After each trial, they filled in a questionnaire and rated the system (1=Strongly disagree; 7=Strongly agree). We used the non-parametric Wilcoxon signed-rank test to evaluate differences. Regarding (Q1) "Using the system, I could easily show information to other persons" participants rated the MobIES system significantly higher (Mdn=7.0) than the MO condition (Mdn=3.0) (z=-3.3, p=.001). Similarly, participants rated MobIES (Mdn=7.0) significantly higher than MO (Mdn=5.0) regarding (Q2) "The system supported sharing of information well" (z=-3.3, p=.001). Further, participants rated MobIES higher (Mdn=7.0) than MO (Mdn=3.5) in regards to (Q3) "The system supported jointly viewing of information well" (z=-3.4, .001). Yet both conditions were rated equally concerning (Q4) "Using the system, I often had to change my focus" (z=-.4, p=.72). One likely reason is that the larger screen space provided by MobIES spanned across two devices and thus required users to change their focus, much as using only the mobile phone requires switching between different views. Regarding Q5 participants rated MobIES significantly higher (Mdn=7.0) when compared to MO (Mdn=5.0) (z=-2.9, p=.004).

Participants rated both conditions using the IBM post study system usability questionnaire (1=Strongly disagree; 7=Strongly agree) that allows calculating four scores: OVERALL (the overall satisfaction score), SYSUSE (system usefulness), INFOQUAL (information quality), and INTERQUAL (interface quality) [7]. All score results are higher for MobIES: OVERALL (MobIES: 6.37; MO: 5.08), SYSUSE

(MobIES: 6.37; MO: 5.25), INFOQUAL (MobIES: 6.30; MO: 5.18), and INTERQUAL (MobIES: 6.34; MO: 4.58). The statements with the largest differences in the ratings cover the issues of system interface and task efficiency (see Fig. 6). S1 and S2 both indicate that participants appreciated the extended interface spanning across two screens as it was perceived as significantly more *pleasant* to use (z=-3.2, p=.001) and the organization of information was rated to be more clear (z=-2.6, p=.01). S3, S4, and S5 show that participants perceived MobIES as significantly more effective (z=-2.7, p=.007), efficient (z=-2.4, p=.01), and faster to use (z=-2.6, p=.008).

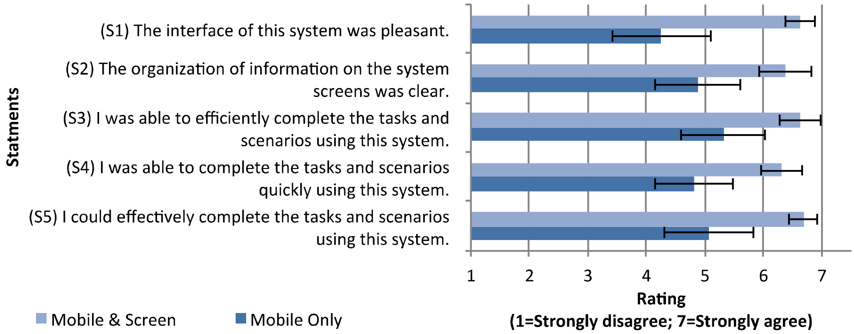


Fig. 6. Questionnaire statements with the largest differences in ratings

Six of the participants emphasized that they liked the level of clarity achieved through the larger screen space. Also, participants pointed out that extending the interface of mobile applications would be helpful to show or share information with others. One user suggested a holder for the mobile phone to leave both hands available for interaction. Four participants pointed out that they liked the ease of use of the system. For instance, P8 stated “It is very easy to switch between using only the mobile phone and using the additional display.” Few participants pointed out that they initially had to look for information after the user interface spanned across two displays. Yet all participants learned how to use the system quickly after a short introduction. Other participants highlighted that they liked the extension but expressed doubts whether an external display would be available when needed.

7 Discussion and Conclusion

MobIES addresses the issue that mobile users temporarily have the need for more screen space in selected situations, for instance, to gain more clarity when viewing large images or maps. The results of our laboratory study strongly indicate that users benefit from using this approach. Parameters that could not be mapped through an experimental setting, such as availability of matching external screens, as well as possible privacy and security concerns need to be considered when deploying such a system. The presented approach is based on a novel application of NFC technology that allows extending existing displays at very low costs. It enables users to take advantage of displays in their environments in order to extend the user interfaces of their mobile applications when needed. However, the presented implementation requires specific software to be available on the mobile and the stationary device, which limits the flexibility of users. To address this, future implementations could include a runtime environment on the stationary display system that executes application logic provided by the mobile client.

In a user study, we compared MobIES with the standard mobile phone option. The results indicate that participants appreciated the degree of information clarity, perceived their task performance to be faster, and highlighted that the system is easy to use. Future investigations will focus on providing a more generalized environment which allows users to take advantage of external displays that are not preconfigured.

Acknowledgement. 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. Baur, D., Boring, S., Feiner, S.: Virtual projection: exploring optical projection as a metaphor for multi-device interaction. In: CHI 2012. ACM (2012)
2. Chittaro, L.: Visualizing information on mobile devices. Computer 39(3), 40–45 (2006)
3. Greenberg, S., Boyle, M., Laberge, J.: PDAs and shared public displays: Making personal information public, and public information personal. PUC 3, 54–64 (1999)
4. Grolaux, D., Vanderdonckt, J., Van Roy, P.: Attach me, detach me, assemble me like you work. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 198–212. Springer, Heidelberg (2005)
5. Hardy, R., Rukzio, E.: Touch & Interact: Touch-based interaction of mobile phones with displays. In: Proc. MobileHCI, pp. 245–254 (2008)
6. Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., Smith, M.: Stitching: Pen Gestures That Span Multiple Displays. In: AVI 2004, pp. 23–31. ACM (2004)
7. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Hum.-Comp. Interact. 7(1), 57–78 (1995)
8. Lucero, A., Holopainen, J., Jokela, T.: Pass-Them-Around: Collaborative Use of Mobile Phones for Photo Sharing. In: CHI 2011, pp. 1787–1796. ACM (2011)
9. Lyons, K., Pering, T., Rosario, B., Sud, S., Want, R.: Multi-Display Composition: Supporting Display Sharing for Collocated Mobile Devices. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 758–771. Springer, Heidelberg (2009)
10. Rashid, U., Nacenta, M.A., Quigley, A.: The cost of display switching: a comparison of mobile, large display and hybrid ui configurations. In: AVI 2012. ACM (2012)
11. Reilly, D., Rodgers, M., Argue, R., Nunes, M., Inkpen, K.: Marked-up maps: combining paper maps and electronic information resources. PUC 10(4), 215–226 (2006)
12. Rekimoto, J., Saitoh, M.: Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments. In: CHI 1999, pp. 378–385. ACM (1999)
13. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: DIS 2012, pp. 318–327. ACM (2012)
14. Schneider, D., Seifert, J., Rukzio, E.: Mobies: extending mobile interfaces using external screens. In: Extended Abracts of MUM 2012, pp. 59:1–59:2. ACM (2012) (Demo Paper)
15. Ullmer, B., Ishii, H., Glas, D.: mediablocks: physical containers, transports, and controls for online media. In: SIGGRAPH 1998, pp. 379–386. ACM (1998)

Communicating in a Ubicomp World: Interaction Rules for Guiding Design of Mobile Interfaces

S. Shyam Sundar[1,2], Xue Dou[1,3], and Sangmee Lee[2]

[1] Media Effects Research Laboratory, College of Communications, The Pennsylvania State University, 122, Carnegie Building, University Park, PA 16802 USA
[2] Department of Interaction Science, Sungkyunkwan University, B307 International Hall, 53 Myeongnyun-dong 3-ga, Jongno-gu, Seoul, 110-745, South Korea
[3] Institute for Media and Communications Research, Keio University 2-15-45 Mita, Minato-ku, Tokyo, Japan #108-8345
{sss12,xud107}@psu.edu, sangmee7@gmail.com

Abstract. As computing resources become accessible anytime anywhere, rules of interaction and engagement between humans are changing. For example, response-time expectations have dramatically decreased in recent years because of the assumption that recipients are constantly checking e-mail and text messages on their mobile devices. Likewise, expectations of context-awareness have become an important part of interactions. For example, the tag-line "Sent from my iPhone" is a means of conveying the context (mobile) of the message sender, which also serves to explain—or offer an excuse for—the brevity of the response. In this manner, there are several design strategies that are needed for managing expectations, as new rules of interaction emerge due to the ubiquity of our access to computing resources. This paper presents a list of 12 such interaction rules based on theory and research in interpersonal communication and psychology. These rules provide design ideas for mobile ubicomp interfaces.

Keywords: Mobile HCI, Ubiquitous Computing, Interpersonal Communication, User Psychology.

1 Introduction

The arrival of "ubiquitous computing" has meant that we are constantly surrounded by a plethora of computing devices that allow us to communicate with others and access information anytime anywhere. While this offers a whole range of fascinating possibilities, we also face significant challenges of reorienting our communications and designing interfaces that better support people's daily activities in a ubiquitous computing (ubicomp) environment. An emerging research need in this area is to understand how ubiquitous computing influences the manner in which we interact with others. For example, we become more and more accessible to others through ubiquitous devices, such as smart phones and tablet computers. However the always-on envi-

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 730–747, 2013.

ronment also burdens our lives since the lines between our social and private lives become blurred [7]. In this paper, we discuss such challenges that ubicomp environments bring to the communication process. We begin with an example that illustrates how ubiquitous computing has changed our expectation of the way we interact with others. We then present some interaction rules that have emerged since the arrival of ubiquitous computing. Everyday examples are provided to illustrate how each interaction rule influences our communication process and the issues that it raises. Also, the paper discusses related studies in the literature that can provide insights into how we should address the issues. Finally, it proposes design recommendations for mobile interfaces and research questions that can be addressed in future studies.

2 Changes in Interaction Rules in Ubiquitous Computing Environments

How exactly did ubiquitous computing change the way we interact with others? A good example is our changing expectation with regard to e-mail response time. When laypersons started using e-mail in the 1990's, the infrastructure was not advanced enough to offer users easy access to their e-mail services, given limited computer availability, unstable Internet, and the high cost of using Internet. Due to this limited access, users did not expect others to check e-mails constantly. It was common to wait several days before getting a reply, so we seldom complained about e-mail responses that took a few days. However, in the first decade of the 21st century, accessing the Internet became much more convenient, and computers became available to many more people, in many forms, and in many places. With improved infrastructure, people now expect others to have easier access, thus raising their expectations for receiving prompt replies. Many users now expect a response the same day. Some users complain about others' lack of e-mail responsiveness, and other users complain about feeling pressured to meet e-mail senders' expectations (e.g., professors complain that students tend to send them a follow-up e-mail if they do not respond to their initial e-mail within a day). With the recent proliferation of ubiquitous computing, particularly smart phones, expectations for getting replies within a short period are even higher. Now, users are able to check e-mails almost anytime anywhere, and they assume that others have constant access to e-mail services as well.

This example demonstrates that advances in communication technologies have influenced our perceptions and expectations for interacting with others. In order to cope with such situations, users make adjustments and announcements in order to prevent conflict and promote communication. In the aforementioned case of e-mail responsiveness, users manage others' increased expectations by setting away-messages when they are not able to reply to e-mails for a couple of days. Some instructors tell students about their e-mail usage habit in advance so that students know what to expect (e.g., "I do not check e-mail after 5pm. E-mails sent after 5pm will be replied next morning").

Instead of such offline "announcements," the applications and devices themselves should be able to serve e-mail senders as well as recipients. This paper argues that mobile interfaces and other ubicomp devices can and should be designed to support users to overcome issues brought about by such changing rules of interaction. In the next sections, we present an initial set of 12 interaction rules and provide design recommendations for interface features that may help users manage their daily lives in ubiquitous computing environments. These rules are grouped under six categories based on conceptual similarity and commonalty in design solutions. We started out with a larger set of rules and identified interpersonal communication principles that are common across them. Those that shared a similar principle were grouped so that we can arrive at common design solutions for each group of rules as a whole.

2.1 Communicating Access to Manage Uncertainty and Others' Expectations

Communication in ubiquitous computing environments can be characterized as "always-on." Users can be reached by others at all times through different kinds of communication devices. At first sight, the ubicomp environment seems to bring a lot of convenience and flexibility for us to communicate with others. However, the "always-on" environment can also increase uncertainty in the communication process and heighten the pressure to meet others' expectations. Below, we describe two interaction rules that pertain to these issues.

Rule # 1: Users make inferences about other users' access to ubiquitous computing. Given the anytime-anywhere assumption, we expect prompt responses to our messages. However, we do not always receive responses promptly. There are some occasions when users are unable to access ubicomp or don't want to be reached by others (e.g., during a meeting) or do not use a particular Ubicomp tool. However, since we do not always tell others explicitly about our access to ubicomp resources and of our availability, those who initiate communications might undergo anxiety and uncertainty (e.g., worry about why the person did not reply to the message; thinking that she doesn't pick up the phone because she is mad at me).

In the same vein, people tend to be mindful of other users' ubicomp behaviors. They observe others' access and use of ubicomp tools in order to infer their individual behavioral norms and thereby manage their own anxiety and uncertainty. A dominant observation pertains to the temporal aspect, i.e., "when" a particular other will access ubicomp. Other observations span a wide range, from tool usage habits to contextual information (e.g., location of recipient, urgency of topic, one's schedule) [36]. For example, if you usually take one day to respond to e-mails, others will infer that you have daily, rather than round-the-clock, access to e-mail and will accordingly adjust their response-time expectations. Similarly, we often expect others to check e-mails less frequently during weekends. That is why we will not get upset if we do not receive a reply until Monday [36].

Another common inference pertains to others' media preferences. Ubiquitous computing environments offer users a vastly increased range of communication media. Previously, users had only a handful of ways for reaching others, such as talking face-to-face, calling through landline phones, and writing letters. The advent of the Internet expanded this choice set to include instant messaging (IM), e-mails, and social networking sites. However, these new types of media are not always available, given users' limited accessibility to computers. The proliferation of ubiquitous computing, particularly the rise of mobile devices, has set us free from such restrictions. We are now able to use a wide variety of communication media at any time and in any place of our choosing. We also have a rich choice set for deciding which medium to use for which occasion.

Selecting an appropriate medium is important since each medium has its own characteristics that make its use appropriate for some situations, but not others [22,39]. For example, if you have a casual acquaintance who happens to be on your list of Facebook friends, you would not think of calling them on a telephone to wish them on their birthday. This gesture is usually reserved for those individuals that we know at a more personal level. Instead, the norm would be to pen a greeting on the person's Facebook wall, so that the birthday wish is delivered in a less intrusive fashion. Of course, such decisions are predicated on inferences made about the recipient's use of these different media channels—something that the latter can control, in order to shape sender's expectations, as we specify in our next rule.

Rule # 2: Users develop their own methods or habits for gaining agency of ubicomp use, as a way of managing their availability to others. Meeting others' expectations is critically important for successful interpersonal communication. For example, expectancy violation theory (EVT) suggests that when an individual's expectation is violated, they actively evaluate the violation. If the violation results in a better-than-expected outcome (e.g., receiving e-mails quicker than expected), they evaluate it positively. However, in the event of a worse-than-expected outcome (e.g., receiving e-mails slower than expected), they evaluate it negatively [29].

In a ubicomp environment, managing others' expectations is a challenge. For example, we sometimes fail to reply to others as promptly as they may have expected. In such cases, the senders might feel frustrated at our lack of responsiveness and therefore form a negative impression. In fact, studies have already noted that responsiveness can influence the impression we leave on others [e.g., 36]. Therefore, it becomes important for users to manage others' expectations of their availability.

Many users do so by engaging in behaviors that block others' access to them (e.g., turn the sound off or switch off mobile phones). Users sometimes also choose who can reach them and provide selective access [1,10]. Previously, they were able to create boundaries between their different circles because the communication was mostly bound to place. For example, people interacted with their co-workers in their offices, while having time with their family or friends once they left the offices. However, ubiquitous computing blurs the boundaries between their circles by making interaction possible regardless of place. Therefore, users are now having to manage their availability based on their circles. An example of such availability management is to have two mobile phones--one for personal communications and one for work-related purposes. In this way, they can switch off the work-related mobile phone after hours, and only allow their friends and family to reach them through their private mobile phones.

