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

LNCS 8117

Human-Computer Interaction – INTERACT 2013

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

1 Part I

ifip

Springer

Lecture Notes in Computer Science 8117

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 I

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 978-3-642-40482-5 e-ISBN 978-3-642-40483-2
DOI 10.1007/978-3-642-40483-2
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 I

Long and Short Papers

3D Navigation

3D Technologies - 3D Object Manipulation

Augmented Reality

Cognitive Workload

Cognitive Workload and Decision Support

Creating Effective 3D Displays

Cross-Cultural, Intercultural and Social Issues

Data Entry Mechanisms and Devices

Design and Evaluation

Design and Evaluation of Prototypes

Design to Support Creativity

Designing for Inclusiveness I

Designing for Inclusiveness II

Designing with- and for People with Special Needs

Display Manipulations

Diversity/ICT in Social Development

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

Table of Contents – Part III

Long and Short Papers (Continued)

Mobile Usage and Techniques

Mobile UX and Privacy Concerns

Model-Based User Interface Design

Multimodal User Interface Design

Multimodality, Cross-Platform Studies

Narratives in Design

Navigation Aids

Novel User Interfaces

Passwords: e-Authentication

Physical Ergonomics

Road Safety

Seniors and Usability

Social Behaviour, Collaboration and Presence

Social Collaborative Interaction

Social Media

Software Development

Table of Contents – Part IV

Long and Short Papers (Continued)

Supporting Physical Activity

Supporting Shared Activities

Sustainability

Tabletop Computing

Text Comprehensibility

Tracking Eyes and Head

Usability Evaluation and Technology Acceptance

User Preferences and Behaviour

User Requirements Capture and Analysis

UX in Work/Educational Contexts

Voice/Sound-Based Computing

Interactive Posters

Industry Tracks

Panels

Special Interests Groups

Tutorial

Workshops

Assessing the Impact of Automatic vs. Controlled Rotations on Spatial Transfer with a Joystick and a Walking Interface in VR

Florian Larrue[1,2], Hélène Sauzéon[2,1], Déborah Foloppe[3], Grégory Wallet[4], Jean-René Cazalets[5], Christian Gross[6], Martin Hachet[1], and Bernard N'Kaoua[1,2]

[1] INRIA, F-33400 Talence, France
[2] University of Bordeaux Victor Segalen - EA 4136, Handicap & Système Nerveux 33076 Bordeaux Cedex, France
[3] LUNAM Université - Université d'Angers - LPPL - (UPRES EA 4638), France
[4] Aix-Marseille Université, CNRS, ISM UMR 7287, 13288, Marseille Cedex 09, France
[5] CNRS UMR 5287 – INCIA - Institut de Neurosciences Cognitives et Intégratives d'Aquitaine 33076 Bordeaux Cedex, France
[6] Institut des Maladies Neurodégénératives - University of Bordeaux Victor Segalen CNRS UMR 5293 - 33076 Bordeaux Cedex, France
{flo.larrue,gregwallet,foloppe.deborah}@gmail.com,
{helene.sauzeon,christian.gross,jean-rene.cazalets,
bernard.nkaoua}@u-bordeaux2.fr,
martin.hachet@inria.fr

Abstract. We present a user study assessing spatial transfer in a 3D navigation task, with two different motor activities: a minimal (joystick) and an extensive motor activity (walking Interface), with rotations of the viewpoint either controlled by the user, or automatically managed by the system. The task consisted in learning a virtual path of a 3D model of a real city, with either one of these four conditions: *Joystick* / *Treadmill* Vs *Manual Rotation* / *Automatic Rotation*. We assessed spatial knowledge with six spatial restitution tasks. To assess the interfaces used, we analyzed also the interaction data acquired during the learning path. Our results show that the direct control of rotations has different effects, depending on the motor activity required by the input modality. The quality of spatial representation increases with the *Treadmill* when rotations are enabled. With the *Joystick*, controlling the rotations affect spatial representations. We discuss our findings in terms of cognitive, sensorimotor processes and human computer interaction issues.

Keywords: Interfaces, Navigation, Virtual Reality, Spatial Cognition, Joystick, Treadmill, Rotation, Body-based Information, Vestibular Information, Human Machine Interaction, Human Factors, User Study, Motor Activity.

1 Introduction

Today, Virtual Reality (VR) enables the simulation of dynamic, three-dimensional, multimodal environments. Moreover, this technology allows immerging users in

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 1–18, 2013.

different simulations close to real situations, where users can interact with the virtual environment (VE) and have a motor and a cognitive activity. VR also permits the access to various data (e.g. completion time and precision), hard to reach in a real environment. Thanks to these advantages, VR is well suited to create therapeutic applications for patients with spatial disabilities diseases. An important question to explore for such application is to evaluate how spatial information is impacted when transferred from virtual to real environments. Several studies already found great results with disabled people about the question of spatial transfer [25][26][27]. In this work, the authors agree that various factors can enhance this spatial transfer. One question not yet resolved concerns the exploration mode used to navigate in a VE [22]. Indeed, sometimes authors have shown great spatial acquisition with an active exploration mode (i.e., user had a sensorimotor interaction with the VE) compared to a passive mode exploration [23] (i.e., user had no interaction with the VE) [21][1][2][8], but others did not [22][23][24]. Moreover, these studies were generally based on a joystick or a mouse/keyboard interface. Nevertheless, different authors demonstrated, that a walking activity allows to optimize the acquisition of spatial knowledge [12][13], but only two concerns the spatial transfer [3][19]. Thus, we first purpose to assess the impact of two motor activities on spatial transfer with two *Input Devices*: a walking interface (using a *Treadmill*) and a *Joystick*. Moreover, the impact of rotation movements during a navigational activity, on spatial transfer, is not yet clear. In a second step, we therefore investigated the role of *Rotation* (*Automatic*-i.e., controlled by the computer- Vs *Controlled*- i.e., managed by the user) with the two *Input Devices* presented above. So, on a spatial transfer task, we used i) a *Treadmill with Controlled Rotation* ii) a *Treadmill with Automatic Rotation* iii) a *Joystick with Automatic Rotation* iv) a *Joystick with Controlled Rotation*. We used six tasks to assess spatial knowledge. To our knowledge, this study is the first one describing impact of translational and directional movements according to different motor activities in a VE, on spatial transfer.

1.1 Spatial Cognition (cognitive and sensorimotor processes)

Spatial cognition refers to cognitive and motor processes requiring to acquire, to store and to restitute spatial knowledge. Processes involved in spatial cognition are necessary for many daily life situations, such as shopping in supermarkets (e.g., finding a product in a section) and driving, and are often affected by neurological diseases (e.g., Alzheimer), brain trauma, etc. For Montello [28], spatial cognition is divided in two components: 1) the motor component, composed of all sensorimotor information acquired during a displacement, with visual, kinesthesic and vestibular information, informing on the position and the orientation of the head/body in an environment; 2) the cognitive component corresponding to the processes used to acquire, store and restitute spatial knowledge. One of the most known spatial acquisition model is the Landmark-Route-Survey model of Siegel and White [9]. For this model, spatial knowledge acquisition of new environments consists of three stages. Firstly, spatial cognition is based on the acquisition of several landmarks in the environment. Secondly, the participant links the landmarks and learns the routes between them. At this level, s/he is able to build a mental representation of a route from a departure point to an arrival point using the various landmarks. These first two levels correspond to egocentric-type representations (*i.e.*, the body serves as a reference). Finally, the participant develops survey knowledge. S/he builds a topographical representation of the environment, including all the associated spatial information (*i.e.,* landmarks and routes), making it possible to infer a representation of the entire environment, making it possible to contemplate shortcuts. At this final level of knowledge, the representation is similar to a "plane view" and is also known as "survey-type" knowledge: the mental representation of the environment is complete and allocentric (i.e., an external point serves as a reference). These three acquisition stages need not follow a strict order but may be obtained in a parallel process [29]. Concerning the sensorimotor component, body-based information required during a navigational activity can be divided in three types of information [14]: 1) the optic flow, consisting of all visual input used to detect forms, textures, semantic landmarks, movements of objects, etc. always in relation with body position, 2) the vestibular system provides translational (acceleration/deceleration of the head and body) as well as rotational information (rotation of the head and body), and 3) the kinesthetic information, which informs about the perception of our members according to our body. In real environments, different authors admitted that vestibular information is important to the creation of egocentric representations (perception of distances, angles or route knowledge) or to store a direction linked to an egocentric point of view [11] [12], while allocentric representations would be more sensible to visual information.