Lying behavior also plays an important role in users' management of their availability. Studies have found that lies are sometimes necessary for facilitating interactions [2,37]. Hancock and colleagues introduced the concept of "butler lies" to describe the type of deception that helps users manage their interaction with others [16]. For example, a user might lie to others that she is busy in order to avoid starting a conversation on IM. Similarly, users sometimes make up reasons to explain why they were not available for a call (e.g., "sorry I couldn't pick up your call, I left my phone at home"). Their field study with 50 IM users revealed that one out of every ten messages has some sort of deception, and about one fifth of the deceptions are butler lies.

Design Recommendations. These rules suggest that users strive to make precise assumptions about others' ubicomp availability, but at the same time try to control their own availability to others. Ubicomp devices should be designed to help users to cope with these contrasting needs, as suggested below.

1. Allow users to communicate availability and media preference in a more fine-grained manner. Current technologies already offer some features that allow users to communicate their availability (e.g., status in IM). However, we propose that such indicators of availability should contain more detailed information, such as users' turnaround time. For instance, before we send an e-mail message to someone, our e-mail interface could inform us how quickly that person will be able to respond to e-mails on average during that day or week. In this way, message senders can make more accurate assumptions about the receivers' access to ubicomp, while receivers meet senders' expectations by setting their own average time by themselves (see left-hand side of Figure 1).

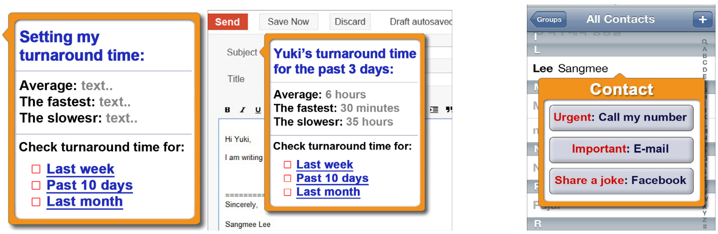


Fig. 1. Prototype of an interface that informs users about the turnaround time of the information recipient (left) and others' media preferences for contacting them (right)

Similarly, the design of ubicomp should help users communicate their media preferences to others. Specifically, a triage mechanism is needed for ubicomp interfaces, functioning as an intermediary channel (or initial filter) for aiding users' selection of particular media. For example, a system can allow users to show their media usage preferences to others. Each user customizes a list of preferences in advance. One user might specify that if a message is urgent, then the sender should call their mobile number, but if it is not urgent but important, then they should e-mail the recipient.

If they simply want to share a joke, then they should connect via Facebook. For another user, the media preferences for the same set of communications could be SMS, IM and E-mail respectively. If each user generates such a list of preferred media for common categories of communication, and if this can be seen by others at the time they decide to initiate communications with the user (such as when they click the user's name on their mobile phone contact list or put the user's e-mail address on their e-mail client), then message senders can choose the appropriate media for their interactions, thereby minimizing mismatches (see right-hand side of Figure 1).

2. Make users aware of their own ubicomp access and usage patterns. The interface could also help the user manage others' expectations of him/her by monitoring his/her response behavior and providing adequate advice and support. For example, the system can inform the user how quickly she typically replies to e-mails based on the urgency of the communicated event. If the user is replying to a non-urgent e-mail rather quickly, the system could notify the user and automatically retain the e-mails in the Outbox for a while before sending it out. In this way, the user can maintain consistency in his/her response-time, which would prevent others from holding false or unrealistic expectations that are likely to be violated.

3. Allow flexible customization. In addition, the design of communication interfaces should allow users to customize their availability based on their relationship with others [28]. Just as social networking sites these days allow users to share information to certain groups of friends, users' availability could be differentially indicated to different groups of contacts. For instance, mobile phone interfaces allow users to group individuals in their contact lists based on their circles. Users, then, can set which group of people can reach them based on time or occasion (see Figure 2). If a user receives a call from a person to whom the user prohibited access temporarily, the system could play a voice message or send a text message to explain their lack of availability.

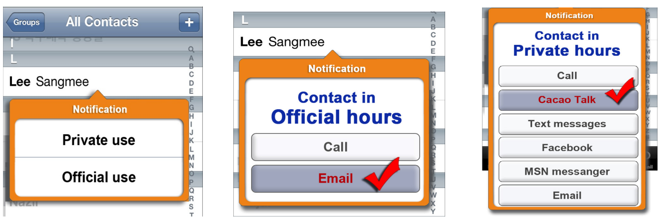


Fig. 2. Prototype of an interface that helps users to choose appropriate media for contact depending on individual or group of receivers

2.2 Providing Context-Awareness

Traditionally, contextual cues are relatively scarce in computer-mediated communication compared to face-to-face (FtF) communication [31]. However, with the development of ubicomp and communication technologies (i.e., video chatting, GPS function in mobile phone), the amount and type of contextual cues have changed, giving rise to new challenges and novel interaction rules aimed at coping with them.

Rule # 3: Users are oblivious to receivers' contextual constraints while communicating with them. Contextual information, such as place, time, and presence of others, provides cues for us to make judgments about what and how to communicate at any given moment [8]. Studies in computer supported cooperative work (CSCW) often highlight the role of contextual cues to establish common ground and facilitate communication among users [5,26]. Although CMC tools are generally low in contextual cues, older communication technologies, such as landline phones, still provide a certain amount of contextual information to users. For example, when we call landline phones to reach someone, we know where the person is (e.g., office or home). The call is sometimes picked up by somebody else, which is another piece of information telling us who is there in addition to the person we want to reach. On the other hand, ubicomp connects a person to another directly. In such a scenario, the amount of contextual cues that is available to users is dramatically reduced. As a result, callers will be less aware of the context of their communication partners. Calling somebody who is driving is a good example. Callers usually do not know whether the person is driving before initiating the call [3]. They might not even pick up on this cue if the receiver does not explicitly mention it (callers usually assume that if receivers pick up the phone, they are situated in a context that is appropriate for them to talk), and therefore may not factor in the divided attention that usually accompanies talking while driving, not to mention other contextual factors such as noise and passenger distractions. As a result, callers may not cadence their communication adequately to suit this situation, i.e., they may talk too fast or too softly, resulting in miscommunication or poor communication, or may not realize that they are on speakerphone and therefore reveal information that ought not to be heard by others in the car.

Rule # 4: Users expect context-awareness of interaction. Since it is becoming common knowledge that ubicomp deprives communication of contextual cues, some CMC interfaces have started enabling users to attach contextual information to their messages. For example, when a user sends e-mails from his/her mobile phones, an automatic signature at the bottom proclaims "sent from my iPhone (or Blackberry)". This kind of cue can communicate the fact that the sender was constrained by their device when sending out this message. If the message receiver knows that the e-mail is sent from a mobile device, they may show more tolerance towards short, cryptic sentences and greater willingness to overlook mis-spelling, grammatical errors and unorganized sentences. Further, such contextual information may serve to add value to the communication, and in some cases lead to greater enjoyment of content. For instance, when users see a picture uploaded in Facebook with the tag-line "uploaded from iPhone," they will infer that the information was shared soon after the event happened. Such recency cues could enhance the perceived freshness or timeliness of the communicated event [38] and thereby determine how the content of the communication is perceived.

Studies have attempted to identify contextual information that is important for influencing users' interaction with others through ubicomp [11,30]. De Guzman and colleagues studied types of contextual information that people use for deciding whether to initiate a call through mobile phones [11]. Their results showed that activity information (e.g., what activities receivers are doing at that time) was used most often to gauge one's availability to accept calls. However, they also found some discrepancies between types of contextual cues used by callers to make calling decisions and types of cues that receivers wished callers to consider before initiating calls. Specifically, receivers want callers to consider their task status (i.e., whether receivers are occupied by other tasks), physical availability (i.e., any physical barriers that prevent receivers from accepting the calls), and social availability (i.e., whether answering a call is socially awkward) more often than what callers normally would do themselves (34% vs. 22%; 21% vs. 18%; 15% vs. 9%, respectively).

It should be noted that people desire control over the degree to which information about their context is made available to others. Just because one is physically accessible does not mean that she is truly available for communication. In such cases, users may not truthfully let others know their contextual information, but would rather make up some white lies to manage their availability (e.g., avoid attaching information such as "sent from my iPhone" on their messages because they do not want others to know their accessibility via mobile phone at that time). Such lying behavior becomes necessary for facilitating interactions [2,37].

The control over the amount and types of contextual information shared is also important for protecting user privacy [9]. Increasingly, location sharing services (e.g., Foursquare) are offering users new ways of socializing with others, but they also bring with them important privacy concerns. According to a recent study, over 50% of individuals are concerned with privacy when they use location-based services [23]. Therefore, privacy needs are paramount when exchanging contextual information.

Design Recommendations. The examples and studies mentioned above illustrate the importance of context awareness for both senders and receivers of communication. Therefore, ubicomp interfaces should be geared towards allowing users control over how they communicate their context.

1. Give users control over when to share contextual information. Users should be able to have the control to release contextual information to others rather than letting devices automatically detect and transmit their situation to others. For example, a device could allow users to customize many types of "situation" settings (e.g., driving, having an important meeting, unable to access Internet/mobile phone) in advance. When necessary, users should be able to turn these settings on and off (see Figure 3).

Once a setting is turned on (see Figure 4), the device would automatically tell | others why a user is not available, and the device would stop attaching unwanted contextual information (e.g., if a user turns on the situation setting of "unable to access mobile phones", then the tag-line "sent from my iPhone" will not be shown on the message even when the user sends the message from his/her iPhone).

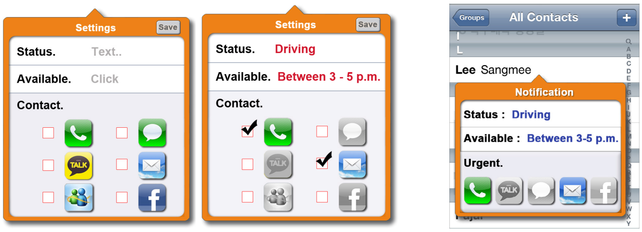


Fig. 3. Prototype of an interface that allows users to specify their availability and media preferences (left) and the manner in which it will be displayed to others (right)

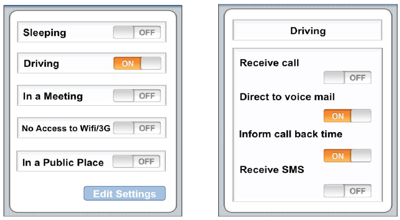


Fig. 4. Prototype of an interface that allows users to customize their situation and communicate their context to others

2. Give users control over how to communicate contextual information. Another important consideration is the manner in which contextual information is communicated. For example, when we share location information, most of the current applications show where a user is by way of detailed geographical information, such as street address or latitude. Sometimes, such information is indeed what users want to share, but not always. Designers ought to be more attentive to how people tailor the information when they decide to communicate it. For example, Lin and colleagues [19] pointed out that users use a rich variety of terms to refer to their location. If one is at Starbucks, she can refer to it as "a coffee shop" or "Starbucks near the public library" (labeled as "semantic information" in their study) instead of providing detailed geographic information. In addition, the decision regarding which term to use is based on various factors, such as the nature of the social relationship between information sharers and receivers, comfort level of sharing (e.g., privacy concern), recipient's familiarity with the place, and place entropy (i.e., how public the place is). Specifically, individuals use semantic information to refer to places when the information is shared with less intimate social groups, when users have higher comfort level of sharing the location, when they are not familiar or extremely familiar with the place (i.e. geographical information use is high when the familiarity level is moderate), and when the place entropy is low (i.e., the place is less public) [19].

2.3 Processing Information in Ubicomp Environments

Ubicomp gives us constant access to abundant amounts of online information from multiple sources. In addition, it lowers the hurdle of information sharing, which also increases the net amount of online information that we can access and use. Nowadays, it is common to see some users reading online news through their computers while checking on their friends' Facebook status updates through mobile devices. As a result, the amount of information that users encounter and deal with in their daily lives has increased significantly. Two interaction rules address this transformed information environment.

Rule # 5: Users will draw meaning from affordances. What happens when people need to deal with excessive amounts of information? Studies in psychology and communication have noted that individuals tend to process information heuristically (i.e., effortless and based on situational cues) rather than systematically (i.e. effortful and analytical) in online environments as a way to deal with the abundance of information [20]. The affordances or interface cues become important in heuristic processing because it helps users to make sense of the information instantly [33]. For example, when users visit a website for the first time, they often check the design of the site to form an initial impression. If the website design looks professional and easy to navigate, they are more likely to think that this website is credible, and thus decide to explore the content further. This positive impression of the site can extend to individuals and organizations that offer the affordances. For instance, great navigability in an organization's website imply greater attention to users' accessibility and convenience, and therefore project an image of caring on the part of the organization [33].

Rule # 6: Users will read meaning into others' use/nonuse of affordances (and ubicomp itself). The fact that one chooses to use an affordance (or not) can, by itself, convey meanings as well. Suppose a restaurant puts their customer ratings on their website. People who see the site will evaluate the restaurant not only through the content of the rating (i.e., four stars, the number of thumbs up), but also by the fact that the rating is even offered in the first place (implying that they care more about customers and place higher value on accountability). Other examples include affordances offered by Facebook, such as whether users use status update, have filled in their profile information, and uploaded pictures on the site. Let's take relationship status update as an example. If you are single but if you do not use this affordance (i.e., not declare that you are single on Facebook), it might signify that you are not proud of being single or not interested in getting into a relationship. On the other hand, if you are in a relationship with someone but do not use this affordance (i.e., not declare that you are in a relationship with so and so on Facebook), your partner might think that you do not care enough about the relationship with him/her or feel uncomfortable telling your friends about the relationship. In this way, the sheer use/non-use of interface affordances can lead to inference of meaning.

Design Recommendation: *Help users become aware of and unpack the meaning of affordances.* The issue with affordances in ubiquitous computing environments is that users are not fully aware of the way others "read" those affordances. Without having a mutual understanding of the meaning of a given affordance, users may face conflicts during the interaction (as in the example of the Facebook relationship-status symbol). In fact, studies have noted that the social nature of information sharing places impression management at the forefront [18,24,38]. Therefore, ubicomp devices should alert users about the consequence of their affordance use (see Figure 5) and help them manage their impressions through appropriate use of affordances. For example, whenever a user starts to use a new site that is open to the public, the interface could make suggestions about the types of affordances that the user can use to project their ideal image to others. The system can also inform users about how other people may interpret the affordance, so that the user can better manage their impressions.

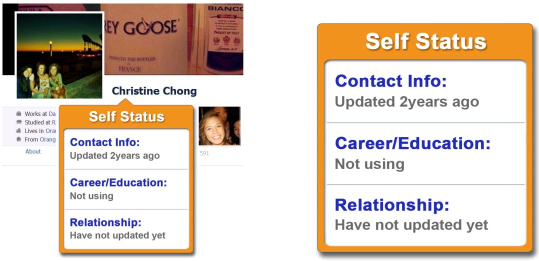


Fig. 5. Prototype of an interface that informs users about others' use/nonuse of affordances

2.4 Providing Awareness of Behaviors in Public

Use of computing devices in public places has increased dramatically in recent years. Nowadays, it is quite common to see people using mobile phones, laptop computers, and portable media players on the streets, at coffee shops, and in public transportation, leading to changes in interaction rules in public places.

Rule # 7: Users' perceptions of what is "appropriate" ubicomp usage behavior in public spaces are changing. Rule # 8: Users' tolerance for others' use of ubicomp during face-to-face communications is increasing. One noticeable change after the increased use of media in public places is our view of others' behaviors and our perceptions of what are appropriate things to do in public places. For example, previously, using mobile phones in public was considered impolite and improper. We used to get annoyed by others' phone conversations. But now, we seldom pay attention to people who are talking on their mobile phones in public places, and treat them as if they are talking with their friends walking next to them (Rule # 7). Likewise, we used to hear that people were irritated by the noise from other people's use of laptop computers in public places (e.g., tapping keyboards). But, we somehow got accustomed to these sounds, and using laptop computers in public places has now become very common and accepted in most places (Rule # 8).

Rule # 9: Our very conception of "public space" is also changing. In addition to the changing notions of "appropriateness" in our public behaviors, our notion of what is "public" itself seems to have undergone a change. Previously, public space was seen more as a place for interacting with others. People might chat with their friends or strangers co-located at that place or watch other people passing by. These types of interactions made individuals aware of others' co-presence in a public place. However, ubicomp access reduces the interaction with co-present others since users often flip open their individual mobile devices in public places. Such behaviors might make them feel that they are in their private space although they are physically located in a public place. As a result, users may feel that it is safe and appropriate to reveal their private information in public, making them vulnerable to privacy intrusions. Studies have noted issues resulting from people's use of personal devices in public places. First of all, individuals seem to decrease their serendipitous social exchanges with others when they use personal devices. Hampton et al. [15] found that wireless Internet users and mobile phone users are less attentive to people nearby and less approachable by others. Such decreases in social exchanges can hurt the formation of social capital [4] and hinder public discourse that is so essential to a democratic society [12]. Secondly, although people have increased their tolerance for others' personal media use in public, some behaviors, such as talking too loudly, can still cause conflicts. A report from the University of Michigan noted that more than 80% of people have been irritated by others' annoying manner of mobile phone use [35]. In addition, use of personal devices can make co-located companions uncomfortable since they may feel left out of the conversation [15]. Since such behaviors can strain relationships and impact impressions, we need solutions to help users to control their seemingly irresponsible and anti-social behaviors in public.

Design Recommendation: *Remind users about the "publicness" of their physical location.* The consciousness of public-private boundaries is fleeting in the ubiquitous computing environment. Therefore, interfaces should provide constant reminders of the "publicness" of their surroundings to users, so that they are constantly aware of what they should or should not say and do at any given moment. In addition, one's access of ubiquitous computing itself has plenty of meanings. Others would judge a user based on when, how, and with whom a user accesses devices. Thus, interfaces should alert users about their frequency of use, location of use, and the type of people around them while they are using the device (see left-hand side of Figure 6).

In this way, users can gauge the appropriateness of their device use and make necessary adjustments to their behavior. For example, using a mobile phone during talking with your school friends at a coffee shop might not be an issue. However, when it comes to a first date at a fancy restaurant, such behavior could imply rudeness or your unwillingness to engage with the date. To prevent this from happening, ubicomp devices could remind users where they are and who else is around. Such contextual information should be both automatically detected and inputted by users in advance. For instance, the location could be detected by the device, based on which it could

Fig. 6. Prototype of an interface that alerts users about the location and the number of people around them (left) and prototype of an interface that can be customized for pushing updates at regular intervals (right)

access the user's calendar and address book to ascertain the identity of the meeting partner. Another way of managing the boundary is to specify how often your system will push your received messages, missed calls, and other media updates to you (see right-hand side of Figure 6). This would effectively work in other public places as well, such as when traveling on a train at night, or when you have an important meeting but are also expecting a time-critical email message.

2.5 Supporting Users' On-Site Planning

The availability of anywhere-anytime information access has changed how we plan our daily schedules and plan for our future events. We propose two interaction rules relating to this topic.

Rule # 10: Users are less likely to stick to schedule. Previously, if we scheduled a meeting with someone, we often decided on a detailed plan (e.g., exactly what time to meet and where to meet). Since last-minute arrangements were difficult to make, individuals tried to stick to the plan and avoid unnecessary changes in their appointments. However, ubiquitous devices enable us to rearrange our schedules and appointments easily, which eventually influences our planning behaviors. Nowadays, it is common for us not to discuss detailed schedules and plans before an event. For example, it has become common for us not to decide on the exact place to meet in advance. Instead, we typically call each other at the meeting time to figure out where our partners are and decide where to meet.

Rule # 11: Users are more likely to access information at the time they need (just in time), and therefore are less likely to prepare in advance. We now depend less on pre-planning since we can schedule things at the point of service through ubiquitous devices. For example, we would previously make a detailed plan when we went on a trip, such as where to eat, which public transportation to take, and how to get to a particular place. Nowadays, we can easily find such information through ubiquitous devices at the time we need it. Such convenience means users are less likely to prepare things in advance (e.g., print out boarding passes before heading to the airport).

Design Recommendation: *Remind users about situational constraints that restrict their Ubicomp access.* Since people are getting so used to accessing information whenever they want, they tend to forget that accessing ubiquitous computing might be difficult in some places (e.g., weak Internet connectivity, low on battery). In addition, some information might not be accessible or hard to find on the Internet, so it is better for users to make necessary arrangements in advance on some occasions. Ubiquitous computing should predict such situational constrains and remind users about them. For example, when a user is entering an area with no wireless connection, the device could alert them so that the user can access whatever information they need before losing the connection. Similarly, once users tell a device what they need to do at their destination, the device could tell users what types of information they can and cannot access at the site so that they can accordingly plan their use of time before and during the event.

Also, in the context of a group rendezvous, it might be efficient and effortless to invite the attendees to a chat room and directly share their respective location information with all other attendees. Users will be able to instantly check where their friends are and how long their friends might take to get to the location. Moreover, they can call or send a message, supporting their on-site planning (see Figure 7).

Fig. 7. Prototype of an interface that provides location and contact information of others arriving to a meeting

2.6 Assisting Digitalization

Rule # 12: Users depend more on digital, rather than tangible, artifacts (what is digital is usable). Data digitalization is important in ubiquitous computing environments. In the last two decades, a variety of things have become digitally available, such as books, music, and videos. As a result, we have increased our dependency on digital content because it offers us many advantages. For example, digital content is easy to carry (e.g., students do not need to carry heavy books if they have a digital book reader with digitalized book content). Also, digital data provide easy handling of information search (e.g., using search function to find a file from folders or words in a document), information sharing (e.g., upload data online), and content editing. Given these conveniences, people, especially young users, have developed a preference for digital data over physical artifacts.

Studies have noted an increasing tendency among young users to prefer digital possessions over tangible ones. Odom et al. [25] found that teenagers prefer digital possessions because of accessibility (i.e., users can ubiquitously access data), accrual of metadata (i.e., users write comments or tag their friends on photos uploaded on Facebook as a way of accumulating social metadata), and easy presentation of self to multiple audiences (e.g., users customize their blog interfaces to convey their tastes and personalities to blog readers).

Design Recommendations

1. Support digitalization and retrieval of digitalized data. Given users' increasing preference for digital data, ubiquitous computing environments should make it easier to digitize things around users. Also, these digitized data need to be easily accessed across users' locations and devices. For example, users' personal devices should automatically share data instead of asking the user to download and sync the data every time manually (e.g., iCloud). The system should also make it easy for users to upload these digitized data onto social network sites, so that users can accrue metadata with their friends and use the digitalized data as a tool to present themselves to others.

2. Be aware of the negative outcomes of intangibility. On the other hand, as pointed out by [25], we also need to pay attention to negative outcomes of intangibility. It is human nature to develop an attachment to their possessions. Such attachment would make them cherish the possessions more (e.g., people refuse to switch their old mobile phone to a new one because they feel attached to the old one) and would also provide psychological comfort to users (i.e., children cannot sleep without their favorite blanket; grandmother remembers her children by seeing the clothes they wore when they were kids.) [21,27]. However, we currently have little knowledge about whether users are still able to establish such attachment with virtual possessions. Researchers and designers need to attend to this potential loss of attachment to virtual possessions and provide design solutions to overcome it.

One solution may be to apply the metaphor of offline possessions and the routines surrounding them. Packing things in our own backpack is a good example that can be applied in the digitalized world. By applying this pattern of behavior, users may have more attachment toward organizing and sharing data among systems. Suppose you are preparing materials for a class, you may need textbooks and apparatus. When you click the category of Apparatus on the interface, you will get a list of apps (see Figure 8). For textbooks, you can manage files based on lectures. If your applications do not support the material in digital form, type the list using *Notice* function so that you can make a list of needed tangible materials that could be prepared in advance. In this way, metaphors of analog possessions that are known to elicit user attachment can be used for designing tools for organizing and managing one's virtual possessions.

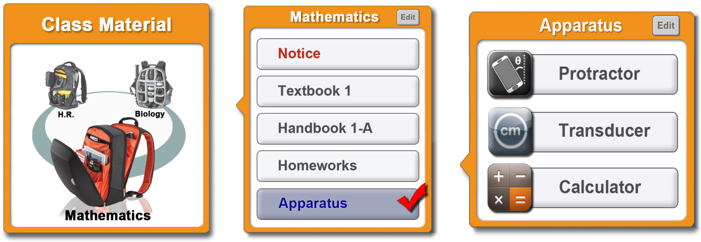


Fig. 8. Prototype of an interface that reminds users of traditional analogue materials necessary for an upcoming task

3 Conclusion

As our access to computing resources becomes ubiquitous, we are reshaping the ways in which we interact with each other. Previously unacceptable behaviors are now perfectly acceptable, but new rules of interaction also bring with them considerable communication challenges. We have called for specific design solutions to address these challenges, but the larger implication is that Ubicomp can benefit from a more unified user interface that treats human-computer interaction and human-human interaction as complementary, rather than competing, activities. The emphasis thus far in Ubicomp has been on ubiquity of access. It is time now to move beyond physical access to psychological access. As we graduate from usefulness and usability concerns into incorporating the social context of users [17], their emotions and sense-making while on the go ought to be considered priorities for research and design. We have to recognize that, given the emphasis on communication phenomena in the aforementioned rules, users will need interfaces that feature designs specifically capitalizing on ubiquity of computing access. These interfaces could come in the form of projected walls or body-parts in the future, but the mobile screen is the most dominant interface at the moment. Whatever we design for it, with the rules of Ubicomp, should be transferable to other Ubicomp interfaces as and when they diffuse into society.

References

1. Aguiléra, A., Guillot, C., Bonin, O.: Mobile Phone Use and the Management of Individual Reachability. In: Proc. of the COST 298 Project (2009), `http://www.abs-center.si/gbccd/papers/P062.pdf`
2. Aoki, P.M., Woodruff, A.: Making Space For Stories: Ambiguity in the Design of Personal Communication Systems. In: Proc. CHI 2005, pp. 181–190. ACM Press, New York (2005)
3. Avrahami, D., Gergle, D., Hudson, S., Kiesler, S.: Improving the Match between Callers and Receivers: A Study on The Effect of Contextual Information on Cell Phone Interruptions. Behaviour and Information Technology 26, 247–259 (2007)
4. Banjo, O., Hu, Y., Sundar, S.S.: Cell Phone Usage and Social Interaction with Proximate Others: Ringing in a Theoretical Model. The Open Communication Journal 2, 127–135 (2008)

5. Bardram, J., Hansen, T.: The AWARE Architecture: Supporting Context-Mediated Social Awareness in Mobile Cooperation. Paper presented at the ACM CSCW 2004, Chicago (2004)
6. Berger, C.R., Calabrese, R.J.: Some Explorations in Initial Interaction And Beyond: Toward a Developmental Theory of Interpersonal Communication. Human Communication Theory 1, 99–112 (1975)
7. Chen, B.X.: Always on: How the iPhone Unlocked the Anything - Anytime - Anywhere Future–and Locked Us in. Da Capo Press, Cambridge (2011)
8. Clark, H.H., Brennan, S.E.: Grounding in Communication. In: Resnick, L.B., Levine, R.M. (eds.) Perspectives on Socially Shared Cognition, pp. 127–149. American Psychological Association, Washington (1991)
9. Consolvo, S., Smith, I., Matthews, T., LaMarca, A., Tabert, J., Powledge, P.: Location Disclosure to Social Relations: Why, When, & What People Want to Share. In: Proc. CHI 2005, pp. 81–95. ACM Press, New York (2005)
10. Davis, S., Gutwin, C.: Using Relationship to Control Disclosure in Awareness Servers. In: Proc. GI 2005, pp. 145–152. ACM Press, New Yrok (2005)
11. De Guzman, E.S., Sharmin, M., Bailey, B.P.: Should I Call Now? Understanding What Context Is Considered When Deciding Whether to Initiate Remote Communication Via Mobile Devices. In: Proc. Graphics Interface 2007, pp. 143–150. ACM Press, New York (2007)
12. Gergen, K.J.: Mobile Communication and the Transformation of The Democratic Process. In: Katz, J. (ed.) Handbook of Mobile Communication Studies, pp. 297–310. MIT Press, Cambridge (2008)
13. Gudykunst, W.B.: Uncertainty and Anxiety. In: Kim, Y.Y., Gudykunst, W.B. (eds.) Theories in Intercultural Communication, pp. 123–156. Sage, Newbury Park (1988)
14. Gudykunst, W.B., Kim, Y.Y.: Communicating With Strangers: An Approach to Intercultural Communication, 4th edn. McGraw-Hill Humanities/Social Sciences/Languages (2002)
15. Hampton, K.N., Livio, O., Goulet, L.S.: The Social Life of Wireless Urban Spaces: Internet Use, Social Networks, and the Public Realm. Journal of Communication 60, 701–722 (2010)
16. Hancock, J., Birnholtz, J., Bazarova, N., Guillory, J., Berlin, J., Amos, B.: Butler Lies: Awareness, Deception, and Design. In: Proc. CHI 2009, pp. 517–526. ACM Press, New York (2009)
17. Höök, K., Isbister, K., Westerman, S., Gardner, P., Sutherland, E., Vasalou, A., Sundstrom, P., Kaye, J.J., Laaksolahti, J.: Evaluation of Affective Interactive Applications. In: Emotion-Oriented Systems, pp. 687–703. Springer, Heidelberg (2011)
18. Juhlin, O., Engström, A., Reponen, E.: Mobile Broadcasting: The Whats and Hows of Live Video as a Social Medium. In: Proc. MobileHCI 2010, pp. 35–43. ACM Press, New York (2010)
19. Lin, J., Xiang, G., Hong, J.I., Sadeh, N.: Modeling People's Place Naming Preferences in Location Sharing. In: Proc. UbiComp 2010, pp. 75–84. ACM Press, New York (2010)
20. Metzger, M.J.: Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. Journal of the American Society for Information Science and Technology 58, 2078–2091 (2007)
21. Mugge, R., Schoormans, J.P.L., Schifferstein, H.N.J.: Emotional bonding with Personalised Products. Journal of Engineering Design 20, 467–476 (2009)

22. Nardi, B.A., Whittaker, S., Bradner, E.: Interaction and Outeraction: Instant Messaging in Action. In: Proc. CSCW 2000, pp. 79–88. ACM Press, New York (2000)
23. Nielsenwire: Privacy Please! U.S. Smartphone App Users Concerned with Privacy When it Comes to Location. The Nielson Company (2011), http://www.nielsen.com/us/en/newswire/2011/privacy-please-u-s-smartphone-app-users-concerned-with-privacy-when-it-comes-to-location/html
24. Oeldorf-Hirsch, A., Sundar, S.S.: Online Photo Sharing as Mediated Communication. In: Proc. of the Annual Conference of the International Communication Association, ICA, Washington (2010)
25. Odom, M., Zimmerman, J., Forlizzi, J.: Teenagers and Their Virtual Possessions: Design Opportunities and Issues. In: Proc. CHI 2011, pp. 1491–1500. ACM Press, New York (2011)
26. Olson, G.M., Olson, J.S.: Distance Matters. Human-Computer Interaction 15, 139–179 (2000)
27. Parkes, C.M., Stevenson-Hinde, J., Marris, P.: Attachment Across The Life Cycle. Routledge, New York (1991)
28. Ståhl, A., Sundström, P., Höök, K.: A foundation for emotional expressivity. In: Proc. of the 2005 Conference on Designing for User Experience, pp. 1–21. AIGA: American Institute of Graphic Arts (2005)
29. Sheldon, O.J., Thomas-Hunt, M.C., Proell, C.A.: When Timeliness Matters: The Effect of Status on Reactions to Perceived Time Delay Within Distributed Collaboration. Journal of Applied Psychology 91, 1385–1395 (2006)
30. Shim, E., Dhara, K., Wu, X., Krishnaswamy, V.: Communication Data Based User Activity Recommendations. In: Proc. ICUIMC 2011, pp. 1–9. ACM Press, New York (2011)
31. Sproull, L., Keisler, S.: Reducing Social Context Cues: Electronic Mail in Organizational Communication. Management Science 32, 1492–1513 (1986)
32. Stephen, W.G., Stephen, C.W., Gudykunst, W.B.: Anxiety in Intergroup Relations: A Comparison of Anxiety/Uncertainty Management Theory and Integrated Threat Theory. Int. J. Intercultural Rel. 23(4), 613–628 (1999)
33. Sundar, S.S.: The MAIN Model: A Heuristic Approach To Understanding Technology Effects on Credibility. In: Metzger, M.J., Flanagin, A.J. (eds.) Digital Media, Youth, and Credibility, pp. 73–100. MIT Press, Cambridge (2008)
34. Sundar, S.S., Knobloch-Westerwick, S., Hastall, M.R.: News Cues: Information Scent and Cognitive Heuristics. Journal of The American Society for Information Science and Technology 58(3), 366–378 (2007)
35. Traugott, M., Soos, S.-H., Ling, R., Qian, Y.: On the Move: The Role of Cellular Communications in American Life: Pohs Report on Mobile Communication. Dept of Communication Studies, University of Michigan, Ann Arbor (2006)
36. Tyler, J., Tang, J.: When Can I Expect an Email Response? A Study of Rhythms in Email Usage. In: Proc. ECSCW 2003, pp. 239–258. ACM Press, New York (2003)
37. Vanden Abeele, M., Roe, K.: White cyberlies: The use of deceptive instant messaging statuses as a social norm. In: Proc. of the Annual Conference of the International Communication Association, ICA, Washington (2008)
38. Vartiainen, E., Väänänen-Vainio-Mattila, K.: User Experience of Mobile Photo Sharing in the Cloud. In: Proc. MUM 2010, pp. 1–10. ACM Press, New York (2010)
39. Woodruff, A., Aoki, P.M.: How Push-To-Talk Makes Talk Less Pushy. In: Proc. GROUP 2003, pp. 170–179. ACM Press, New York (2003)