1.2 Spatial Cognition, Interfaces and Rotational Movements in VR

Literature concerning walking activity in VR is not very consistent. However, most of the studies agree that the extent of body-based information provided by a treadmill locomotion interface (compared to a joystick) was considered largely favorable for spatial learning in a VE [11][13][14], due to the improvement in egocentric [11] and allocentric spatial representation [14], as well as navigational measurements [19]. Recently, Ruddle et al. [14] assessed the role of both translational and rotational vestibular information on survey knowledge, using different locomotion interfaces (translational displacements with walking or treadmill Vs. no translational displacements with joystick), sometimes with the possibility of really turning the head (i.e., rotational vestibular condition or not) during rotational movement. Performances revealed an improvement of survey knowledge with a walking activity, but little effect about rotational vestibular information was observed. For Waller et al.[11], the low level of body-based information provided by the design of the locomotion interfaces of the desktop VEs (i.e., keyboards, mouse or joysticks) do not allow the increase of spatial knowledge acquisition. For some authors [16], the manipulation of translational and rotational movements with a joystick demands a strong motor and cognitive attentional levels, which could interfere on spatial learning acquisition. Even if authors promote a walking interface to optimize spatial knowledge, it seems it is still possible to navigate in a VE with a joystick [1][2], without vestibular information[3]. However, in all the studies presented we did not find research, whatever the interfaces used (and body-based information provided), where the possibility to perform rotation of the user's viewpoint was disabled and managed automatically by the system.

1.3 Spatial Cognition and Spatial Transfer from Virtual to Real Environments

One important challenge of VR is to detect the factors promoting knowledge acquisition in VR to improve daily life activities in the real life. Different authors already showed great spatial transfers with "normal"[1][2][3] or patients with disabilities people [27]. Numerous factors like visual fidelity [2], retention delay [1], game experience [16] increase this transfer. However, concerning the motor activity of the interfaces used, most parts of studies used a passive exploration mode, or a joystick/mouse/keyboard interface (active exploration mode) to navigate in the VE. And the results point out sometimes great performances for the active exploration mode, [21][1][2][8], and others did not [22][23][24]. Moreover, these interfaces don't provide vestibular information, known to improve spatial acquisition. We found only two studies which used a walking interface to study spatial transfer. The first [19] revealed a better spatial transfer with a walking interface compared to a joystick, concluding on the importance of vestibular information. The second study [3] assessed the impact of the motor activity in a spatial transfer task. They compared a Brain Computer Interface (allowing to navigate in a VE with no motor activity), a treadmill interface (enabling vestibular information), and a learning path in the real environment. The results revealed similar performances, whatever the learning conditions, indicating that the cognitive processes are more essential than a motor activity. Results revealed also that a walking activity (and vestibular information) enables spatial knowledge transfer similar to the real life.

2 Method

VR was assessed as a spatial learning medium using a spatial learning paradigm that involved acquiring a path in its virtual replica [1][2][3]. In our experiment, the acquisition path in the VE was assessed according to four conditions: (1) *Treadmill with Controlled (head) Rotation* (optic flow, rotational and translational vestibular information); (2) *Treadmill with Automatic Rotation* (optic flow, translational vestibular information and no rotational vestibular information); (3) *Joystick with Controlled (hand) Rotation* (optic flow); (4) *Joystick with Automatic Rotation* (optic flow). Following VR-based path acquisition, the participants completed six tasks to assessing their spatial knowledge and spatial transfer.

2.1 Setup

The environment.
The real environment was a 9km2 area. The VE was a 3D scale model of the real environment, with realistic and visual stimuli. The scale of the real environment was faithfully reproduced (measurements of houses, streets, etc.) and photos of several

building facades were applied to geometric surfaces in the VE. Significant local and global landmarks (e.g., signposts, signs, and urban furniture) and urban sounds were included in the VE (see Figure 1). VE was laboratory-developed using Virtools Dev 3.5™. Irrespective of the interfaces conditions, the itinerary was presented to participants on the basis of an egocentric frame of reference, at head height. It was characterized by an irregular closed loop, 780 m in length, with thirteen crossroads and eleven directional changes.

Fig. 1. Screenshots our real (left) and our virtual environment (right)

Material.
The material used in the darkened laboratory room was a DELL Precision M6300 laptop computer (RAM: 3GHz; processor: Intel Core 2 Duo T9500 2,60 Ghz) with an Nvidia Quadro FX 1600M graphics card (256Mo), a 2 x 1.88 meter screen, a projector (Optoma/ThemeScene from Texas Instrument) with rear projection. The participants were placed two meters from the display screen.

2.2 Interface Modeling

The Treadmill Input Device.
The two *Treadmill* conditions (with *Automatic* and *Controlled Rotation*) included an HP COSCOM programmable (speed, declination and acceleration) treadmill with serial Cable Ownership coupled to a Software Development Kit and an MS-EZ1 sonar telemeter. This interface enabled participants to modify the VE's visual display in real time to match his/her walking speed, with a maximum of 6 km/h. Acceleration and deceleration were applied by means of a Sonar MS-EZ1 telemeter that monitored the participant's displacements on the treadmill. The treadmill surface was divided into three parts: one for accelerating (the front of the treadmill), one for walking normally (the middle of the treadmill), and one for decelerating (the back of the treadmill). No acceleration or deceleration information was sent to the treadmill when the participant was in the walk zone. In contrast, when the participant walked into the acceleration or deceleration zone, the sonar detected length changes in the participant's position, and instructed the computer to accelerate or decelerate until the participant returned to the walk zone. Finally, the participant remaining in the deceleration zone for a prolonged period induced a stop in the environment. In the two

Treadmill conditions, participants were able to walk, accelerate, decelerate, and stop in the VE, thus receiving physical input including optic flow, as well as kinesthetic and translational vestibular information.

For the condition *Treadmill with Controlled Rotation*, the participant walked on the treadmill and was informed that his/her point of view in the VE would be controlled by head rotation (providing rotational vestibular information). Head rotation movements were captured in real time by motion capture (12 OPTITRACK video-cameras, Motion point™). When a participant turned his/her head, the system updated the visual optic flow at a rate correlated with the head movement rotation angle (the greater the rotation angle, the faster the modification in rotational optic flow). Thus, this condition enabled translational and rotational vestibular information.

The *Treadmill condition with Automatic Rotation* was the same as the condition *Controlled Rotation*: the participant controlled its translational displacement but, on a pre-determined path; directions changes were automatically managed by the system at each intersection. The interface did not allow any rotational movement control, enabling only translational vestibular information.

The Joystick Input Device.
In both *Joystick* conditions (with *Controlled* or *Automatic Rotation*), displacement was controlled by a Saitek ™ X52 Flight System. Forward speed, ranging from 0 to 6 km/h, was proportional to the pressure on the device, which was also used to control translational movement. Consequently, the *Joystick* conditions differed from the *Treadmill* conditions in providing optic flow, but no vestibular information.

The *Joystick with Controlled Rotation* condition added horizontal joystick movements, coupled to changes in rotational optic flow to simulate turning in the VE to mimic direction changes during walking. Turning speed was proportional to the magnitude of horizontal joystick movement, similar to natural head movement.

For the *Joystick with Automatic Rotation*, participants were informed that rotational movement was not available; turning at intersections would be automatic.

2.3 Procedure

Each participant completed a three-phase procedure: (1) spatial ability tests and orientation questionnaire, to assess the participant's characteristics (see below); (2) learning phase: training interface and the route-learning task under one of the four conditions; (3) restitution phase, consisting of six spatial knowledge-based tasks.

Spatial Ability Tests, Orientation Questionnaire
The GZ5 test [4] was used to measure spatial translation ability of participants; the Mental Rotation Test (MRT) [5] to measure spatial visualization and mental rotation abilities; and the Corsi's block-tapping test [6], was used to assess the visual-spatial memory span. Three self-administrated questionnaires including seven questions each (for which responses were given on a 7-point scale) were filled in by the participant. One questionnaire assessed general navigational abilities and spatial orientation in everyday life, a second evaluated the ability to take shortcuts, and the third was dedicated to the ability to use maps.

Learning Phase

Interface Training.

Before VR exposition, each participant participated to a training phase in a different environment, to get used to interacting with one of the four interfaces that he/she will use. The training phase was finished when the participant was able to use the interface in another VE.

Learning path in the VE.

For the two conditions with *Controlled Rotation*, participants walked at their own speed and managed their directions in the VE. The directions at each intersection were indicated verbally by an experimenter situated behind the participant. For the two conditions with *Automatic Rotation*, participants mastered their speed with the *Joystick* or the *Treadmill*, but were not able to perform rotations; they were automatically managed at each intersection by the computer. Moreover, a path learning software was developed to analyze the participant's position, time, speed, collisions and interactions during the learning path. In addition, after VR exposure, the participants completed a simplified simulator sickness questionnaire (SSQ) [7] to measure the negative side effects of being immersed in graphically-rendered virtual worlds, and a questionnaire about the ergonomic of the interface used and the participant's habits.

Restitution phase.

Six tasks were performed by each participant, with a counterbalanced order.

Egocentric photograph classification task: twelve real photographs of intersections, in a random order, were presented to the participants. Participants were required to arrange the photographs in a chronological order along the path they had learned. The time limit for this task was ten minutes. The results were scored as follows: one point for a photo in the correct position, 0.5 point for each photo in a correct sequence, but not correctly placed along the path (e.g., positioning photos 4-5-6 in the right order but not placing them correctly in the overall sequence earned 1.5points). This paper-pencil task assessed the participants' ability to recall landmarks and route knowledge within an egocentric framework ([1][2][3]).