Flow Specification Patterns of End-User Programmers: Lessons Learnt from a Health Mobile Application Authoring Environment Design

Filipe Fernandes, Luís Duarte, and Luís Carriço

LaSIGE & Department of Informatics, University of Lisbon
Edifício C6, Campo-Grande, 1749-016 Lisboa, Portugal
{ffernandes,lduarte}@lasige.di.fc.ul.pt, lmc@di.fc.ul.pt

Abstract. This paper discusses a set of interaction patterns encountered during the development of an authoring tool for mobile therapeutic applications. Unlike static paper artefacts, mobile applications can be enriched via the inclusion of complex behaviors. Typical examples include the definition of simple sequential interaction among all screens or the involvement of basic rules and triggers. As part of an ongoing project in which we are designing an authoring environment for mobile applications in clinical interventions, we studied how clinicians with no programming background were able to intertwine different screens from an application according to different rules. We were especially interested in comparing the approaches adopted using a low-fidelity prototype and using a high-fidelity version of the authoring tool. Results show that, despite a few technology induced strategies, users tend to mimic their actions using the paper based prototype in the corresponding hi-fi version.

Keywords: Authoring Tool, Non-Expert Programming, Participatory Design.

1 Introduction

Clinical applications for mobile phones have been shown to improve a patient's state [5][7][14][16][18]. The increasing dissemination and power of smart-phones has further emphasized this possibility [4][13]. Success cases for pathologies and therapy procedures as diverse as autism [2], fear therapy [7], aphasia [13] or obsessive-compulsive disorder [10] are testaments on the benefits of technology.

However, a significant number of these applications fall short to success for longer intervention periods [19]. Several factors can account for this outcome, among which the inability to personalize and adapt content [17]. For instance, an application's presentation is typically the same for all users who download it. Yet, the expectations of a potential 8 year old user are quite different from those of a 45 year old patient [7]. Also, the evolution of the patient's health status often requires adjustments that applications are not ready to accompany. For example, monitoring thresholds vary, support messaging and data collection should be adapted to new clinical assessments [16][17].

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 748–755, 2013.

The origin of this application stiffness builds on many factors. The complexity of the technology and of the application domain is certainly one of those reasons. In fact we believe that it is one of the most important factors: the dichotomy and complexity of knowledge involved. Information technology (IT) engineers and researchers understand technology and are able to handle its complexity. Clinicians on the other hand comprehend patients and the protocols they must put forward to provide them a better quality of life. Combining the two knowledge sources is no easy task. It gets worse because both knowledge domains evolve rapidly as well as the ultimate target, the patient wellbeing.

The usage of authoring tools is a possible solution. Past works prove valuable evidence in tackling similar situations [2]. These tools aim at joining two knowledge sources to a middle ground. They provide domain experts the mechanisms to customize and deeply adapt the applications, refining and embedding it with clinical decisions. For that, the tools must hide the technology complexity under a hopefully well-defined set of components, developed by IT staff. In the end they offer means for end-user programming [20]. Finding that adequate middle-ground can be complicated. It is not just about usability. It is also about programming and domain concepts, and ultimately the domain experts' perception of its combination. Available systems often recur to state charts or similar representations of the programmable elements [9]. Unfortunately and despite the existence of a few systems in this area, some design aspects were left unaddressed by researchers and IT experts alike. Our research focuses on one of these issues: the way information is organized as far as connections between programmable elements is concerned. We believe that sequential transitions between screens of a mobile application may be sufficient for some domains of intervention (e.g. mobile app prototyping [4][11][12]). Nevertheless, other domains possess a more critical nature which requires richer and more complex connections.

In this paper we present the design process of DETACH – DEsign Tool for smartphone Application Composition in the Health domain. The tool targets therapists and clinicians in general, and aims to provide them support to compose ubiquitous applications. For that, it offers a set of building blocks, as predefined, yet customizable, mobile app screens (e.g. a message panel, a mood selection panel). These can be assembled and customized to address the particular patients' needs. As such, domain experts are the tool's end-users and thus responsible to define the applications according to the adequate clinical procedures and evolution. Particularly relevant was the tool development process. We aimed at understanding what concepts and behaviors a domain expert formulates and expresses to grasp the programming endeavor itself. We were particularly focused in assessing how these end-users connect and define the control of flow between these building blocks, as this emerged as the most difficult task to assimilate. The assessment started in the early design stages and the understanding of users' behavior in low-fidelity prototyping, and transitioned into how that behavior was translated into the high-fidelity (hi-fi), and its evaluation with end-users.

2 Sketching DETACH: Participatory Design of Low-Fidelity Prototypes

The DETACH's development process started with the creation of specific mobile applications, addressing cognitive behavior therapy procedures. Currently several of those applications are being used in clinical trials and for all of them therapists provide continuous screening and comments [5]. Several brainstorming sessions were conducted including a team of expert HCI researchers and the therapists specializing in different types of interventions and pathologies.

DETACH's development process was heavily rooted in participatory design [1][13] and thinking aloud sessions not only with therapists but clinicians in general (Fig. 1). We have conducted a substantial set of meetings to understand: a) their current requirements and expectations for a tool of this nature; b) how modern smartphones can improve existing therapeutic practices without disrupting established procedures; c) interaction patterns employed by our stakeholders when organizing different elements of an application.

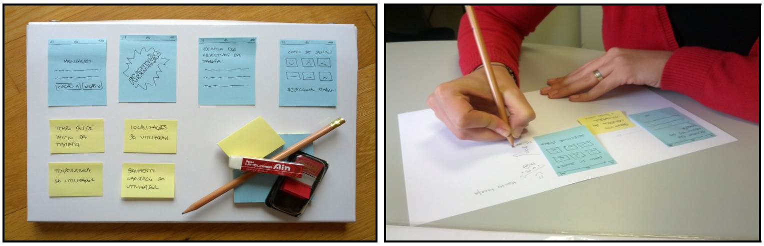


Fig. 1. Application representation made by one of the participatory design users

Design sessions lasted for approximately 45 minutes per subject, with each subject being accompanied by an expert UI researcher. A total of 15 users (aged M = 41; SD = 7.3) had volunteered themselves to participate in these trials.

For each trial, participants were asked to design an application to support a child undergoing a fear therapy [7] treatment and containing 4 screens: screen 1 allowed the child to select his / her current emotional state from a set of pictorial representations (e.g. happy face, frowny face, etc.); screen 2 presented a question (with "yes" / "no" answers) asking whether the child was still feeling scared or not; screen 3 was presented if the child answered "no" in screen 2 and displayed a descriptive support text regarding typical procedures for fear therapy; screen 4 was a congratulatory animation which was displayed either after screen 3 or if the child responded "yes" in screen 2. Subjects were assisted as required if they were insecure about screen details and other technical details. They were given full freedom to organize the screens.

Each session started with the researcher explaining the purpose of the authoring tool and its main features. Clinicians were then briefed regarding the mobile application they would be creating during the session. Finally they were provided with all the necessary material to create the screens: pen, pencil, eraser, different colored post-its and white sheets of paper (Fig. 1). In order for us to study how clinicians would deal

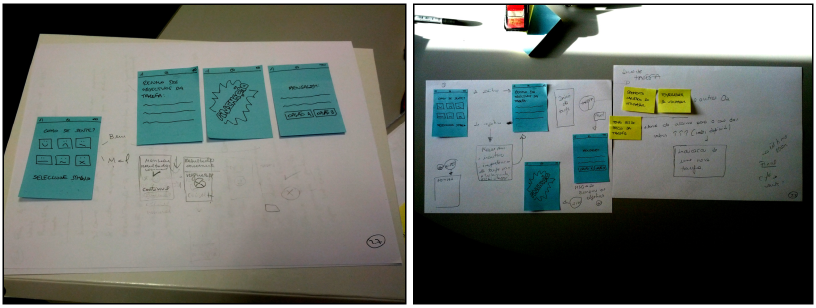


Fig. 2. Examples of screen transition representations on low-fi prototypes

with the provided material, some of the sessions had available additional template screen mockups and/or a guidance text of a specific application for them to represent.

2.1 Results and Discussion

Upon completing the creation of the mobile application using the low-fi DETACH prototype, subjects were asked to explain us the rationale behind their strategy to organize and connect the screens. Some notable approaches emerged from this trial:

- Subjects position each screen at their will on the sheet. Transitions between each screen are represented by arrows, complemented with side notes either using plain text or a small post-it above the arrow line (Fig. 2– right).
- Some participants preferred to organize screens on a sequential fashion, touching each other to pinpoint the transitions. Special connections such as the one stemming from screen 2 to screen 3 were addressed by separating one of the screens slightly from the main group and annotating the rules associated with that transition (Fig. 2 – left).
- New screens emerged. Interestingly some of them addressing physiologic sensing. More importantly a new type of transition and condition was created by some (more active) clinicians, featuring an interruption of the normal flow. The example in the figure (Fig. 2– right, lower left corner) refers to: whenever heart beat raises above something show this screen.

3 Testing the High-Fidelity Prototype

Based on stakeholder feedback, we designed the first DETACH's high-fidelity prototype (Fig. 3). The tool is a web application offering a workspace on which therapists can populate screens based on different templates and connect them according to a set of rules whose complexity may vary (e.g. from basic screen sequence to transitions based on previous patient answers). The top section displays the available screen templates (e.g. multiple choice answer, animation display, etc.). Therapists can drag a template into the center canvas and configure each screen's particular elements in the rightmost panel.

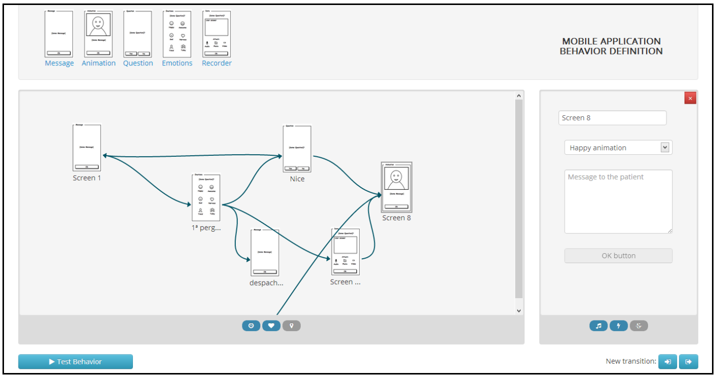


Fig. 3. Resulting representation of one of the thinking aloud sessions for the initial application

Transitions between screens (represented by arrows) are defined by selecting a screen, hitting a button and finally stipulating the rules which trigger the transition (e.g. if the transition considers the selected element as the destination or its source, how is the transition triggered).

3.1 High-Fidelity Prototype Trials

The high-fidelity prototype participatory design sessions occurred roughly 4 months after the low-fi ones. This trial was performed by the same clinicians who had been present in the previous experimental period. To provide the most pleasant and relaxed experience to our subjects, testing occurred again in the clinicians' offices. The high-fidelity prototype trials were carried out in a laptop (HP Pavilion dv9890ep model).

Regarding screen organization on the prototype's canvas, subjects adhered to two main strategies: a) akin to the low-fi trial, users positioned the screens sequentially, typically reflecting creation order; b) users delineated imaginary columns upon which screens were positioned – if a screen generates two navigational branches (i.e. users can transition towards two different screens from the same origin) then all destination screens would fit inside the same virtual column in the canvas, with the branches collapsing into another screen if that was the case. A substantial number of participants did not show any particular spatial strategy here (as observed in Fig. 3).

3.2 Results and Discussion

We were able to identify a substantial number of strategies employed by our participants regarding the organization of the elements they were able to interact with. While there was a slight degree of variation in the presented approaches, we were able to cluster them into well-defined interaction patterns.

Touch to Connect. One of the approaches observed rooted itself in the low-fi version of the prototype. Clinicians (approximately 25%) often aggregated screens which had a sequential transition nature together in the canvas. A particular behavior was noted:

subjects often organized the screen post-it in a way resembling a deck of cards, with each screen slightly touching each other it had a connection with. Transition rules were either described using side textual notes or by adding smaller post-its covering the screens involved. On the hi-fi prototype, these same users attempted to link screens using a different strategy: they dragged one screen towards another, "touching" it. The expectation was that a new transition was established between the "touched" screen and the dragged one. This strategy clearly shows a sequence oriented thinking towards building mobile applications. When confronted with the possibility of adding additional rules for these transitions (e.g. based on patient inserted content / answers) they argued that this still felt like the most natural way to interact with the elements in the canvas. They also reiterated the environment should leave the possibility of editing the transitions to add behaviors on top of the basic sequential connections (a feature already present, but not tested by these subjects).

Origin-Destination Paradigm. The most popular strategy adopted by users was inspired by the way they typically fill-in a postcard, an e-mail or a letter: they define the origin of the connection, the destination and then any related content with them. Approximately 55% of our subjects adopted this approach arguing "this is the way I naturally write" and "the way I did on the paper version". One may ask if using the same subjects and the previous experience with the low-fi prototype could influence this result: in part, we agree, but we must also note that a substantial number of clinicians did not follow the same connection strategies; also the timespan between both trials dissipates some of the "training" acquired in the first trial.

A minority of the clinicians (roughly 10%) approached this paradigm by completely switching the connection's order definition: they started by selecting the destination screen and then they picked the screens which would transit to the former. When asked to verbalize why they adopted this strategy, they argued "it made sense, considering a patient can reach the same screen from different branches", so "defining the destination first felt straightforward". Here, we must state such decision may have been influenced in part by the mobile application they were asked to create, since it featured a screen which could be reached from two different navigational branches. However, this behavior was only noted in the hi-fidelity prototype.

Connection from Screen Elements. The last adopted strategy pertains only to the hi-fidelity prototype participatory design sessions. When selecting and configuring each screen, some clinicians attempted to generate connections from the screen's components themselves (e.g. each answer, a button, etc.), justifying their behavior stating "the patient will transition to another screen if he / she presses this button". Even though no participant had previous programming experience, this is an approach reminiscent of existing Integrated Development Environments (IDE) such as Visual Studio or Eclipse and highly connected with event-based programming. In these tools, users may click a component, such as a button, to configure the application's behavior when the button is pressed. It is interesting that despite the absence of experience, some participants actually prefer this strategy.

Implications for DETACH. The main findings from these design sessions pertain to the variety of approaches clinicians were able to adopt to accomplish the same goal. Nevertheless we must prioritize the strategies which gathered more followers while at the same time not neglecting the preferences manifested by some of our participants. As such, DETACH's primary approach towards the definition of screen connections will follow the origin-destination paradigm. All other strategies will be incorporated into the environment as user preferences. Ideally, the application should be configurable for each user, storing their preferences and preferred workflow strategies.