Egocentric distance estimation task: Each participant was asked to give a verbal estimate of the VR walked distance (in meters) and the figure was noted by the experimenter. This task quantified the participants' knowledge of the distances walked between the starting and ending points, which is known to be accurate when participants have acquired well-developed route knowledge [8].

Egocentric directional estimation task: This task was computer-based and consisted of presenting a series of twelve real photographs of intersections, taken from an egocentric viewpoint, in random order. Each photograph was displayed at the top of the screen, above an 8-point compass. The participant had to select the compass direction in which they were facing on the learned path when the photograph was taken. We assessed the percentage of errors and the angular error was averaged. Directional estimates are expected to be accurate when participants have acquired well-developed route knowledge [9].

Allocentric sketch-mapping task: Participants were required to draw a freehand sketch of the visualized route. The time limit for this task was ten minutes. One point was scored for each correct change of direction. This paper-pencil task is known to measure survey knowledge [1][2][3].

Allocentric point starting estimation task: This computer-based task consisted of presenting a series of twelve real photographs of intersections, taken from a walker's point of view, in random order. Each photograph was displayed at the top of the screen, above a 8-point compass and the participant was instructed to select the compass direction of the starting point of the learned path. We assessed the percentage errors and the mean angular errors. These direction estimates are expected to be accurate when participants have memorized a well-developed, map-like representation of the environment [10]. This task measures survey knowledge.

Real wayfinding task: This task consisted of reproducing the learned path in the real environment; this task measures the spatial transfer of participants. During this restitution task, position and time data were acquired using a Magellan™ GPS CrossOver, and a video was recorded using an AIPTEK™ DV8900 camera mounted on a bicycle helmet worn by the participant. Direction errors were calculated and expressed in percentages. When a participant made a mistake, s/he was stopped and invited to turn in the right direction. This wayfinding task is based on the use of landmarks, as well as route and survey knowledge [1][2][3], and may be considered as a naturalistic assessment of navigational abilities based. In addition, the path learning software enabled to analyze the participant's position and time data in the real environment.

Participants.

72 volunteer students participated in this experiment (36 men and 36 women). Students were randomly divided in one of the four learning conditions: 18 students were assigned to the *Treadmill with Controlled Rotation* condition, 18 to the *Treadmill with Automatic Rotation* condition, 18 in the condition *Joystick with Controlled Rotation*, and 18 in the *Joystick with Automatic Rotation* condition. All the participants had normal or corrected-to- normal vision and were native French speakers, right-handed, and had at least an A-Level or equivalent degree. Their ages were ranged from 18 to 30 years. We controlled video game experience of participants: each learning condition were composed of half gamers (who played a minimum of three hours by week during more than one year), and the other half of non video game players (who never played regularly to video games, and who were not old video gamers). The four composed learning conditions were balanced for gender and the video-gamer distribution ($\chi 2$ procedure $p>.05$). In addition, there was no significant difference in spatial abilities among the four groups, as assessed with the GZ-5 test, the Mental Rotation Test (MRT) and the Corsi's block-tapping test (respectively, $p>0.180$ $p>0.640$; $p>0.200$). No differences were found for the orientation questionnaire ($p>0.800$), neither for shortcuts questionnaire ($p>0.600$), or the map questionnaire ($p>0.800$).

3 Results

We used a two-way ANOVA analysis [2 (*Input Devices*: *Treadmill* Vs *Joystick*) x 2 (*Rotation*: *Controlled* Vs *Automatic*)], with between-subject measures for each factor. Bravais-Pearson test was used to assess correlations.

3.1 Learning Phase

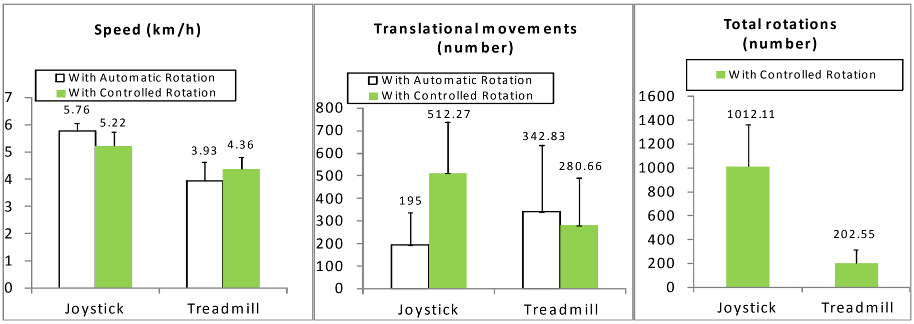


Fig. 2. Learning data according to the Input Devices (Joystick Vs Treadmill) and the Rotation (Automatic Vs Controlled)

For *the Speed during learning phase*, a significant effect of the *Input Devices* was found [F(1,68)=114.53; p<0.0001; η^2=0.63],with higher speed during the learning phase with the *Joystick* compared to the *Treadmill*. In addition, the factor “*Rotation*” had no significant effect (p>0.900), but the two-way interaction ("*Input Devices* x *Rotation*") was significant [F(1,68)=13.36; p<0.001; η^2=0.16]. With the *Treadmill*, speed learning was faster for with *Controlled Rotation* compared to the condition with *Automatic Rotation*. With the *Joystick*, the results are inversed; speed restitution was faster in *Automatic Rotation* condition compared to the *Controlled Rotation* condition.

Concerning the *total translational movements* (i.e, the number of accelerations/decelerations demands) during the learning path, the ANOVA analyses revealed a significant difference concerning the *Rotation* factor [F(1,68)=5.8; p<0.02; η^2=0.08]. The total number of translational movements was highest in *Controlled Rotation* condition. An interaction “*Input Devices* x *Rotation*” was found [F(1,68)=12.84; p<0.0001; η^2=0.16]. With the *Joystick*, there were more translational movements with *Controlled Rotation* compared to the condition with *Automatic Rotation*. With the *Treadmill*, the results were inversed: we found more translational movements with the *Automatic Rotation* compared to the *Controlled Rotation*.

Concerning the *number of rotations* performed by the participant, they were summed only for the two conditions with *Controlled Rotation*. We used an unpaired two-tailed Student's t-test (dof = 34). We found a significant difference (t(9.27); p<0.0001; η^2=0.72): more rotations were performed with the *Joystick*.

For the questionnaire about *the ergonomics of the interface used*, a question concerns the difficulties to perform rotations. Statistical analyses revealed a *Rotation* effect [$F(1,68)=7.60$; $p<0.01$; $\eta^2=0.10$]. Participants felt logically more difficulties to control their rotations in the condition *With Automatic Rotation*. An interaction *"Input Devices* x *Rotation"* was likewise found [$F(1,68)=4.60$; $p<0.05$; $\eta^2=0.06$], revealing that it is only in the *Treadmill with Automatic Rotation* condition participants revealed rotational difficulties. In the *Joystick* conditions, the results were similar, whatever the possibility or not to perform rotations.

3.2 Spatial Restitution Tasks

For the *Egocentric Photograph classification task*, the ANOVA revealed no significant effect (*Input Devices*, $p>0.600$; *Rotation*, $p>0.700$; and "*Input Devices* x *Rotation*", $p>0.400$). Performances were close in all four VR learning conditions.

Concerning the *Egocentric Distance Estimation Task*, the ANOVA revealed a significant effect for each factor [*Input Devices* effect, $F(1,68)=4,81$; $p<0.05$; $\eta^2=0.07$; *Rotation* effect, $F(1,68)=12,27$; $p<0.001$; $\eta^2=0.15$], with the *Joystick* conditions overestimating distances compared to the *Treadmill* conditions, and the groups with *Controlled Rotation* overestimating distances compared to *Automatic Rotation*. The two-way interaction effect was significant [$F(1,68)=4,44$; $p<0.05$; $\eta^2=0.06$]. The distances were only overestimated in the *Joystick with Controlled Rotation* condition compared to the other VR conditions.

For the *Egocentric Directional estimation task*, the ANOVA for mean angular error revealed that the two effects taken separately were not significant (*Input Devices* effect, $p>0.800$; *Rotation* effect, $p>0.800$), but the two-way interaction was significant [$F(1,68)= 3.99$; $p<0.05$; $\eta^2=0.06$]. With the *Joystick*, egocentric estimations were more accurate with *Automatic Rotation* compared to the *Controlled Rotation* condition, while for the *Treadmill*, performances were better in *Controlled Rotation* compared to the *Automatic Rotation* condition. It is to note that the results of the *Joystick with Controlled Rotation* and the *Treadmill with Automatic Rotation* conditions are very close. The results for the *Joystick with Automatic Rotation* and for the *Treadmill with Controlled Rotation* are also very close.

For the *Allocentric Sketch mapping task*, the ANOVA did not reveal any significant effects for the *Input Devices* or *Rotation* factors ($p>0.800$; $p>0.300$; two-way interaction, $p>0.100$). The performances did not reveal any differences among the four VR learning conditions.

Concerning the *Allocentric starting point estimation task*, the only significant *Input Devices* effect [$F(1,68)=4,38$; $p<0.05$; $\eta^2=0.06$] revealed by ANOVA was that the *Joystick* condition resulted in poorer performances than the *Treadmill* condition. No other effects were significant (*Rotation*, $p> 0.200$; two-way interaction, $p>0.800$).