4 Conclusions and Future Work

Our research points that non-expert programmers embrace desktop and paper metaphors in a virtual environment. We observed that our subjects employed similar screen organization strategies in the authoring environment regardless of operating a low-fi or hi-fi prototype during the participatory design sessions. It is important to stress that DETACH's final design will reflect these findings, to alleviate the technology transition impact which our stakeholders will be subject to.

We are finishing the development of DETACH, currently focusing on the mobile application emulator and the XML specification that will be used in Android smartphones to recreate the therapists' designs. The final step of this research will encompass a set of clinical trials in which we will assess whether authored applications can foster patient commitment when compared to previous non-authored digital artifacts.

References

1. Bergvall-Kåreborn, B., Ståhlbrost, A.: Participatory design: one step back or two steps forward? In: Procs. of PDC 2008, pp. 102–111. Indiana University, Indianapolis (2008)
2. Boujarwah, F.A., et al.: REACT: intelligent authoring of social skills instructional modules for adolescents with high-functioning Autism. SIGACCESS Access. Comput. 99, 13–23 (2011)
3. Boyd-Graber, J., et al.: Participatory design with proxies: developing a desktop-PDA system to support people with aphasia. In: Procs. of CHI 2006, pp. 151–160. ACM (2006)
4. Carriço, L., et al.: Therapy: location-aware assessment and tasks. In: Procs of AH 2012, Article 2, 5 pages. ACM Press (2012)
5. Carriço, L., et al.: A Mixed-Fidelity Prototyping Tool for Mobile Devices. In: Procs of AVI 2008, pp. 225–232. ACM, New York (2008)
6. Das, A.K.: Computers in Psychiatry: A Review of Past Programs and an Analysis of Historical Trends. Psychiatry Quarterly 79(4) (2002)
7. de Sá, M., Carriço, L.: Fear therapy for children: a mobile approach. In: Procs of EICS 2012, pp. 237–246. ACM (2012)
8. Grasso, M.A.: Clinical Applications of Hand Held Computing. In: 17th IEEE Symposium on Computer Based Medical Systems, pp. 141–146 (2004)

9. Hartmann, B., et al.: d.tools: Visually Prototyping Physical UIs through Statecharts. In: Ext. Abstracts of ACM Symposium on User Interface Software and Technology, UIST 2005 (2005)
10. Herman, S., Koran, L.: In vivo measurement of obsessive-compulsive disorder symptoms using palmtop computers. Computers in Human Behaviour 14(3) (1998)
11. Klemmer, S.R., et al.: SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In: Proc. of UIST 2000, San Diego, California, USA, pp. 1–10. ACM (2000)
12. Landay, J.: SILK: Sketching Interfaces Like Krazy. In: Procs. of CHI 1996, pp. 398–399. ACM (1996)
13. Matthews, M., et al.: In the mood: engaging teenagers in psychotherapy using mobile phones. In: Procs. of CHI 2011, pp. 2947–2956. ACM (2011)
14. Mattila, E., et al.: Mobile Diary for Wellness Management - Results on Usage and Usability in Two User Studies. IEEE Transactions on Information Technology in Biomedicine 12(4), 501–512 (2008)
15. Moffatt, K., et al.: The Participatory Design of a Sound and Image Enhanced Daily Planner for People with Aphasia. In: Procs of CHI 2004, ACM (2004)
16. Morris, M., et al.: Mobile Therapy: Case Study Evaluations of a Cell Phone Application for Emotional Self-Awareness. Journal of Internet Medical Research 12(2) (2010)
17. Offer, D., et al.: To whom do adolescents turn for help? Differences between disturbed and non-disturbed adolescents. Journal of the American Academy of Child and Adolescent Psychiatry 30(4), 623–630 (1991)
18. Przeworski, A., Newman, M.G.: Palmtop computer-assisted group therapy for social phobia. Journal of Clinical Psychology 60(2), 179–188 (2004)
19. Robinson, S., et al.: Aftercare intervention through text messaging in the treatment of bulimia nervosa - Feasibility pilot. Intl. Journal of Eating Disorders 39(8), 633–638 (2006)
20. Lieberman, H., et al. (eds.): End-User Development. Springer (2006) ISBN-10 1-4020-4220-5
21. Timms, P., et al.: Cognitive Behavioral Therapy. In: Leaflet for the Royal College of Psychiatrists' Public Education Editorial Board (March 2007)

MoCoShoP: Supporting Mobile and Collaborative Shopping and Planning of Interiors

Julian Seifert[1], Dennis Schneider[2], and Enrico Rukzio[1]

[1] Ulm University, Ulm, Germany
{julian.seifert,enrico.rukzio}@uni-ulm.de
[2] University of Duisburg-Essen, Essen, Germany
dennis.schneider_1@uni-due.de

Abstract. We present MoCoShoP, a system that supports the collaborative process of shopping and planning furniture and interior items. The system consists of a mobile application running on the users' mobile phones and an interactive surface application deployed on shared planning desks in the furniture retail store environment. Users belonging together share a virtual shopping cart. By scanning labels attached to furniture items with their phones that are of interest, users can inspect item details (e.g., dimensions, available colors) with the mobile application and add items to their shopping cart. The shared planning desk allows users to collaboratively review collected items and create possible arrangements of items on a floor plan. Finally, users can store furniture arrangements for later inspection. In this work, we contribute the design and a prototype implementation of MoCoShoP. Results of a first evaluation indicate that users appreciate how they can collect and share data during the process of shopping and how it supports collaborative planning.

Keywords: Mobile phones, interactive surface, collaboration, shopping assistant, collaborative planning.

1 Introduction

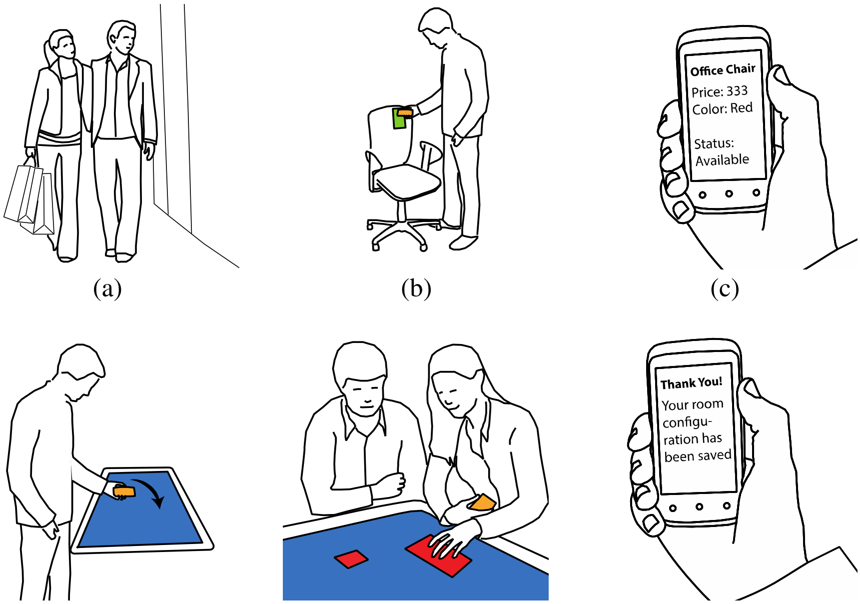
Online shopping is more popular than ever and recent numbers indicate that this trend is continuing [3]. The list of reasons for this success includes high flexibility for customers who wish to compare prices of products, access to detailed information on products (e.g., availability, possible configurations, dimensions), and social aspects such as easy access to other customers' ratings and reports on experiences with a product.

Many types and groups of products are well suited for online shopping. For instance, previews on media files such as music or movies can be provided and thus, customers get a clear idea of what they are going to purchase. However, other artifacts cannot be previewed in an adequate way due to their specific physicality or other inherent aspects that cannot be communicated. Accordingly, many customers prefer visiting retail stores as they allow the touching, testing, and experiencing of a product.

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 756–763, 2013.

This is in particular the case for pieces of furniture that must fit into an existing setting of other previously acquired pieces of furniture. Additionally, they need to meet the customer's personal criteria such as taste or comfort. In retail stores, customers can check these criteria and gain hands-on experience with products. On the downside, retail stores have different drawbacks compared to online shops: detailed product information such as prices, available configurations, etc. are difficult to access. Also, planning how different products would fit into a room with existing pieces of furniture is difficult.

We present MoCoShoP, a system that allows customers to experience the advantages of retail stores (e.g., physical and hands-on experiencing of products) and combines these with the benefits of online shopping (e.g., information access, social shopping). MoCoShoP provides a mobile client application that runs on the customers' mobile phones, which allows for access of product information via network and provides a shared shopping cart (e.g., with family members) if desired. Further, the system provides an interactive planning desk which supports collaborative creating of product arrangements and floor plans containing the collected products. In the following, we illustrate the usage of MoCoShoP with a usage scenario.



(a) (b) (c)

(d) (e) (f)

Fig. 1. Usage scenario for MoCoShoP: Multiple users go shopping together (a). Users pick up information by scanning labels (b) and (c). Users transfer collected items to a planning desk (d) and create plans containing interesting products (e). Finally, they save a planning arrangement and purchase items (f).

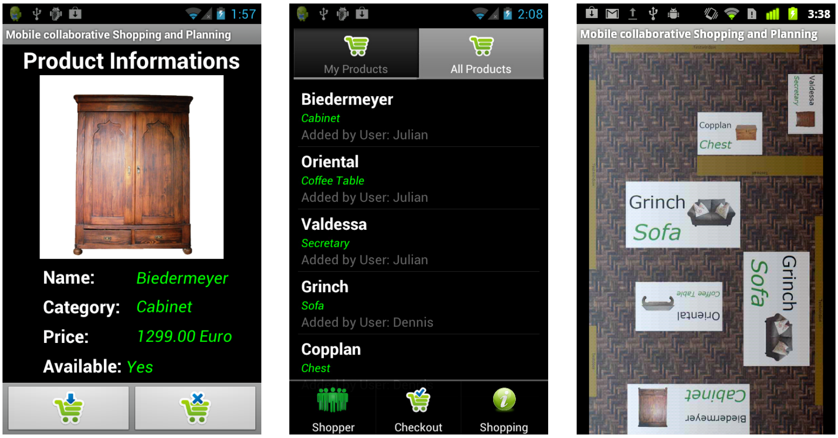
Scenario: Alice and Bob are planning to buy additional pieces of furniture for their office. In order to look for possible items, they go to a furniture retail store (see Fig. 1a). Both Alice and Bob use the MoCoShoP mobile client on their mobile phones to scan and check out prices and available settings of products (Fig. 1b and 1c). When they have collected and added enough items to their cart, they approach the collaborative planning desk and transfer the items to the desk through a touch gesture (Fig. 1d). On the planning desk, Alice and Bob try different configurations and floor plans with selected products (Fig. 1e). When they agree on a configuration including which items to buy, they save the configuration back to their mobile phones (see Fig. 1f) allowing for further item collection or for the purchase of the selected items.

In this work, we contribute the design and a first implementation of MoCoShoP. Further, we contribute the results of an initial user study.

2 MoCoShoP Application

The design goals of MoCoShoP are (a) supporting quick information access in retail environments, (b) providing awareness of other users actions to support collaboration, (c) support for collaborative planning and reviewing of potential room plans including purchasable furniture items.

In order to meet these design goals, MoCoShoP includes two components for interaction: a personal mobile client application for each user and a shared interactive planning desk.



(a) (b) (c)

Fig. 2. The MoCoShoP mobile application: (a) Product details screen after scanning a product label. (b) Shared shopping cart overview. (c) A floor plan of a configured room including arranged pieces of furniture.

2.1 Personal Mobile Client

The mobile client runs as an application on the user's mobile phone. It allows users to scan product labels in order to access related detailed information. In order to scan a product ID, the user holds the phone close to the corresponding label which allows the phone to read a Near Field Communication (NFC) tag that is integrated into the label. NFC is based on the Radio-Frequency Identification technology and allows storing of data on a chip that is powered via a capacitive field created by the reading device. This technology is included recently in an increasing number of smartphones (e.g., Nexus 4). As an alternative, printed barcodes could be used to include a larger number of potential smartphones (e.g., the iPhone). When a product label has been scanned, the application retrieves product details and provides an overview (see Fig. 2a). Users can choose to add the product to their shopping cart or simply reject the product. Multiple users can create a joint shopping session which allows them to add products to a shared shopping cart (see Fig. 2b). By selecting an item from the product list in the shopping cart, users can inspect the corresponding product information or delete the item. The mobile client also allows the storage of product lists and floor plan configurations that were created on the shared planning desk (see Fig. 2c).

2.2 Collaborative Planning Desk

When users have added potentially interesting products to their shopping cart, they can transition their shopping activity towards a planning activity which is supported by MoCoShoP through the collaborative planning desk. The planning desk is an application that is running on an interactive multi-touch surface, allowing multiple users to work together. First, one user of a group touches the planning desk on the device border with their mobile phone. The mobile phone reads a specific NFC tag which initiates the transfer of collected product IDs to the planning desk application.

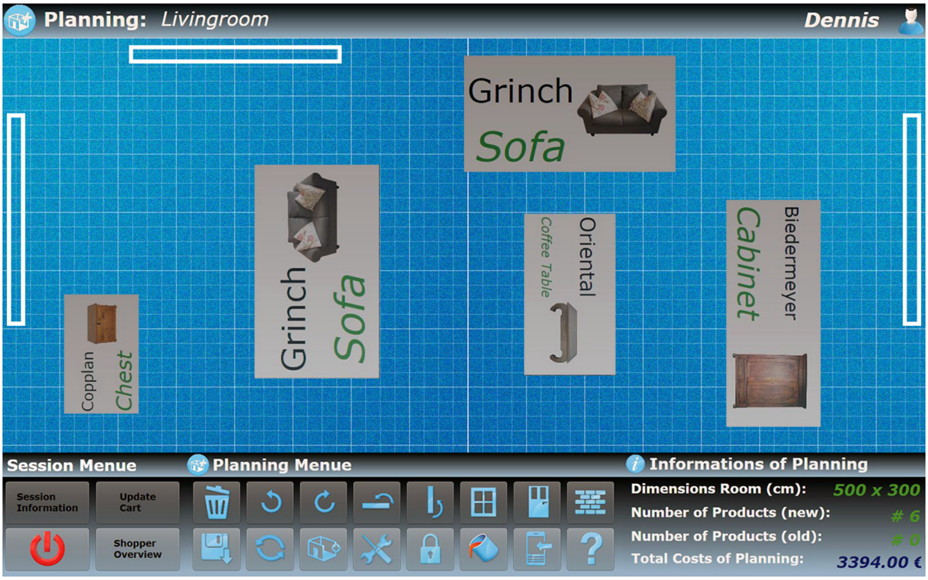


Fig. 3. The collaborative planning desk application provides a touch-based interface

The planning desk application provides a large canvas which represents a floor plan of the room which the user would like to configure, and thereby plan which pieces of furniture would fit into it (see Fig. 3). The application allows users to quickly rearrange and configure such a floor plan. The interface provides information such as how much money the items cost that are included in the current configuration. In addition, the application provides a number of tools that support the users throughout the planning task. For instance, buttons which rotate items, align, or delete them are provided. Finally, when users are satisfied with their design, the store, the floor plan, and the data are transferred back to their mobile devices.

3 User Study

We conducted an initial user study in order to gain insights on if and how users would appreciate such a collaborative shopping and planning system such as MoCoShoP. In particular, our aim was to gain an understanding of how the system would support collaboration during the shopping and the planning process of furnishing when compared to the current practice of using pen and paper in order to collect information and plan during the shopping process.

Session Organization. Initially, participants were introduced to the aim of the study. Then, participants performed two practical task in counterbalanced order. Once they used the MoCoShoP system and once they used only pen and paper. This pen and paper condition was selected for comparison as it represents an approach most users are familiar with. In order to investigate the collaboration support by MoCoShoP, participants would perform these tasks as pairs of two. After finishing each task, participants were asked to fill out a questionnaire regarding their experiences with the used approach.

Practical Tasks. Participants performed one task with each condition (MoCoShoP; pen and paper). The tasks required participants to select, collect, and plan furniture items for a room (a living room and a bedroom). Both tasks were similar in terms of the actions required: first, users were given instructions such as how much money they could spend and what pieces of furniture should be included. Second, the two participants started walking through the experimental shopping environment. We equipped two laboratory rooms with 69 labels attached to the walls representing available furniture items (see Fig. 4a). There, participants looked for items suitable for their planning task. Whenever participants found interesting items they could add them to their shopping lists. When using MoCoShoP, they used smartphones which were provided with the mobile client application installed. In the pen and paper condition, participants were required to take notes manually (see Fig. 4c). Further, participants should plan a room layout including the selected pieces of furniture one time with the MoCoShoP planning desk (see Fig. 4b) and one time using pen and paper (see Fig. 4d).