For the *Wayfinding task (transfer task)*, two data (speed restitution and percentage of direction errors) were collected. For the speed restitution, the ANOVA results revealed a significant effect for *Rotation* [$F(1,68)=4,22$; $p<0.05$; $\eta^2=0.06$], i.e., the group *with Controlled Rotation* performed better than the one *with Automatic Rotation*. No other difference was found (*Input Devices* effect, $p>0.800$; two way-interaction, $p>0.900$). For the direction error measurements, the ANOVA results revealed a significant two-way interaction [$F(1,68)=4.00$; $p<0.05$; $\eta^2=0.06$].

Analysis revealed that for the *Treadmill* condition, performances were better *with Controlled Rotation* compared to the condition *with Automatic Rotation*. With the *Joystick*, the performances were more accurate *with Automatic Rotation* than with the *Controlled Rotation* condition. The best performances were performed with the *Treadmill* with and the *Controlled Rotation*. Other effects were not significant (*Input Devices*, p>0.800; *Rotation*, p>0.300).

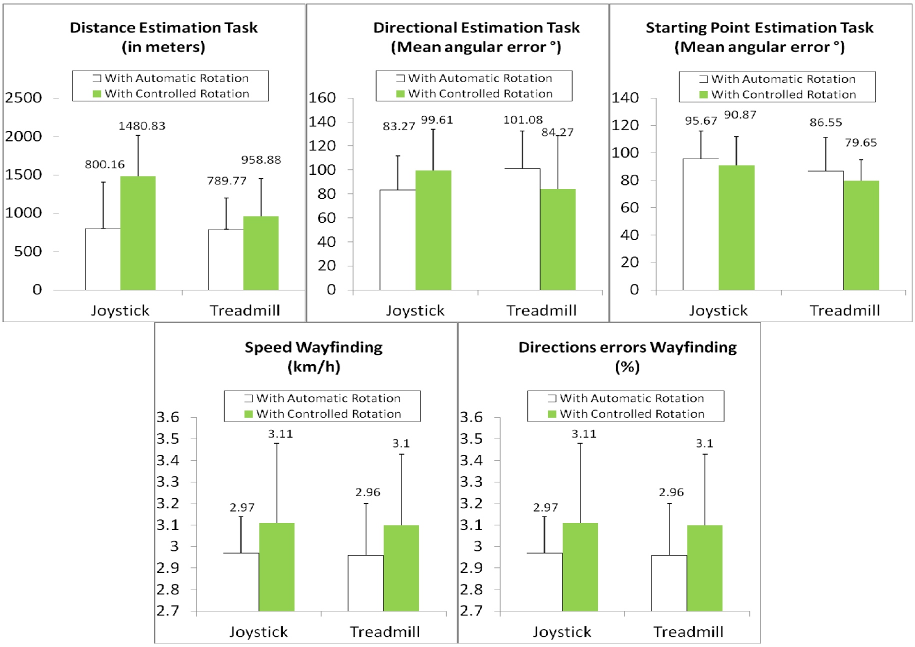


Fig. 3. Significant results for our spatial restitution tasks (Input Devices Vs. Rotation)

3.3 Correlations

We present only the principal correlations between data learning (translational movements, performed rotations), paper pencils tasks, the three orientations questionnaires, and the six spatial restitution tasks. No correlations were found about *spatial pencil papers tasks* and spatial restitution tasks. Concerning the *orientation questionnaires*, we found a negative correlation between the questionnaire assessing the abilities to use maps and the starting point estimation task (p=0.02, r=-0.37). No correlations were found concerning the other spatial restitution tasks. *For the data acquired during the VE learning path*, a positive correlation was found between the time to finish the learning path and the sketch mapping task (p=0.04, r=0.34). No correlations were found between rotations, translational movements and spatial restitution tasks.

4 Discussion

To recall, the goal of this study is to understand the impact on spatial transfer, of two *Input Devices* with two different levels of motor activities (the manipulation of a *Joystick* Vs a *Treadmill*) and the possibility or not to control rotations. In manipulating these two factors, we are forced to manipulate body-based information. More precisely, whatever the type of *Rotation* (i.e., *Automatic* Vs *Controlled*), the *Joystick* enables visual information, but no vestibular information (no head displacement). The only difference was, in the condition *Joystick with Controlled Rotation*, participant was able to master their translational displacements and their directions to explore freely the VE, while in the *Joystick with Automatic Rotation* condition, participant followed a predetermined path, and was able to control only their speed displacement; the rotations being controlled by computer at each intersection. *The Treadmill with Controlled Rotation* condition enabled to perform rotations (rotational head movements) during the learning path, providing translational and rotational vestibular information. In the *Treadmill with Automatic Rotation* condition, head rotations were blocked and translational vestibular information were activated. As provided in *Joystick with Automatic Rotation*, participant was able to control his/her displacement speed, and directions changes at each intersection were managed by the computer. We present our results according to the egocentric, allocentric and transfer tasks used.

4.1 Egocentric Tasks

For the *Photograph classification task,* whatever the *Rotation* or the *Input Devices* factors used, no statistical differences were found for this task. To recall, this task consisted in ordering chronologically twelve photos of the real environment, with an egocentric point of view. So, the motor activity and the possibility or not to perform rotations of our four interfaces seems to have little impact on this task. These results are in accordance with literature. For example, Wallet et al.[2] found, on the same type of task, that visual fidelity of a VE was more important than the interface used. We may wonder that different body-based information and attentional levels provided by our four interfaces had a little impact on tasks that do not require the recall of an action. It could mean also that the visual fidelity of our VE was perceived in the same way by the participants, whatever the interfaces used. If some differences appear they cannot be explained by this factor.

Concerning the *Egocentric Distance Estimation Task (*which consisted in evaluating the total distance travelled during the learning path phase), the results showed a significant difference for *Input Devices*, in favor of the *Treadmill* (vestibular information present) compared to the *Joystick* (i.e., no vestibular information because no head movements). These results are in accordance with different authors [11][12] who demonstrated that vestibular information is important to correctly estimate distances. An effect *Rotation* was also found, showing a distance overestimation for the conditions *with Controlled Rotation* compared to the conditions *with Automatic Rotation*. Statistical analyses revealed an interaction "*Input Devices* x *Rotation*". The distance estimates were very close for the *Treadmill with Controlled* or *Automatic Rotation* conditions, and the *Joystick with Automatic Rotation*. An overestimation for the *Joystick with Controlled Rotation* condition was found, explaining the *Rotation* effect described above. Finally, the *Rotation* (and rotational vestibular information), had no impact for the *Treadmill*. These results are coherent with [14] where the importance of the translational vestibular information on distance estimation is confirmed, and where no effect of rotational vestibular information was found. In contrast, in the two *Joystick* conditions (only visual information provided), we can see an overestimation only with the *Controlled Rotation* condition. For several authors, visual information would be sufficient to estimate distance [15], explaining maybe the results with *Automatic Rotation*. These results are new and difficult to explain. This may be due to the fact that two directions and the visuomotor coordination requested could interfere on visual and cognitive processes. Maybe the visuomotor coordination of the joystick was higher for gamers(compared to no gamers). It would be interesting to add a condition comparing video game experience to improve our conclusions. It is important to note that in a walking activity, the rotation seems not to be important for distance estimation. But, with a *Joystick*, the *Controlled Rotation* affects the distance estimation.

For the *Egocentric Directional Estimation task* (which consisted for participant to indicate the direction he took, according to real photographs of intersections), results showed an interaction "*Input Devices* x *Rotation*". With the *Joystick*, the condition with *Automatic Rotation* gave the best performances, while with the *Treadmill*, contrary to the previous task, the best performances were with the *Controlled Rotation* condition (and rotational vestibular information). Finally, according to the motor activity of the interface used, the *Rotation* factor had a different impact. Our results corroborate the results found by other authors [12] [13] showing that 1) vestibular information improves egocentric representations 2) rotational vestibular information, rotational head movements allow increasing egocentric and perception-action representations. However, for the *Joystick*, the *Controlled Rotation* decreases once again egocentric perceptions. Moreover, statistical analyses showed that in the *Joystick* conditions, translational movements strongly increased with the *Controlled Rotation* (compared to the *Automatic Rotation* condition). For the two *Treadmill* conditions, the number of translational movements was quite similar whatever the type of *Rotation* used. When we compared the number of rotations (*Joystick and Treadmill with Controlled Rotation* conditions), statistical analyses revealed almost five times more interactions for the *Joystick* than for the *Treadmill*. These results concerning the number of interactions seem to support our assumptions about the visuomotor difficulties to control two directions with the joystick. Moreover, to the question of the difficulties to perform rotations, statistical analyses showed a significant effect of *Rotation*, where the group with *Automatic Rotation* had more difficulties to improve rotations compared to the *Controlled Rotation* group. These results seem to be logical because in the *Automatic Rotation* condition, participants were not able to control their rotations. An interaction "*Input Devices* x *Rotation*" also appeared. Surprisingly, this result showed that the rotation difficulties felt by participants concerned only the condition *Treadmill with Automatic Rotation*. In other words, when the interaction is natural, as a walking activity, the rotation seems to be necessary for a 3D navigational task. On the other hands, with the *Joystick*, the results to this question were very close, *with Automatic or Controlled Rotation*. Thus, participants did not felt the need to improve their rotations with the two-*Joystick* conditions. With a *Joystick*, it seems that the *Controlled Rotation* is not always necessary, confirming the great participants performances in the *Joystick with Automatic Rotation* condition. These results can be interpreted in different ways: 1) a walking activity is more natural than a joystick, explaining the similar number of interactions *With Automatic* or *Controlled Rotation*. But the *Controlled Rotation* and rotational vestibular information improves the egocentric representations 2) it seems to indicate that the two directions manipulated with the joystick is difficult. Maybe the attentional levels of participants are divided between the control of the joystick and the visual perception of the VE. Participants could have visuomotor coordination hand difficulties [16]. Moreover, unlike with the *Treadmill*, participants did not interact in the same manner with the *Joystick* and the *Automatic* or the *Controlled Rotation*. 3) Maybe the *Controlled Rotation* of the *Joystick* took in account to many rotational movements of the hand. Adding a condition where the *Controlled Rotation* of the *Joystick* took into consideration fewer rotations should give new information about our results. It is to note that no correlations were found between translational movements/rotations and our six spatial restitution tasks.

To summarize:

- The distance perception is optimized in a walking activity (whatever the *Rotation* factor). A *joystick with Automatic Rotation* permits also to assess correctly distances (only with visual information).
- The use of a *Joystick* and a *Controlled Rotation* decreases egocentric representations. With a joystick, only the control of translational movements seems to be sufficient to acquire egocentric spatial knowledge, similarly to a walking interaction close to the real life (i.e., *Treadmill with Controlled Rotation*).
- In a walking activity, the rotation, rotational head movements optimize the creation of egocentric representations. The absence of rotational vestibular information with a *Treadmill* affects negatively spatial egocentric representations.

4.2 Allocentric Tasks

Concerning the *allocentric sketch mapping task*, no significant differences were found. We can still observe a positive correlation between this task and the time to learn the path. The higher the completion time, the better the performances. These results support the L-R-S model of Siegel and White [9], who admitted survey representations to be improved with a long and repeated exploration of the VE.

For the *Allocentric Starting-point estimation task*, the results indicated better performances with the *Treadmill* compared to the *Joystick*, whatever the *Rotation* Factor. A walking activity, natural and transparent could optimize allocentric representations. Due to the significant effects absence of the *Rotation* factor with the *Treadmill*, we can suppose that for increasing allocentric representations, translational vestibular information is more important than rotational vestibular information. These results are consistent with the findings of [14][17], regarding the importance of walking activity in the development of allocentric representations. However, these results could be different with a joystick condition with gamers participants, accustomed to use a joystick. Once again, it seems to be interesting to add this condition as a factor. Moreover, a positive correlation between the questionnaire assessing the abilities to use maps and this allocentric task can be observed. It means the allocentric representations would be strongly linked with allocentric participants experience and cognitive processes used to manipulate allocentric representations [3].

Due to the different results on the allocentric tasks, it is difficult to summarize this part. In one case (the sketch-mapping task), we did not find motor activity effect. It is already admitted the allocentric representations are related to different cognitive processes and to the manipulation and the repetition of spatial representations [9]. On the other hand (allocentric starting-point estimation task), we found a great impact of the walking activity on allocentric representations, meaning a walking activity improved the creation of allocentric representations [14]. A hypothesis concerns the allocentric tasks used. Indeed, different authors argue that drawing could be an ability to sketch correctly a route [18]. Maybe these two allocentric tasks did not assess the same cognitive processes or spatial representations. The debate about the impact of motor activity on allocentric representations still exists. However, due to the absence of *Rotation* effect, we can state that the rotational component seems to be negligible for tasks mainly driven by allocentric spatial representations [20].

4.3 Spatial Transfer (The wayfinding task)

To recall, this task consisted in reproducing in the real environment, the learned path. We collected two data: the mean speed to finish the task and the directions errors percentage. For the mean speed, the statistical analyses showed a *Rotation* effect; the mean speed was higher for the conditions with *Controlled Rotation*, compared to the condition with *Automatic Rotation*, whatever the *Input Devices* used. It can be supposed that the free learning exploration of the VE at each intersection (whatever the *Input Devices* used) is close to a real learning, optimizing the speed transfer in the real environment. For the percentage of direction errors, we observed an interaction "*Input Devices* x *Rotation*". With the *Treadmill*, performances are the best with *Controlled Rotation*, while with the *Joystick*, best performances have been observed with *Automatic Rotation*. *The Treadmill with Controlled Rotation* condition allows the participants to optimize performances in term of speed and errors percentage. Grant and Magee [19], in a spatial transfer task, already found these results, in comparing a walking interface and a joystick, but both with *Controlled Rotation*. Nevertheless, the *Rotation* factor was not controlled in their study, the superiority of walking interface over joystick may have been induced by freedom of rotation rather than the physical engagement provided by the walking interface, as demonstrated in our study. We can also suppose that the *Treadmill with Controlled Rotation* is very close to a real walking activity, and optimize the spatial transfer performances. However, once again, the *Controlled Rotation* negatively impacts spatial performances with the *Joystick*. These results are very similar to *the Egocentric Estimation Task*, and the *Rotation* factor generates different impact according to the *Input Devices* and the motor activity provided: in a walking situation, a *Controlled Rotation* (with rotational vestibular information) increases performances, while with a *Joystick*, a *Controlled Rotation* affect negatively spatial acquisition. As for *the Egocentric Estimation Task*, we suppose performances in the *Joystick with Controlled Rotation* could be due to the difficulties to manage two directions with the hand, generating visuomotor problems [16]. Considering a similar joystick condition with game experience could give more information about spatial transfer.

To summarize:

- Translational and rotational vestibular information provided by *Controlled Rotation with the Treadmill* optimize spatial transfer [19]. Only translational vestibular information decreases spatial transfer performances.
- With the *Joystick*, the *Automatic Rotation* enabled the best performances.

5 Conclusion

According to our experiments, the motor activity during an interaction and the manual control of rotations had different impacts on spatial transfer. Translational and rotational vestibular information provided by the *Treadmill with Controlled Rotation* optimizes spatial egocentric and transfer performances, as well as the *Joystick with Automatic Rotation*. The question concerning the allocentric representations is most contrasted. In one case a walking activity enhances performances (starting-point estimation task), while in another, no differences were found (sketch mapping task). Further investigation is required to clarify this point.

The novelty of this research concerns the bad performances, whatever the tasks performed, with the *Joystick and the Controlled Rotation*, though often used in videogames or on spatial cognition research. The *Joystick Input Device* may offer an advantage for spatial learning under specific conditions (translational control), close to the *Treadmill with Controlled Rotation*, but not others (translational and rotational controls). All our results showed better performances in the *Joystick with Automatic Rotation* condition (close to *the Controlled Rotation Treadmill*) compared to the *Automatic Rotation Joystick* condition. One hypothesis if that vertical and horizontal hand movements do not provide adequate metaphors of translational and rotational displacements to implement a dialogue between the cognitive and sensorimotor systems that contribute to spatial learning. A condition where participants can only manage the direction of their displacement would give some information about the visuomotor coordination of two directions with the hand. This challenges the debate on the possible advantage of active navigation with a joystick (compared to simple observation), where some studies detected a benefit for spatial learning performances [21][1][2][8]) but others did not ([22][23][24]). Moreover, joystick interfaces are more widely used than treadmills, since they are less expensive, easier to implement from a technological standpoint. This device is also often adapted to the user's needs, notably for people with mobility issues. This the case with the elderly, patients with Parkinson's or Alzheimer's diseases, or sensorimotor injuries ([25][26][27]). Thus, clarifying the impact of joystick use represents a research challenge and is essential to resolve fundamental issues for clinical neuropsychological applications.