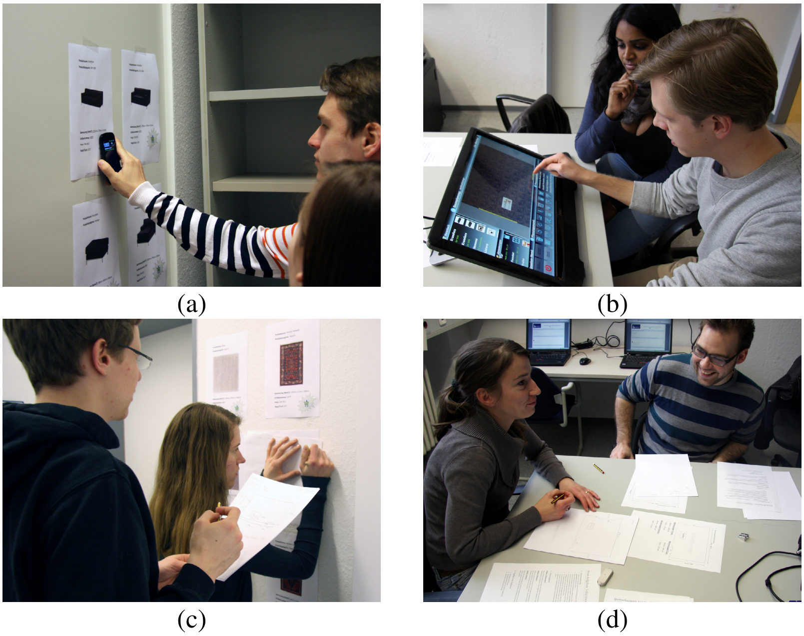


Fig. 4. Interaction during the evaluation tasks: (a) Using the personal mobile client to collect product information. (b) Collaboration on the planning desk. (c) and (d): Collecting information and planning a room outline using pen and paper.

Apparatus Implementation. We implemented a prototype of MoCoShoP for the study. The mobile client application was developed for the Android platform running on a Samsung Nexus S (4" screen, 800×480 px) mobile phone that provides an NFC module for the scanning of product labels. The collaborative planning desk (Dell ST2220T, 22" screen (1920××××1080 px)) was developed based on the Microsoft Surface 2.0 SDK which provides support for multi-touch interfaces. For the storage and management of product information, a web server provided an interface for the retrieval of corresponding information. Further, a session management server was implemented to store information related to shopping sessions (e.g., list of items in a shared shopping cart).

Participants. We recruited 14 participants who worked in pairs of two. They were aged between 23 and 33 years old and seven of them were female.

3.1 Results

All participants expressed that they liked how fast it was to access product information by scanning a label. Several users expressed that a shared shopping cart is helpful in situations when collaborators split up to search for different products by creating a kind of awareness for the other users' activity or location. As expected, most participants appreciated the flexibility provided by the planning desk application which allows users to create many different confections easily.

13 participants stated that shopping and planning furniture items is a collaborative activity they perform together with other people. This reinforces the identified design goal that collaboration support is needed for shopping for furniture items.

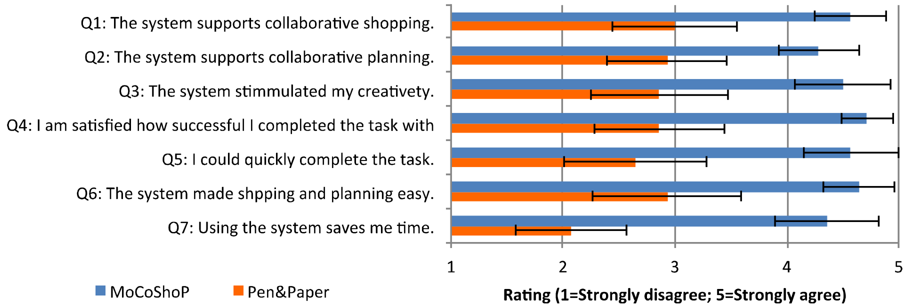


Fig. 5. Post-hoc questions comparing MoCoShoP and the pen and paper condition (Error bars indicate the standard deviation)

Participants rated MoCoShoP significantly higher (on a 5-point scale; 5=best; tested using the Wilcoxon Signed Ranks test) compared to the pen and paper condition regarding the support for collaborative shopping ($z=-3.13$; $p=.002$), collaborative planning ($z=-2.87$; $p=.004$), and perceived creativity stimulation ($z=-3.1$; $p=.002$) (see Fig. 5). Further, participants rated MoCoShoP significantly higher in terms of successful task completion ($z=-3.22$; $p=.001$), time required to complete the task ($z=-3.21$; $p=.001$), support to make the task easy ($z=3.21$; $p=.003$), and the perceived system ability to save the user time ($z=-3.33$; $p=.001$).

4 Related Work

The concept and interaction techniques applied for MoCoShoP are grounded in a number of existing and related works. Early work by Rekimoto investigated the pick and drop interaction technique [4]. The touch and interact technique advances the touch-based interaction to mobile phones based on NFC technology [2]. PhoneTouch generalizes cross-device (touch-based) interaction [5] as adopted by MoCoShoP.

Mobile phones have been demonstrated to be suitable devices for mobile recommendation systems to overcome the limitations of traditional retail stores [6].

Additionally, mobile phones have been used [1] for the visualization of customer-specific information on products (e.g., a diabetes shopping assistant). Similar to MoCoShoP, the system SoloFind allows users to collect information on products in a retail store for further inspection on a kiosk computer [7]. In contrast, MoCoShoP incorporates different classes of devices for specific tasks, allows in- formation access via the mobile device, and supports collaboration on the shared planning desk.

5 Discussion and Conclusion

We presented MoCoShoP, a system that supports customers in retail stores during the process of collecting information on potentially interesting pieces of furniture, and further, during the process of planning how the collected products could fit into their devised layout. While the personal mobile devices are used for information collection, the large interactive surface is used for collaboration and shared discussion.

Our prototype implementation of MoCoShoP demonstrates that the effort for deploying such a system is moderate and existing environments can be easily augmented: product labels with either integrated NFC tags or simply printed barcodes are low-cost factors and interactive surfaces to be used as planning desks will be relatively cheap as technology matures. MoCoShoP combines the benefits of e-commerce and traditional retail stores to improve the user experience.

Acknowledgement. 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. Broll, G., Palleis, H., Richter, H., Wiethoff, A.: Exploring multimodal feedback for an NFC-based mobile shopping assistant. In: NFC 2013, Zurich, Switzerland. IEEE Computer Society (2013)
2. Hardy, R., Rukzio, E.: Touch&Interact: touch-based interaction of mobile phones with displays. In: MobileHCI 2008, pp. 245–254. ACM (2008)
   Lipsman, A.: $41.9 Billion in Q3 2012 U.S. Retail E-Commerce Spending (November 2012), `http://www.comscore.com/Insights/Press_Releases/2012/11/comScore_Reports_Q3_2012_U.S._Retail_E-Commerce_Spending` (last access March 25, 2013)
3. Rekimoto, J.: Pick-and-drop: a direct manipulation technique for multiple computer environments. In: UIST 1997, pp. 31–39. ACM (1997)
4. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: DIS 2012, pp. 318–327. ACM (2012)
5. von Reischach, F., Guinard, D., Michahelles, F., Fleisch, E.: A mobile product recommendation system interacting with tagged products. In: PERCOM 2009, pp. 1–6. IEEE Computer Society (2009)
6. Wiethoff, A., Broll, G.: Solofind: chains of interactions with a mobile retail experience system. In: CHI EA 2011, pp. 1303–1308. ACM (2011)

Usability and Utility Needs of Mobile Applications for Business Management among MSEs: A Case of Myshop in Uganda

Rehema Baguma[1], Marko Myllyluoma[2], Nancy Mwakaba[2], and Bridget Nakajubi[3]

[1] College of Computing & Information Sciences, Makerere University
rbaguma@cit.mak.ac.ug
[2] Nokia Research Centre Nairobi
{marko.myllyluoma,ext-nancy.mwakaba}@nokia.com
[3] Grameen Applab-Uganda Office
bridget@applab.org

Abstract. This paper discusses the usability needs of mobile applications for basic business management for Micro and Small Scale Enterprises (MSEs) in developing countries. This is based on results from a user study carried out in Uganda on 30 MSEs. The study was carried out on MyShop, an easy to use mobile business management application for cash transactions and book keeping designed for micro-entrepreneurs. The study investigated learning to use MyShop, the support MyShop gives to the user and its usefulness, and value addition to users. The study also covered the pleasure and stimulation MyShop gives to users. Results from the study show that MSE owners/shopkeepers would like an application that is easy to use such as have an intuitive navigation and a simple and clear language. They would also like an application that supports their unique context like multiple people operating a shop, selling goods on credit, ownership of multiple businesses, use of low end phones and regular load shedding. In terms of value addition, MSEs would like the application to assist them in managing the daily operations and developing the business in the long term such as marketing, time saving and control over business.

Keywords: Usability, mobile applications, utility, MSEs, MyShop, business management.

1 Introduction

Micro and Small Scale Enterprises (MSEs) play important roles in the economic growth and sustainable development of every nation (Moore et al., 2008). The development and growth of MSEs can provide solutions to the problem of high unemployment facing many countries especially developing countries. Maseko and Manyani (2012) noted that MSEs have low start-up costs, low risk and can exploit untapped knowledge bases of creativity in the population for new product development. According to East African Community (EAC) (2010), MSEs are often considered a key source of productivity, growth and job creation. They promote equitable

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 764–773, 2013.

distribution of income because they are owned by relatively low income entrepreneurs, most of whom are women. MSEs are therefore important instruments for both income distribution and equitable participation in the process of economic development. For governments, well-managed and healthy MSEs are a source of revenue in form of taxes. Uganda has 0.8 million MSEs that contribute 20% to GDP and employ over 1.5 million people (EAC, 2010). In Uganda, MSEs are categorized according to number of employees, value of capital investments and turn over as follows:

Table 1. MSE Definition used in Uganda

	Number of Employees (Max)	Capital Investments (Max in UGX)	**Revenue (Max in UGX)**
Micro	4	12m (USD 4800)	12m (USD 4800)
Small	50	360m (USD 144000)	360m (USD 4800)

From EAC (2010)

As mobile phones are multiplying in the developing world, many social economic services and applications are being innovated by Mobile Network Operators (MNOs), not for profit and for profit organizations. Hellstrom (2010) noted that in a region where the vast majority have limited resources among other constraints, mobile phone solutions provide a means of extending a number of services to the poor who are in rural and remote areas. However, Corbett (2008) remarked that as sales, access, and coverage continue to grow, it is yet to be seen whether the mobile phone technology will play a significant, sustained role in alleviating poverty in the developing world. In most developing countries such as Uganda, while operators' focus has been people with a high disposable income and revenue generating services like urban youth and entertainment, services and development oriented applications have not been prioritised. Therefore the number of relevant applications for economic and social development is still limited.

On the other hand, efforts to develop mobile applications for banking and managing business internationally have been focused on applications for high end phones such as iPhone (e.g. Mint and Splash Money), Android (Xero) and ipad (quickbooks mobile). These sophiscated mobile applications are out of reach for MSEs in developing countries due to their high cost, hi-tech nature (most MSE owners are computer illiterate), energy efficiency (battery consumption) and data dependency (online connection). As a result there is need for applications that can run on low end phones that most MSEs in developing countries use. In addition, most available applications are mainly about personal financial activities such as banking than managing business activities of a business. For example iPhones's Ledger Docs is used for digital capturing of ledger receipts. Mint, another iPhone application, updates account balances, transmits real-time alerts to the phone for low account balances, over-budget, and tracks strange buying and selling activity. Android's Xero keeps the user updated about business finances and cashflow, creates real time reports as well as integrated invoicing for precision management.

To-date, most research on MSEs and mobile phone usage has been focused on the communication function of the phone and mobile commerce. Examples of such include; Kwaka (2012) who reports about mobile phone usage by MSEs in semi-rural

Ghana; Melchioly and Aebø (2010) reports about added value concerning economic growth and entrepreneurship to SMEs in terms of communication with suppliers and customers, market solicitation, social cohesion and operational cost reduction; Esselaar et al (2008) reports on ICT usage and its impact on profitability of SMEs in 13 African Countries. Others are Escobari and Donner (2009), Matambalya and Sussana (2012), Dorflinger et al (2009), Edim and Muyingi (2010) and Giridher et al (2009).

Hence little so far has been done about using mobile phones for business management. Palsey (2011) reports about a mobile book keeping application developed as part of her Honours studies to enable MSEs keep records of their business transactions and automatically generate financial reports. However, the application in question is not available in the public domain. Therefore it is not clear whether it went beyond the university laboratory to be used by the target users.

Business management for MSEs like other business types is important for planning, budgeting and forecasting, taxation, access to finance among other functions. EAC (2010) noted that the quality of information that a firm produces and maintains determines its access to finance. However, EAC (2010) found out that majority of Micro and Small Scale Enterprises do not keep proper records due to lack of appropriate skills. Therefore efforts to provide simple business management tools tailored to the needs and characteristics of MSEs are greatly needed.

This paper discusses attributes for an easy to use and beneficial (utility) mobile application for business management among MSEs in developing countries. This is based on results of a user study carried out in Uganda on MyShop. MyShop is mobile business application for cash transactions and book keeping designed for micro-entrepreneurs. The study covered ease of use and utility evaluation of MyShop. It investigated learning to use MyShop and the phone (Nokia c3-01), support MyShop gives to the user, the pleasure and stimulation MyShop gives to users and the usefulness, and value addition of MyShop to users.

Nielsen (2012) refers to usability as a quality attribute that assesses how easy the user interfaces are to use. According to Schumacher and Lowry (2010), usability is the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in a particular environment. Utility on the other hand refers to the design's functionality i.e. whether a product or service does what users need (Nielsen, 2012). Ease of use and utility are equally important and together determine whether something is useful. Nielsen (2012) noted that it matters little that something is easy to use if it is not what you want and it is also no good if the system can hypothetically do what you want, but you cannot make it happen because the user interface is too difficult. Therefore the attributes discussed in this work are concerned with ease of use and utility needs of mobile applications for business management among MSEs. In particular, this is discussed in the perspective of the MyShop application.

2 Methodology

The ease of use and utility evaluation of MyShop was carried out on a sample of 30 MSEs dealing in products and services in and around Kampala city, Uganda. The study was conducted between November 2011 to March 2012. The study areas included one urban location and three peri-urban areas namely: Kampala city centre, Kawempe trading centre, Kyaliwajjala trading centre and Natete trading centre.

The rationale for selecting one urban location versus 3 peri-urban ones was because most MSEs in Uganda operate in peri-urban and rural areas.

The 30 participants were chosen purposively based on the following attributes:

ability to read and write in English. The minimum requirement was completion of Ordinary Level (Middle School) but completion of Advanced Level (High School) was preferred. This attribute was important because the application was in English. The second requirement was ability to use Short Message Service (SMS) or mobile money service. This gave an indication of experience with data applications. In addition, some individual user characteristics were preferred in order to be sure that the participant would perform and fulfill all expected tasks. These included: able to attend trainings, willingness to use the application in day to day business, willingness to commit at least 1 hour daily to update information and commitment to meet data collection requirements. Other desired attributes included: located 10 - 15 kms from central Kampala, small scale dealing in products and services, 1+ years of operation; 10+ transactions/day, owner operated or employing not more than 5 people and between 200 and 2000 USD start up capital.

Three weeks to the beginning of the study, participants were trained (half day) in using the application and phone. After the training, they were given phones loaded with the application to use for three weeks before beginning of interviews. In addition, they were given a user guide covering using the main menu and application settings, and a summary of how each of the application's functions work.

The study was conducted through weekly field visits to the study sites during which, the study team observed how participants were using the application in their cash and book keeping activities. In addition, 3 questionnaires on ease of use, reliability and enjoyment, and utility were administered to participants to elicit quantitative and more detailed qualitative data. There was no requirement for approval from an Institutional Review Board before running the study.

3 Discussion of Findings

3.1 About MSEs That Participated

The study covered one urban area and three peri-urban areas around Kampala city. The study population was 30 MSEs dealing in products and services. Of the 30 participants, 53% were male and 47% female. Fifty percent had a degree or diploma, 33% had completed Advanced Level (High School) while 17% had completed Ordinary Level (Middle School). All the participants could read and write in English and could use SMS or mobile money. The business premises were located 10-15 KMs from central Kampala and 83% had been in operation for 1+ years. All participating businesses had 10 + transactions/day, were owner operated or employing not more than 5 people and 59% of the businesses were worth 10 million Uganda Shillings and above (approximately $4,200 USD). In the findings discussed below, the names given are not actual names of participants for purposes of protecting the identity and businesses of participants.

3.2 Usability and Utility Needs of Mobile Applications for Business Management among MSEs: A Case of Myshop in Uganda

This section discusses ease of use and utility needs of mobile applications for business management among MSEs in developing countries. This is based on results from a user evaluation study of MyShop. MyShop was developed by Nokia and has four main functionalities: selling stock (called “Sell”), recording sales and expenses (called “Records”), purchases and inventory management (called “Stock”), and monitoring of business performance over time (called “Reports”). It is available for free in Nokia’s online store on http://store.ovi.com/content/249179.

The needs discussed are categorized into: ease of use, support to the user, usefulness and value addition to MSE owners/shopkeepers and the pleasure, and stimulation to the user. According to the Webster online dictionary, ease of use is the property of a product or thing that a user can operate without having to overcome a steep learning curve. High ease of use makes products or systems intuitive to use to the average user. In the context of computer/mobile applications, ease of use simply means an application that is effortless to understand, find what is needed and do what one wants to do. In addition, the user’s context such as characteristics of the environment, language preference, cultural or community practices, unique business characteristics, etc., need to be understood and supported for better usability. These attributes are categorized as support to the user in this work. Further more, for a product, service or system to be useful, it should be both easy to use and of value to the users (Nielsen, 2012). Hence value attributes are also important and are classified as value to the users in this work. Besides ease of use, supporting users’ context and being of value, a product, service or system should be pleasing and motivating to users to succeed.

3.2.1 Ease of Use

This subsection discusses attributes that can make mobile applications for business management effortless to use for MSE owners/shopkeepers. These include:

Using the application as transactions happen: MSE owners/shopkeepers prefer to use the application as transactions happen. In the MyShop usability evaluation study, 80% of the shopkeepers interviewed preferred to use the application as transactions happen. They noted that this helps them avoid forgetting details of transactions performed as Mariam, a shopkeeper from Kyaliwajjala trading centre remarked: “*I use it as I transact my business throughout the day because I need to ensure that all records have been entered*”. In line with this need, developers of mobile applications for business management among MSEs need to focus on making it easier and faster to record transactions because efficiency is very important to shopkeepers.

Auto Computation of Tasks Such as Profits and Losses: MSE owners like the auto computation of tasks by computer/mobile applications. This is much easier and less error prone compared to manual computations on paper. Jolly a shopkeeper in Kawempe trading centre equipped: “*auto calculation of sales, profits and losses makes my life easy*”.

Intuitive Navigation: MSE owners want a well organized interface that they can easily navigate and find functions needed. Participants in the MyShop study found it easy to use the simple navigation structure of MyShop (4 items at the first level and

3-6 items inside each of the screens on the second level). Betty a saloon operator in Kawempe trading centre said: "*I like the fact that it is easy to navigate through the application and know where to find what function when I need it"*.

Simple and clear language: MSE owners/shopkeepers would like mobile applications written in simple and clear business management language. This is preferred because most of them have low levels of education and do not have formal training in book keeping. In the MyShop study, 68% strongly agreed that the language used in MyShop is clear and simple which made the application easy to use. Violet, a drug shop attendant in Kawempe trading centre noted: *"the application uses simple business language and the English is clear"*.

Representative icons: Using icons that are a clear representation of intended function makes such applications easy to use for MSEs. In the MyShop study, most participants found the icons used meaningful in relation to the intended function. Jumba, an operator of a crafts shop in Kyaliwajjala trading centre said: "*the icons explain well what function they represent e.g. using a symbol of money notes to mean sell clearly shows that whatever is under that function is money coming in"*. A key thing about the understandability of the icons is that they are not global abstract metaphors but tangible everyday items. For example, instead of a shopping cart, there is a basket and for stock, the initial metaphor of a delivery truck, was replaced with a pile of boxes.

Easy to learn: MSEs find applications that are easy to learn easy to use. In the MyShop research study, 24% strongly agreed and 48% agreed that learning MyShop was easy which made use of the application easy. This was attributed to the ease of finding functions, simple and clear language, representative icons and an easy to operate phone.

3.2.2 Support to the User

This subsection discusses usability attributes that address contextual needs of MSEs from mobile applications for business management. These include;

Multiple user accounts: Some MSE owners have employees hence it is important for the mobile application to support multiple user accounts with varying privileges. This will enable shop owners/managers to restrict access to information on certain aspects such as reports. In the MyShop study, 80% of the participants preferred using the application alone (which meant running a parallel manual system) because the version used did not support multiple user accounts.

Portability: MSEs prefer the mobile application to a computer application due to its portability. This helps them carry out offsite business management activities like restocking more easily. In the MyShop study, 35% strongly agreed and 57% agreed that they enjoy using MyShop in business partly due to its portability. Agnes, a grocery shop attendant in Kampala city centre noted: "*I go with it when shopping for new stock and I use it to remind me about restocking needs"*.

Easy to operate phone: The phone(s) on which the application works must be easy to operate. In the MyShop study, 28% strongly agreed and 40% agreed that the phone on which MyShop was loaded, was easy to use. This was attributed to the semi-touch screen phone (small touch screen and large physical keys) and participants' previous experience with other Nokia models. Isaac, a video library operator in Kawempe trading centre noted that: *"It is very easy to find functions with the touch screen"*. This finding is not isolated because White (2010) noted that less educated and less technology savy mobile phone users find touch screens easier to use because they directly manipulate interface objects rather than using third party interaction aids like a mouse and keyboard.

Short introduction to bookkeeping: EAC (2010) in a study about MSEs in E. African countries found that majority of micro and small enterprises do not keep proper records due to lack of appropriate skills. In the MyShop study, the "Reports" module was used less because most participants confused the "Records" module to offer similar functions to "Reports" and or did not understand how it works. For example Christine, a beverages shop operator in Kyaliwajjala trading centre said: "*I really did not understand how the reports function works and since I can assess the business performance with records then I do not need to use reports"*. The user manual in the version of MyShop used did not cover book keeping. Including an introduction to book keeping in the help module, can benefit those new to book keeping concepts.

Provide cloud or local backup storage: To prevent against data loss in case of loss of the phone or a problem with the hardware or software, it is important to provide for data back up locally or in the cloud. The MyShop version that was used in the study did not have this provision.

Device independence: It is important to make such applications compatible with low end phones so that users are not limited on device choice. In the study, the Nokia c3-01 used had a short battery life (which was a problem to participants due to regular load shedding). In addition, Nokia c3-01 costs approximately USD 200 hence is expensive for most MSE owners but the current version of MyShop runs on lower end Nokia devices that cost as low as USD 50.

Provide for selling services: The MSE sector has both product and service businesses with varying operational procedures. Therefore it is important for such applications to provide for both. One way is to have a product and service version with services offered and details of offering in the place of stock. The MyShop version used was only applicable to MSEs dealing in physical products such as retail shops. Therefore it was a challenge using it on businesses dealing in services like video libraries.

Provide for credit transactions: Selling on credit is a common practice especially for retail shops in Uganda and other African countries. The MyShop version used in the study did not have this provision and participants expressed a need for it. Betty, a cosmetics shop operator in Kawempe trading centre said: *"It is very difficult to keep track of items sold on credit, yet we sometimes offer goods to on credit"*.

Local language version: Most MSE owners are not comfortable with English due to low levels of education. UNESCO (2007) reported that one of the challenges of delivering mobile phone based services is that 41% of the population in developing countries is non-literate and even the literate among the poor are typically novice users of computer technologies. Due to this limitation, the study purposively selected participants whose level of education ranged from Ordinary Level (Middle School) and above. Therefore for this application to succeed in the market like Uganda there is need for it to have a version in a language. In the study, participants requested for a version in Luganda, the most dominant local language in Uganda.

Use one install for two or more businesses: Multiple business ownership is common among business people in Uganda and Africa. Ronald, a video library operator in Natete trading centre said: *"I have a video library and a phone accessories shop hence I would prefer being able to use one copy of MyShop for both businesses"*. Therefore it is important to make it possible for one install to be used for two or more businesses for example through a multi business edition.

Link application to mobile money: Like other business people, MSEs make a lot of financial transactions such as paying suppliers, utilities, employees, receiving payments from customers, etc. However, the banking infrastructure in Uganda like other developing countries is still severely limited and very few people can meet requirements of banking institutions (FinScope, 2007). Participants in the study revealed that some of the payments they make in their businesses such as paying suppliers, rent, and receiving payment from some customers are settled using mobile money. Therefore such a function would make the application more useful and relevant for MSEs.

3.2.3 Usefulness and Value addition to MSE Owners/Shopkeepers

This subsection discusses attributes that address relevancy and value addition attributes MSEs desire from mobile applications for business management.

Marketing function: MSEs would like an application that can on top of basic book keeping, perform related functions such as marketing. This can be achieved through mass dissemination of marketing information via Short Message Service (SMS) or multimedia messaging. During the study, some participants used the phone camera to take pictures of their stock for marketing to potential customers. Moses, an operator of an arts and crafts shop in Kyaliwajjala trading centre noted: "*I take pictures of current stock and sold out items to show to potential customers"*.

Reduction of workload: MSEs would like an application that reduces their work load thereby making work easier. 60% of the study participants agreed that using MyShop had made work easier in their business such as in stock taking and tracking profits and losses. Deborah, a supermarket operator in Kyaliwajjala remarked: *"I track the stock automatically within the application which less tiresome*".

Time saving: MSEs desire applications that help them save time in the management of their businesses creating more efficiency. In the study, over 60% agreed that MyShop saves time in business management because of automated functions. Joshua an airtime supplier from city centre remarked that: "*MyShop has promoted efficiency in business information recording/book keeping"*.

Safe keeping of business information: MSEs would like an application that keeps their business information safe and secure. Katende, a shopkeeper in Kawempe trading centre remarked: "*My business is individual therefore I like it that MyShop keeps my business information away from third parties"*. Therefore it is important for such applications to restrict access to some information like reports.

Control over business: MSEs would like the mobile application to help them better control their business such as keeping track when away and easy assessment of performance. Jackson, a mobile video and disco operator in Natete trading centre noted: "*the business' performance is easily assessed since everything is automated"*. This can be further enhanced by making the app an online service that syncs across multiple shopkeepers so that the business owner can monitor while away.

3.2.4 Pleasure and Stimulation to the User

This subsection discusses attributes that will make mobile applications for business management pleasing and motivating to use for MSE owners/shopkeepers.

Exciting experience: MSE would be attracted to an application that gives them an exciting experience. This can be achieved through adaptation of the application to their business needs. In the MyShop usability study, 17% had an extremely exciting experience while 62% had an exciting experience. This was attributed to relevancy of MyShop and the semi-touch screen phone which most of them were using for the first time. Miriam a shopkeeper in Kyaliwajjala trading centre said: *"MyShop helps me make business decisions based on facts e.g. every day, I look at MyShop to find out what needs restocking before I restock".*

Enjoy using the application: MSEs would like an application that they enjoy using. To achieve this, the application must be relevant to their needs and context. In the MyShop study, 35% strongly agreed and 57% agreed that they enjoy using MyShop in business due to its portability and auto computation of tasks. Doris, a mini supermarket operator in Kyaliwajjala said: *"I go with it when shopping for new stock and I use it to remind me about restocking needs".*

4 Conclusion

The potential value of Mobile applications for business management among MSEs is enormous but for this to be realized such applications must be usable and have high utility to the target MSEs. For MSEs in developing countries such as Uganda, such applications must be effortless to use, relevant to the needs and context of MSEs add value to their businesses and be pleasurable and stimulating to use. Future efforts on such applications should focus on usability and utility for the target market segments.

Acknowledgements. We extend our sincere thanks to Nokia that allowed us to use results from an originally non-academic usability evaluation study on MyShop in Uganda. We also extend thanks to Grammeen Applab-Uganda Office in particular Sean Paavo Krepp, the Country Director and Annete Bogere, the Content and Services Manager. Grammen-Applab coordinated the study in Uganda as a local partner on behalf of Nokia. Our since thanks also go to Gabriel White who worked as a User Experience Consultant and Richard Nuwagaba, and Andrew Gagwera who worked as Research Assistants on the study.

References

1. Donner, J., Escobari, M. (no date). A review of the research on mobile use by micro and small enterprises (MSEs) ICTD 2009, Qatar, April 17-19 (2009)
2. Dorflinger, J., et al.: Mobile Commerce in Rural South Africa-Proof of Concept of Mobile Solutions for the Next Billion Mobile Consumers. In: Proc. WOWMOM 2009 (2009)
3. Dim, A.E., Muyingi, H., N., A.: Mobile Commerce Application for Rural Economy Development: A case study for Dwesa. In: Proc. SAICSIT 2010. ACM Press (2010)

4. Esselaar, S., Stork, C., Ndiwalana, A., Een-Swarray, M.: ICT Usage and Its Impact on Profitability of SMEs in 13 African Countries (2008)
5. Faris, S.: Bookkeeping Made Easy...And Mobile With The Help Of Some Great New Apps. (2012), http://www.smallbiztechnology.com/archive/2012/03/bookkeeping-made-easy-and-mobile-with-the-help-of-some-great-new-apps.html/ (retrieved on June 13, 2012)
6. Gilbert, C.: Usability Evaluation. In: Soegaard, M., Dam, R.F. (eds.) Encyclopedia of Human-Computer Interaction. The Interaction Design Foundation, Aarhus (2012), http://www.interaction-design.org/encyclopedia/usability_evaluation.html
7. Giridher, T., et al.: Mobile Applications for Informal Economies. In: Proc. IMCSIT 2009. IEEE (2009)
8. Hellström, J.: The Innovative Use of Mobile Applications in East Africa, Sida Review 2010, vol. 12. Edita (2010)
9. Kwakwa, A.P.: Mobile Phone Usage by Micro and Small Scale Enterprises in Semi-Rural Ghana. International Review of Management and Marketing Journal 2(3) (2012)
10. Melchioly, R.S., Sæbø, Ø.: ICTs and Development: Nature of Mobile Phones usage for SMEs Economic Development - An Exploratory Study in Morogoro, Tanzania. Technical Commission 9 – Relationship between Computers and Society, March 22-23. IFIP, vol. 201, Workshop at Makerere University, Uganda (2010)
11. Maseko, N., Manyani, O.: Accounting practices of SMEs in Zimbabwe: An investigative study of record keeping for performance measurement (A case study of Bindura). Journal of Accounting and Taxation 3(8), 171–181 (2012)
12. Moore, et al.: Managing small business: An entrepreneurial emphasis, 14th edn. (2008)
13. Nielsen, J.: Usability 101: Introduction to Usability (2012), http://www.useit.com/alertbox/20030825.html (retrieved on October 1, 2012)
14. Schumacher, R.M., Lowry, S.Z.: NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records. National Institute of Standards and Technology (2010)
15. SMEs get productivity from mobile applications, http://www.computerweekly.com/news/2240088236/SMEs-get-productivity-from-mobile-applications (retrieved on September 25)
16. The East African Community (EAC), Study on the Promotion of Micro, Small and Medium Enterprises (MSMEs) in the East African Region, consolidated report by Ernst and Young. (2010), http://www.ugandainvest.go.ug (retrieved on June 30)
17. UNESCO. UNESCO Institute of Statistics Data Centre, regional literacy rates (2007), http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=y201 (retrieved on June 12, 2012)
18. Wamuyu, E., Maharaj: Usage of mobile technologies in Kenya. The African Journal of Information Systems (2011)
19. Webster Online Dictionary, http://www.websters-online-dictionary.org (retrieved on October 2, 2012)
20. White, G.: Designing Mobile services for the Non-literate communities. In: The Proceedings of the 2nd International Conference on M4D 2010, Kampala, Uganda (2010)
21. Xero Accounting Software, http://www.getapp.com/xero-application (retrieved on September 20)

Using Video Prototypes for Evaluating Design Concepts with Users: A Comparison to Usability Testing

Matthijs Zwinderman, Rinze Leenheer, Azadeh Shirzad, Nikolay Chupriyanov, Glenn Veugen, Biyong Zhang, and Panos Markopoulos

Eindhoven University of Technology, Department of Industrial Design,
Eindhoven, The Netherlands
{mzwinderman,r.a.leenheer,a.shirzad,n.chupriyanov,
gveugen,b.zhang}@gmail.com, P.Markopoulos@tue.nl

Abstract. We present an exploratory study that compared user feedback obtained from evaluating a mobile application versus a reverse engineered video prototype of this same application. The comparison included qualitative and quantitative data analysis. Questionnaire responses regarding user acceptance (UTAUT [8]) and the overall user experience (AttrakDiff [4]) showed no differences. Qualitative analysis of the comments by potential users yielded similar results. Some differences regarding comments pertaining to the fit of the application to its context of use and appreciation of hedonic qualities, warrant investigation in future research. Usability tests seem better suited to identify issues referring to the functionality and data accessed. Overall the results confirm the validity and usefulness of video prototyping, and can help us develop a better understanding of its strengths and weaknesses.