References

1. Wallet, G., Sauzeon, H., Rodrigues, J., Larrue, F., N'Kaoua, B.: Virtual / Real Transfer of Spatial Learning: Impact of Activity According to the Retention Delay. St. Heal. (2010)
2. Wallet, G., Sauzeon, H., Pala, P.A., Larrue, F., Zheng, X., N'Kaoua, B.: Virtual/Real transfer of spatial knowledge: benefit from visual fidelity provided in a virtual environment and impact of active navigation. Cyberpsychology, Behavior and Social Networking 14(7-8), 417–423 (2011)
3. Larrue, F., Sauzeon, H., Aguilova, L., Lotte, F., Hachet, M., N'Kaoua, B.: Brain Computer Interface Vs Walking Interface in VR: The impact of motor activity on spatial transfer. In: Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology (VRST 2012), pp. 113–120. ACM, New York (2012)
4. Guilford, J.P., Zimmerman, W.S.: The Guilford-Zimmerman Aptitude Survey. Journal of Applied Psychology (1948)
5. Vandenberg, S.G., Kuse, A.R.: Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and Motor Skills 47(2), 599–604 (1978)
6. Wechsler, D.: Manual for the Wechsler Adult Intelligence Scale-Revised. Psychological Corporation, New York (1981)
7. Kennedy, R., Lane, N., Berbaum, K., Lilienthal, M.: Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. The International Journal of Aviation Psychology 3(3), 203–220 (1993)
8. Waller, D., Hunt, E., Knapp, D.: The Transfer of Spatial Knowledge in Virtual Environment Training. Presence: Teleoperators and Virtual Environments 7(2), 129–143 (1998)
9. Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. Advances in Child Development and Behavior 10, 9–55 (1975)
10. Ruddle, R.A., Volkova, E., Mohler, B., Bülthoff, H.H.: The effect of landmark and body-based sensory information on route knowledge. Memory & Cognition 39(4) (2011)
11. Waller, D., Richardson, A.R.: Correcting distance estimates by interacting with immersive virtual environments: effects of task and available sensory information. Journal of Experimental Psychology: Applied 14(1), 61–72 (2008)
12. Klatzky, R.L., Loomis, J.M., Beall, A.C., Chance, S.S., Golledge, R.G.: Spatial Updating of Self-Position and Orientation During Real, Imagined, and Virtual Locomotion. Psychological Science 9(4), 293–298 (1998)
13. Chance, S.S., Gaunet, F., Beall, A.C., Loomis, J.M.: Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration. Presence: Teleoperators and Virtual Environments 7(2), 168–178 (1998)
14. Ruddle, R.A., Volkova, E., Bülthoff, H.H.: Walking improves your cognitive map in environments that are large-scale and large in extent. ACM Transactions on Computer-Human Interaction 18(2), 1–20 (2011)
15. Riecke, B.E., Cunningham, D.W., Bülthoff, H.H.: Spatial updating in virtual reality: the sufficiency of visual information. Psychol. Res. 71(3), 298–313 (2007)
16. Richardson, A.E., Powers, M.E., Bousquet, L.G.: Video game experience predicts virtual, but not real navigation performance. Computers in Human Behavior 27(1) (2011)
17. Ruddle, R.A., Lessels, S.: The benefits of using a walking interface to navigate virtual environments. ACM Transactions on Computer-Human Interaction 16(1), 1–18 (2009)
18. Golledge, R.G.: Human wayfinding and cognitive maps. In: Wayfinding Behavior, pp. 5–45. The John Hopkins University Press, Baltimore (1999)

19. Grant, S.C., Magee, L.E.: Contributions of Proprioception to Navigation in Virtual Environments. Human Factors: The Journal of the Human Factors and Ergonomics Society 40(3), 489–497 (1998)
20. Ruddle, R.A., Péruch, P.: Effects of proprioceptive feedback and environmental characteristics on spatial learning in virtual environments. International Journal of Human-Computer Studies 60(3), 299–306 (2004)
21. Péruch, P., Vercher, J., Gauthier, G.M.: Acquisition of Spatial Knowledge Through Visual Exploration of Simulated Environments. Ecological Psychology 7(1), 1 (1995)
22. Wilson, P., Foreman, N., Gillett, R., Stanton, D.: Active Versus Passive Processing of Spatial Information in a Computer-Simulated Environment. Ecological Psychology (1997)
23. Wilson, P.N.: Active Exploration of a Virtual Environment Does Not Promote Orientation or Memory for Objects. Environment and Behavior 31(6), 752–763 (1999)
24. Wilson, P.N., Péruch, P.: The influence of interactivity and attention on spatial learning in a desktop virtual environment. Current Psychology of Cognition 21, 601–633 (2002)
25. Rizzo, A., Schultheis, M., Kerns, K., Mateer, C.: Analysis of assets for virtual reality applications in neuropsychology. Neuropsychological Rehabilitation 14(1), 207–239 (2004)
26. Skelton, R.W., Bukach, C.M., Laurance, H.E., Thomas, K.G., Jacobs, J.W.: Humans with traumatic brain injuries show place-learning deficits in computer-generated virtual space. Journal of Clinical and Experimental Neuropsychology 22(2), 157–175 (2000)
27. Lloyd, J., Powell, T.E., Smith, J., Persaud, N.V.: Use of a virtual-reality town for examining route-memory, and techniques for its rehabilitation in people with acquired brain injury. In: Proceedings of the 6th International Disability, Virtual Reality and Associated Technologies, Esbjerg, pp. 175–182 (2006)
28. Montello, D.R.: Navigation. In: Shah, P., Miyake, A. (eds.) The Cambridge Handbook of Visuospatial Thinking, pp. 257–294. Cambridge University, Cambridge (2005)
29. Ishikawa, T., Montello, D.R.: Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. Cognitive Psychology 52(2), 93–129 (2006)

Designing Intuitive Multi-touch 3D Navigation Techniques

Damien Marchal[1], Clément Moerman[2,3], Géry Casiez[3], and Nicolas Roussel[4]

[1] CNRS
[2] Idées-3Com
[3] University of Lille and
[4] Inria Lille Villeneuve d'Ascq, France
{damien.marchal,clement.moerman,gery.casiez}@lifl.fr,
nicolas.roussel@inria.fr
http://ns.inria.fr/mint/MoveAndLook

Abstract. Multi-touch displays have become commonplace over recent years. Numerous applications take advantage of this to support interactions that build on users' knowledge and correspond to daily practices within the real world. 3D applications are also becoming more common on these platforms, but the multi-touch techniques for 3D operations often lag behind 2D ones in terms of intuitiveness and ease of use. Intuitive navigation techniques are particularly needed to make multi-touch 3D applications more useful, and systematic approaches are direly needed to inform their design: existing techniques are still too often designed in ad-hoc ways. In this paper, we propose a methodology based on cognitive principles to address this problem. The methodology combines standard user-centered design practices with optical flow analysis to determine the mappings between navigation controls and multi-touch input. It was used to design the navigation technique of a specific application. This technique proved to be more efficient and preferred by users when compared to existing ones, which provides a first validation of the approach.

Keywords: 3D navigation, multi-touch, interaction technique, design rationale.

1 Introduction

Multi-touch displays have become commonplace over the recent years. Smartphones, tablets, interactive kiosks and systems of other sorts can now detect and react to the presence of two or more contact points on the screen surface. Numerous applications take advantage of this to support *reality-based interactions* [13] that build on users' knowledge and correspond to daily practices within the real world. 3D applications are also becoming more common on these platforms, including games, virtual tours, and CAD applications for both specific, e.g. interior design, and general purposes. But the multi-touch techniques for 3D operations often lag behind the 2D ones in terms of intuitiveness and ease of use. Navigation particularly seems to be the Achilles heel of multi-touch 3D applications. Existing techniques are still too often designed in ad-hoc

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 19–36, 2013.

ways. Intuitive techniques are direly needed to make the applications more useful, and systematic approaches direly needed to inform their design.

The Merriam-Webster Dictionary defines intuitive as "*attaining to direct knowledge or cognition without evident rational thought and inference*". Based on an extensive review of the relevant literature, Ingram et al. also identify *direct manipulation* as the most influential factor determining the intuitiveness of multi-touch systems [10]. Direct manipulation is commonly supported by 2D multi-touch applications, due to the trivial mapping between 2D tasks and the multi-touch input space. However, finding a *direct* mapping between 3D tasks and this 2D input space is much more difficult. To clarify the requirements for intuitive 3D navigation techniques, we propose to turn first to cognitive accounts of the feeling of directness.

Hutchins et al. identify two underlying phenomena that give rise to this feeling: a small cognitive distance, and direct engagement [9]. The cognitive distance is the one "*between the user's intentions and the facilities provided by the machine*". It encompasses the semantic distance, concerned with the meaning of available interactions, and the articulatory one, concerned about their form. For the semantic distance to be small, the system should provide users with adequate commands to concisely express what they want to do. For the articulatory distance to be small, the system should provide an adequate mapping between user actions and the commands. This mapping should not be arbitrary, but should rather favor similarities between user action and command meaning. Lastly, for direct engagement, the system should provide continuous representations of the objects of interest and promptly react to user actions on them. Ultimately, the degree of directness relates to the one to which the system supports skill-based rather than rule-or knowledge-based behaviors [22].

Navigation concerns viewpoint control and is the aggregate of *wayfinding* (cognitive planning of one's route), *travel* (the motor aspects) and *inspection* (for particular proximal views of objects). The importance of each sub-task depends on the considered application. We did not consider wayfinding sub-tasks in this work. We rather focused on multi-touch support for traveling and, to a lesser extent, inspection. Travel techniques support the motor aspects of 3D navigation, allowing users to control the position and orientation of their viewpoint [2]. Viewpoints are typically modeled using seven parameters: the camera's field of view, three Cartesian coordinates (its position) and three Euler angles (its orientation). Controlling these parameters requires a rich command vocabulary because of their number and the different ways to use them. Turning around is pretty straightforward, for example, as it requires the control of a single viewpoint parameter (Figure 1, R_y). Wandering around a horizontal space requires the control of three parameters at the same time (T_x, T_z & R_y). But some navigation tasks require quite a complex coordination of controls, especially when the desired motion is conceptually tied to other reference points. Orbiting around an object, for example, couples planar circular motions with a rotation around an orthogonal axis, both centered on the object (T_y, T_z & R_x for O_x; T_z, T_x & R_y for O_y). Temporary transformations can also be useful, such as adjusting the field of view (FOV) of the camera to remotely inspect a distant place or have a closer look at a nearby detail.

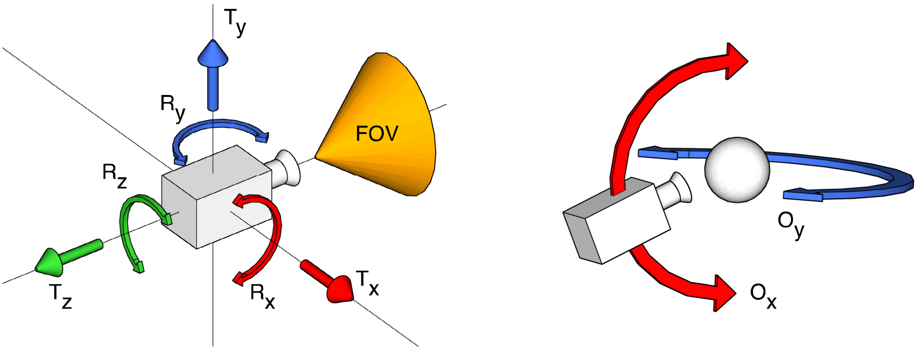


Fig. 1. Typical viewpoint controls: dolly, sidestep and fly (T_z, T_x and T_y), tilt, pan and roll (R_x, R_y and R_z), zoom (FOV) and orbit (O_x and O_y)

To reduce semantic distance, multi-touch navigation techniques should support not only elementary viewpoint controls, but also coordinated ones, including the complex coordination required by externally referenced tasks and temporary ones. To reduce articulatory distance, they should be based on a mapping favoring similarities between user gestures and commands meaning. To support direct engagement, they should provide a close and continuous visual-motor loop. In this paper, we present a methodology for designing multi-touch 3D navigation techniques that meet all these requirements. After discussing related work, we describe our methodology and explain how it was used to design the navigation technique of a particular application. We then provide some implementation details for that application. We finally report on a study that compared this technique with existing ones and provides a first validation of the approach.

2 Related Work

Navigation in virtual 3D worlds has been extensively studied in immersive and desktop environments. Bowman et al. [2] and Christie et al. [4] provide detailed reviews of the relevant concepts and techniques in these contexts, many of which are also relevant to multi-touch environments. Navigation techniques map user actions on one or more input devices to viewpoint controls such as those of Figure 1. Theoretically, a technique could allow users to operate all controls at once. However, it is rarely the case since few input devices (or device combinations) have enough degrees of freedom, and their control properties seldom match the perceptual structure of the navigation tasks [12]. Viewpoint controls thus tend to be organized in groups, which can raise issues about consistency and mode switching. The following review focuses on input-to-control mappings for multi-touch systems, but also discusses pen or mouse-based navigation techniques that could easily be adapted to these systems.

2.1 Basic Viewpoint Control

Different techniques have been proposed to freely and precisely control the viewpoint. These are typically used when navigation is a primary task of the application, one without which it would not be the same. They most often provide only elementary input-to-control mappings to move and orient the viewpoint.

Games using a first-person perspective are probably the most popular 3D applications supporting this kind of navigation on multi-touch platforms. A common practice is to use one or two on-screen joysticks for moving and turning (R_y, T_z or $T_x, T_z + R_y, R_x$). Virtual joysticks make control grouping explicit and make it possible to use non-linear transfer functions for a trade-off between speed and control. They can also be complemented by buttons, sliders or other widgets for discrete or continuous actions on other controls (e.g. FOV, T_y). Coordinating interactions on multiple screen locations is not necessarily easy, though, especially without haptic feedback. Compound controls such as orbiting are thus usually difficult in these settings. Instead of spreading controls across the screen, some techniques combine them using modes. In the RealMyst[1] game, for example, touching the screen and moving horizontally or vertically changes the orientation of the viewpoint while holding still moves forward. Such an approach is of course only practical for a small number of modes.

Multi-touch devices can be used to interact with a 3D scene displayed elsewhere. The ability to use a different view of the scene, a different orientation, or a different physical shape offers new interesting possibilities. The *Finger Walking in Place* (FWIP) technique allows to navigate in a CAVE by mimicking walking movements with fingers on a horizontal multi-touch device [14, 15]: repeated single-touch sliding gestures move the viewpoint forward or backward (T_z), while multi-touch turn gestures rotate it left or right (R_y). The *Follow my Finger* (FmF) technique uses a horizontal multi-touch table to navigate in a scene shown on a vertical screen [1]. The table shows a 2D bird's-eye view of the scene with a camera icon that users can move (T_z, T_x) and orient (R_y) using the 2D Rotate'N Translate technique [16]. The *CubTile* [24] takes the idea of aligning the perceptual structure of the tasks with the input device in the opposite direction. This device combines 5 multi-touch surfaces arranged as a cube so that gestures performed on multiples sides at the same time define a 3D gesture that can be used for 3D interaction. Although designed for object manipulation, the CubTile may well be suitable for navigation tasks.

2.2 Viewpoint Control Facilitation

It might well be the case that the 3D environment in which a user wants to navigate is extremely large [19]. Or the user might be willing to quickly get a glimpse of the scene from different perspectives. Or (s)he might be engaged in repeated tasks requiring frequent switches between two or more viewpoints. In these situations, navigation is just a mean and not an end. One wants to transition between viewpoints but does not necessarily care about all the details of the transition. Even with sophisticated transfer functions, basic viewpoint controls are not enough: one needs faster and integrated techniques to move and orient the viewpoint.

The *Point Of Interest* (POI) desktop technique was precisely designed for rapid controlled movement through a 3D space [17]. After selecting a POI with the mouse,

[1] `http://www.cyanworlds.com/iOS_realMyst/`

it allows to quickly move there by simply pressing a key, the system taking care of the transition with an animation adjusting the viewpoint position (using a logarithmic function of the remaining distance) and reorienting it to face the POI. *UniCam* [25] is a mouse or stylus-based system that uses simple gesture recognition to facilitate a variety of complex navigation tasks including orbiting around specific points, click-to-focus on points and edges, and region zooming. *Navidget* [8] expands on these ideas by allowing the user to not only specify a point of interest but also control the final position and orientation of the viewpoint, rather than inferring them. The system uses single-stroke symbolic gestures and animations in constant time to combine travel and inspection with the ability to go back to a previous viewpoint configuration.

A difficulty when trying to control the viewpoint without any external representation (as in FmF) is that by definition, it cannot be seen. As a consequence, it can only be indirectly manipulated. There is, however, a way to change this: by giving users the impression they can grab the whole scene and manipulate it. Instead of indirectly moving the viewpoint to a particular place, for example, they would thus manipulate the whole scene so that the place comes into the viewpoint. To support this, one needs to make sure that any object touched by a finger remains under it as long as it stays in contact with the surface. This approach has its roots in Gleicher & Witkin's early work on *through-the-lens camera control* [7] and recently received renewed attention after Reisman et al. adapted it to the interactive manipulation of 3D content on 2D multi-touch systems under the name *screen-space* [23]. The *DabR* system [5] uses it in a strict way to support the direct manipulation of basic viewpoint controls, for example. A drawback of the screen-space approach is that its output (viewpoint transformation) is not always predictable due to ambiguities in potential mappings between points in screen space and the 3D scene. To avoid these ambiguities, the *Drag'n Go* technique [20] assigns viewpoint controls to input gestures based on *kinematic correspondence*, i.e. the similarity of the input and output paths [3].

Fu et al. assembled an impressive set of viewpoint control facilitation techniques more or less inspired by the above ones for exploring large-scale 3D astrophysical simulations [6]. Yet this assemblage seems quite ad-hoc. The fact is that designers have little information to rely on when creating a new application.

2.3 A Lack of Systematic Approach

Intuitive navigation techniques are needed to make multi-touch 3D applications useful. Different techniques have been proposed to support basic viewpoint controls and facilitate more complex ones. But comparing these techniques is hard, considering the little information available on their design process and performances. Although the initial design motivation is usually clearly stated in the corresponding papers, the design rationale is largely undocumented. Which decisions were made during the design process, and why, for example? How did the authors come up with the proposed mapping between user actions and viewpoint controls? Why did they decide to group them this way? Without these explanations, one might wonder if there was actually a design process. The authors of UniCam somehow acknowledge this problem: "*Our choice of how to gesturally map the 3 DOFs of camera translation to 2D mouse movements involves some apparently arbitrary choices. (...) In lieu of an* *explanation, we note that from our experience with gestural interaction, the most reliable technique for insuring usable interactions is through empirical evaluations.*" [25].

We collectively need more explanations on the design of these techniques. Evaluations are also important and need to be properly reported. As illustrated by Table 1, few of the techniques we reviewed have been evaluated and even less have been compared to others. We definitely need more comparative evaluations. Without explanations of what is being done and comparisons with existing solutions, there can be no progress in the understanding of the problems and their solutions. Systematic approaches are direly needed to inform the design of new techniques.