Keywords: Video prototyping, comparative usability study, experiment, user centered design methodology.

1 Introduction

The design of interactive systems benefits from early exposure of design concepts to users to obtain feedback and improve them iteratively. During early phases of the design process fast, and low-cost design representations are better suited than fully functional systems or realistic prototypes: this allows ideas to be tried out, changes to be made efficiently, and ideas that do not work well to be abandoned without much loss of effort. Video prototypes or video scenarios are such representations, showing on video envisioned scenarios of use.

Video as a medium offers several possibilities. It can draw attention to what users do with technology rather than technical workings, can be unencumbered by technological challenges that hamper implementation, and does not require technical expertise to be understood or created. Importantly the technique makes it just as easy to represent mobile interaction, ubiquitous computing, tangible and embodied interaction, etc. Shooting interaction in a particular location or social setting makes it easy to visualize context of use. The continuity of the medium can help explicate and explain

P. Kotzé et al. (Eds.): INTERACT 2013, Part II, LNCS 8118, pp. 774–781, 2013.

detailed interaction design issues [1] that are typically brushed over in more discrete representations such as text scenarios or storyboards.

Given the apparent usefulness and popularity of this prototyping technique the question arises whether the feedback it helps obtain from representative users provides valid insights. The video prototype inevitably represents a normative view of how the designer imagines that such a system should be used; it typically will show flawless execution of interaction tasks by expert users as the designer imagines will be possible. Further the viewer cannot report back on actual use of the system and on actual experiences, but on projections from a current context, and an imagined use of the system.

In this paper we compare feedback obtained from users who were shown a video prototype (from here on viewers) to that provided by test participants in a usability testing session.

2 Methodology

An experiment was designed to compare the feedback obtained from usability testing a widely available mobile application and a video prototype that was reverse engineered to represent the interaction with this application. The comparison focused on what interaction and user experience designers typically look for: overall appreciation, usability, perceived usefulness and acceptance (how likely it is that they will use the application), measures of the user experience, and indications of any potential improvements to the design.

A between subjects design was chosen; participants would either watch a video prototype of an actor carrying out some tasks, or carry out the same tasks in a usability test. Quantitative and qualitative data was collected and compared. The qualitative analysis included an exploratory phase of open coding where hypotheses were generated, followed by a closed coding (directed content analysis) to validate these hypotheses.

2.1 Materials

In order to make a useful comparison, we chose to evaluate a product that would be both novel to test participants and sufficiently complex. Novel, since users seeing the video-prototype should give feedback based on the video and not on previous experience. The interaction should not be trivial or familiar so that explaining it to viewers would indeed rely on the video prototype rather than a comment or prior knowledge. We chose as a test case to evaluate Google Goggles[1], a smartphone application. With this application a user can take pictures from objects and analyze and extract information from it. For instance, the application can recognize text and translate it or find the name of a painter based solely on a photograph of the painting. This application is still a novelty for the broad public, even though it is widely available in the iPhone and Android 'markets'. Also the means of interaction is rather unfamiliar to the broad public relying on the camera rather than entering information with buttons and touch.

[1] `http://www.google.com/mobile/goggles`

Fig. 1. Props used for the five tasks the participants had to perform

In this sense, an evaluation based on the video prototype is practically identical (as far as the users are concerned) with the situations in which video prototypes are evaluated as part of an actual design process.

2.2 Tasks and Procedure

The comparative evaluation covered the main application functions which were mapped to five distinct evaluation tasks:

- Translating the ingredient list of a recipe printed on paper.
- Translating the name of two milk cartons (half-skimmed and whole milk).
- Getting the title and artist name of a famous painting.
- Transferring the details of a business card to the phone`s address book.
- Getting the price of a camera online (for this task both a picture of the camera or a description card could be used).

The products used for the test are shown in Figure 1.

2.3 Participants

In total thirty individuals participated in the study (11 female, 19 male, mean age=22.34, std=2.94, min=18, max=28), who were recruited among students in the university. Participants were divided over two groups: a product group (P) and a video group (V). In the 'product' group participants performed the five tasks with the Goggles application which was pre-installed on an Android based HTC phone. In the video group they watched a video showing actors carrying out these tasks.

2.4 Measures

In both groups the participants then completed two questionnaires, the AttrakDiff [4] for evaluating user experience aspects and UTAUT [8] for evaluating user acceptance, and participated in a semi-structured interview.

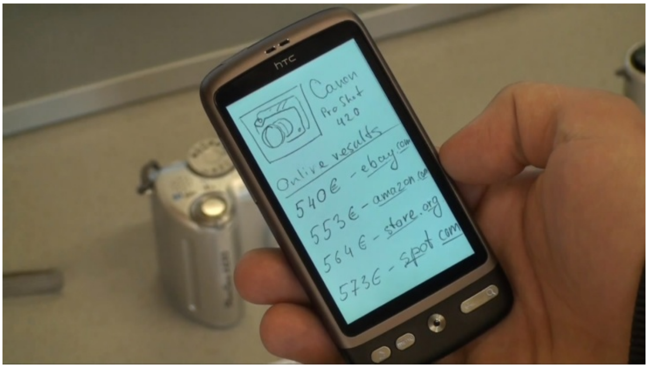


Fig. 2. Screenshot from the reverse-engineered video-prototype of Google Goggles™. The prototype combined an actual phone with drawn sketches of the user-interface. This increased the feeling that the application was not actually developed yet.

2.5 Video Prototype

The video-group saw a video where an actor performed the same tasks, each in a fitting context. The video-prototype[2] was a short movie lasting 2 minutes and 38 seconds. The video features one of the researchers acting out foreign exchange student using a Goggles application in a supermarket, an art-exhibition, and an electronics-store.

Instead of filming actual use of the Goggles application which would not be representative of video prototypes, a mockup was reverse-engineered to give the impression of the application still being in the conceptual phase of design. The prototype consisted in a website showing hand-drawn figures simulating the Goggles application and the results it returns during operation (see figure 2 for an impression). As the pages` figures were loaded on a website actors could simulate interaction: clicking on a drawn button in the video gives the impression of going from one screen to another. The video thus created can be described as a low-fidelity video-prototype; higher fidelity would mean that the more visually refined graphics would be used for filming which would be not very distinguishable on the video from the actual application. We note here that earlier studies suggest that high and low representations in video prototypes yield trigger similar feedback by viewers regarding user acceptance of the concept and also regarding how critical their comments are over the concept shown [2].

2.6 Analysis Method

Data from the first five participants in each group was analyzed qualitatively to identify potential patterns characterizing the two testing procedures. The comments made in the semi-structured interview from these sessions were annotated and transcribed on sticky notes. These were then clustered inductively according to common characteristics into an affinity diagram by a team of five researchers who continued structuring

[2] `http://www.vimeo.com/36969549`

the diagram and moving sticky-notes, creating, combining or removing groups until consensus was reached. Based on the resulting groups in the affinity diagram the following expectations were formulated:

E1. Participants who use the product give more comments on the interface.
E2. Viewers of the video prototype make more comments on the context of use
E3. Participants will provide a similar number of comments as to when and where they will use the application
E4. Viewers of the video prototype suggest more new features
E5. Participants who use the product suggest more improvements

The data from the first coding session was not used after this. The interviews from the remaining participants were transcribed in full and then chunked into 242 coherent comments (e.g., "it was very intuitive", or "at the moment I don`t see it as a useful tool, more as a fun application", or "I think the product tries to make life too easy"). These chunks were matched to the expectations by two observers working independently of each other. The inter-rater reliability for the raters was found to be Kappa = 0.60 ($p < .0.001$), 95% CI (0.508, 0.688). This is "moderate agreement" [5]. Chunks (on sticky notes) where the observers disagreed were discarded, as well as all chunks that could not be matched to any of the above expectations.

The data from the questionnaires was processed and analyzed statistically. A two independent sample non parametric test (Mann-Whitney U) was used to compare the scores of the two groups on the four factors of the AttrakDiff questionnaire. One participant in the video group indicated having had prior experience with Google Goggles™. This participant was excluded from the results. Another four participants did not fill in all the questions on the form and were therefore also excluded. This left thirteen participants in the "Product group" and twelve in the "Video group". Because the hypotheses were not directional, tests were "two-tailed".

3 Results

3.1 Qualitative Results

We summarize below the qualitative data (excluding the 5 first participants per group), providing also indications of the size of different clusters. Table 1 shows a brief summary of how the total numbers of comments produced were classified.

Only one comment could be tied to E1, which makes it impossible to draw any conclusions.

E2 was partially corroborated by the directed content analysis. There were 21 chunks from the product group and only 14 from the video group which fits our expectations, but many of these referred to the same issue. Participants in the test and viewers of the prototype identified a similar number of unique remarks regarding context (7 in the product group and 6 in the video group) and the nature of the comments they made about it was very similar as well.

Table 1. The number of comments that could be matched, after directed content analysis and after removing cases of disagreement between coders, for each expectation per group

Expectation	Product Group	Video Group
1	1	0
2	7	6
3	3	4
4	5	6
5	5	3

Both groups made almost the same amount of comments concerning the expected use of the application, which confirms our expectations (E3). The product group (3 comments) and the video group (4 comments). While the number of comments was similar, the video group`s comments related to future use are overall more positive than the ones made by the product group. This could reflect the difference between idealized operation by an expert user on the video versus actually attempting to operate the application first hand.

Contrary to our expectations (E4) the video group did not come up with more new features compared to the product group. The two groups produced a similar amount of suggestions (5 in product group, 6 in video group). In addition, the type of features that were suggested was similar (such as 'scanning buildings' or 'getting allergy information').

The slight difference between the numbers of suggested improvements by application group compared to the video group confirmed E5 (5 comments in the product group versus 3 in the video group). There is also a difference in the type of comments: the product group is more in-depth, they talk about usability issues in speed, capture results, text editing, button placement, etc. Comments from the video group are broader, such as "it should give other information as well, besides the painter".

3.2 Quantitative Results

Results showed no statistically significant difference between the groups on the factors Pragmatic Quality, Attractiveness, and HQ. As these are multi-dimensional constructs we, examine potential differences in the subscales they consist of.

A significant difference between the two groups was found on the factor "Hedonic Quality (HQ) Identity". HQ Identity consists of items such as 'professional', 'stylish' and 'presentable'. The sum of ranks for the `Product group' was significantly higher than the sum of the 'Video group' (15.96 vs. 9.79, $p<0.05$). This can be explained by the low visual refinement of the application as shown on the video.

HQ Stimulation consists of items like 'creative', 'captivating' and 'challenging'. The participants who used the actual application scored higher on this factor (one-tailed t-test, $p<0.05$). However, this difference was not hypothesized a priori, so it has to be confirmed in future studies. In hindsight, this is a result that should have been predicted. Stimulation seems like a quality one would mainly experience when using the application and not when only watching a video of someone else using it.

For the UTAUT four factors were examined: Performance expectancy, Effort expectancy, Attitude towards using technology and Behavioral intention to use the system. A two independent sample non parametric test (Mann-Whitney U) found no significant difference between the two groups for any of these factors.

4 Discussion

Video prototyping is a useful prototyping tool for exploring design ideas, and in this study it was shown that overall, using a video prototype for evaluation has led to practically the same insights as a user test with an actual product. This is quite a positive result, since making a video prototype is far cheaper and less time consuming than building a fully working prototype.

Comments concerning context appeared more pronounced in reaction to a video prototype than in actual use. Presumably this is because the specific video highlighted contextual aspects of use rather than filming usage out of context. Good video prototypes are likely to follow a similar practice using spaces and props to provide an idea of how users experience a product in their physical and social environment. However no strong claims can be made regarding the superiority of video prototyping as the unique issues identified were practically the same for the two groups.

Our results suggest that detailed design improvements are more likely to be identified by usability testing. Presumably this is because participants in the video-group were shown a film of a non-functional system where every scan and image based retrieval operation was shown to work immediately and impeccably. We should emphasize that while video prototyping is good for evaluating the overall design concept and direction, it is no substitute for testing a prototype with high refinement regarding functionality and data access. Still, in this case it did help identify numerous useful suggestions on improving the product. Given that it can be produced at a fraction of the cost, video prototyping is confirmed as a useful technique for early in the design process.

In comparison of the answers participants gave in the questionnaires regarding acceptance and the overall user experience there were also no significant differences found on all subscales (except for 'Hedonic Quality' for which further corroboration is needed by future research). This seems to be in agreement with earlier research [3], where variations on the refinement of the prototype shown on video did not impact the overall appreciation of the concept. The current study suggests that this extends to comparisons with actual product use and in reference also to feature suggestions, context of use, and expected use.

Related research [7] has found that there can be an interaction effect between different user groups (based on demographics, knowledge, etc.) and the prototyping medium with regards to the feedback obtained by users. It would be interesting to extend this comparison to other kinds of systems, and to different user groups. This study's user group was relatively familiar with the medium (smart phones); it would be interesting to check whether these results can be reproduced in cases where test-participants are very unfamiliar with the tested technology, requiring a lot more from their imagination and empathy to envision situations of use by watching a video.

Overall this study suggests that video prototypes help obtain feedback from users that is quite similar to that gathered when user testing the final product. Of course the study has examined only one application and its presentation as a video prototype; to generalize our conclusions one would have to reproduce these results for different applications and videos. Nevertheless, the fact though that no major differences are found does have face validity; it is exactly the intent of a prototype (be it on video or not) to capture what is essential from the design concept in a way that will solicit valid feedback by users. Detailed comparisons such as the one presented can inform us regarding the appropriateness of different prototyping media for different evaluation aims.

References

1. Bardram, J., Bossen, C., Lykke-Olesen, A., Nielsen, R., Madsen, K.H.: Virtual video prototyping of pervasive healthcare systems. In: Proceedings DIS 2002, pp. 167–177 (2002)
2. Bojic, M., Goulati, A., Szostak, D., Markopoulos, P.: On the effect of visual refinement upon user feedback in the context of video prototyping. In: Proceedings of the 29th ACM International Conference on Design of Communication, SIG DOC 2011, pp. 115–118. ACM (2011)
3. Diefenbach, S., Hassenzahl, M., Eckoldt, K., Laschke, M.: The impact of concept (re)presentation on users' evaluation and perception. In: Proceedings NordiCHI 2010, New York, NY, USA, pp. 631–634 (2010)
4. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. Mensch & Computer, 187–196 (2003)
5. Landis, J.R., Koch, G.G.: An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics, 363–374 (1977)
6. Mackay, W.E., Ratzer, A.V., Janecek, P.: Video artifacts for design: bridging the Gap between abstraction and detail. In: Proceedings DIS 2000. ACM, pp. 72–82 (2000)
7. Sellen, K.M., Massimi, M.A., Lottridge, D.M., Truong, K.N., Bittle, S.A.: The people-prototype problem: understanding the interaction between prototype format and user group. In: Proceedings CHI 2009, pp. 635–638. ACM, New York (2009)
8. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Quarterly, 425–478 (2003)

Author Index