Table 1. Summary of the characteristics of the most relevant techniques and systems discussed in this section. Rows prefixed with a star correspond to those specifically designed for multi-touch interaction.

Technique or system	Design motivation	Reported Evaluation
*FWIP [14,15]	walking metaphor	comparative (vs. joystick) & usability testing
*FmF [1]	2D directness	None
POI [17]	speed and control	none
UniCam [25]	integrated suite	empirical?
Navidget [8]	ease of use and control	comparative (vs. standard 3D viewer)
*Screen-space [23]	3D directness	none
*DabR [5]	3D directness	none
*Drag'n Go [20]	multi-scale navigation	comparative (vs. POI, DabR, keyboard+mouse)
*Fu et al [6]	large scale navigation	usability testing

3 Design Methodology

Considering a set of application-specific tasks, how can one map the associated viewpoint controls to the input handles provided by a multi-touch system? In this section, we report on the design of such a mapping for a particular application. Although the resulting technique is specific to that application, we believe our design methodology should be of general interest. The application we worked on is one for reviewing interior designs (Figure 2) that typically runs on computers equipped with a multi-touch screen. Our goal was to design an intuitive navigation technique for this application, as defined earlier, i.e. one with reduced semantic and articulatory distances and a close and continuous visual-motor loop. In the following, we explain how standard user-centered practices and optical flow analysis helped us identify application controls and input handles and define the mappings between them.

3.1 Identifying Navigation Tasks and Associated Controls

A way to reduce the semantic distance *by design* is to work with users to define the navigation commands from their perspective, rather than the application developers'.

Developers usually think of viewpoint control in terms of parametric modifications of the camera model, as these are ultimately the only controls available. But they typically have little insight into the ways these should be grouped. Users, on the other hand, typically think in terms of high-level situated tasks (i.e. context-specific) that can help structure the design space. To provide users with adequate commands to concisely express what they want to do, application designers must identify these tasks.

Fig. 2. Sample interior design. Reviewing such a scene requires the ability to quickly navigate through it (including moving from one floor to the other), to orient the viewpoint in a precise way (e.g. to check the view from the couch) and to inspect particular objects (e.g. the ones on the table).

Based on our specific application context (Figure 2), the related work we previously described and interviews of potential users, we decided to focus on the following tasks, in decreasing order of importance:

- *Move around* - Users need to be able to move around the virtual interior the way they do in the real world, i.e. mostly by moving forward or backward (T_z) while possibly turning left or right (R_y). Although commonly supported by video games, sidestepping (or *strafing*,T_x) is rarely used in real world situations and thus of lesser importance. Altitude control (T_y) is also pretty limited in the real world without assistance, and thus also of lesser importance.
- *Look around* - Adjusting the viewpoint orientation is another important task that must be supported by the considered application. Users need to be able to look left and right (R_y) as well as up and down (R_x). The third rotation of the camera (R_z) does not seem necessary as people have limited control on it in the real world and it does not change what they see but only how they see it.
- *Circle around P* - When focusing on a particular object or area, users need to be able to look at it from different sides. This is typically achieved by orbiting around a previously specified point (P) in the horizontal plane (O_y).
- *Scrutinize P* - Looking at a particular point of interest (P) from different sides might not be enough. One might want to have a closer look at it. In real-life, one can bend over or use optical tools such as a magnifying glass or binoculars to temporarily modify one's field of view (FOV).

Table 2 summarizes the viewpoint controls of Figure 1 associated with the above four tasks to be supported by our particular application. Having identified the viewpoint model's degrees of freedom we want to control and taken a few first steps into their organization, we must now turn to the input device (the multi-touch screen) and examine the handles it provides for that control.

Table 2. Relevance of viewpoint controls to high-level navigation tasks, by decreasing order of importance.●: relevant,◐: partially relevant, ○: not relevant.

	T_x	T_y	T_x	R_x	R_y	FOV	O_y
Move around	◐	◐	●	○	●	○	○
Look around	○	○	○	●	●	○	○
Circle around P	○	○	○	○	○	○	●
Scrutinize P	○	○	○	○	○	●	○

3.2 Identifying Input Handles

Although some touch sensing technologies provide rich information about contact regions, including their shape or the applied force, most multi-touch APIs only expose the 2D coordinates of their centroid. One might thus think that using n fingers, users should be able to control $2 \times n$ degrees of freedom. But in reality, it is never the case. Multi-touch systems can't distinguish between fingers, so degrees of freedom cannot be univocally associated to them. The order of appearance of contacts or hit-testing with specific on-screen areas can be used for these associations. But in the end, interaction will always be constrained by limited finger individuation: it is quite difficult to move one finger without some degree of involuntary movement at one or more of the others [11].

Instead of considering contacts individually, one can group them using different methods (e.g. hit-testing, proximity, hand identification) and extract from the collated movement information global parameters to be associated with degrees of freedom to control. A common practice is to consider multi-touch gestures on objects as *Rotate-Scale-Translate* (RST) manipulations and to determine and characterize the principal transformation involved - e.g. $R(\alpha, C)$ for a *turn* gesture, $S(\kappa, C)$ for a *pinch* or a *spread*, and $T(x, y)$ for a *swipe*[2]. Figure 3 shows a simplified view of the state machine we used, based on this approach. The machine differentiates four interaction states (shown in gray): one for single-touch interaction, and three differentiating multi-touch interactions based on the first principal transformation detected.

Having described the desirable viewpoint controls for our application (Table 2) and the different handles provided by a multi-touch screen, i.e. $T(x, y)$ for single-touch interactions and $R(\alpha, C)$, $S(\kappa, C)$ & $T(x, y)$ for multi-touch ones, we will now examine the mappings between them.

[2] The parameters associated to each transformation will be explained in the next section.

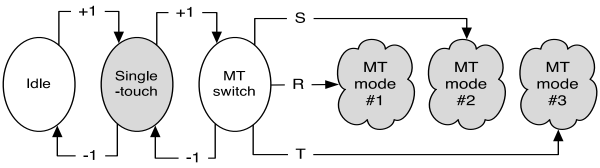
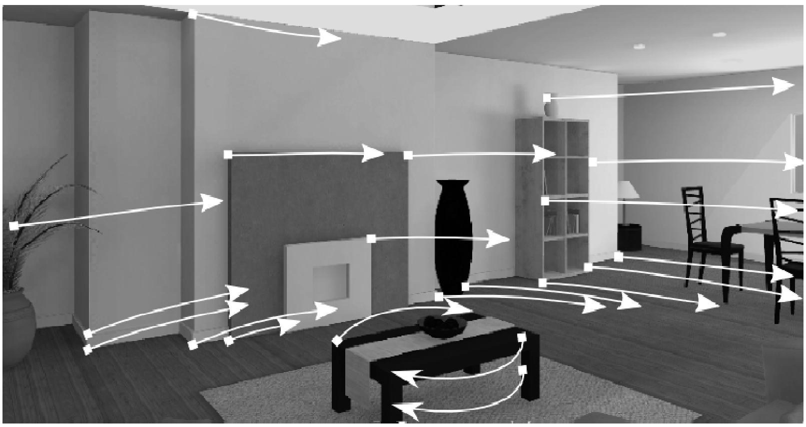


Fig. 3. Simplified view of a multi-touch state machine based on an RST classifier. +1 and -1 transitions are triggered by contact addition and removal. The R, S and T transitions are triggered based on the first principal transformation detected. Note that once in mode #1, #2 or #3, multi-touch interaction does not have to be restricted to the transformation initially detected.

3.3 Choosing the Right Mappings

A way to reduce the articulatory distance and ensure tight coupling between perception and action is to choose the mappings between viewpoint controls and multi-touch gestures so that contact trajectories match the scene transformations caused by viewpoint modifications. To achieve this, we created video clips illustrating the effect of the 7 viewpoint controls of Table 2 on a particular scene. We used OpenCV to compute and visualize the optical flow of each of these videos (Figure 4). We then compared these flows to those corresponding to multi-touch RST manipulations, i.e. $T(x,y)$ when using a single finger and $R(\alpha,C)$, $S(\kappa,C)$, $T(x,y)$ otherwise.



T_x	T_y	T_z	R_x	R_y	FOV	O_y

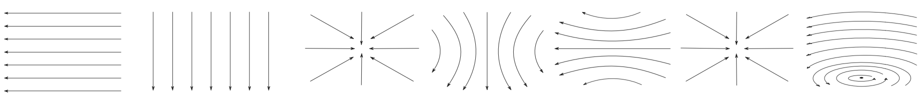


Fig. 4. Top image: optical flow computed over time with OpenCV when circling the viewpoint along the vertical axis (R_y). Bottom images: stylized renderings of the flows corresponding to the 7 viewpoint controls of Table 2 (arrows could all point in the opposite direction).

Table 3 summarizes the compatibility between gesture flows and optical flows using a three-level scale. Starting from this table, we applied the following heuristics to choose between alternative mappings:

- As MT-$R(\alpha, C)$ is the most compatible gesture with O_y, we decided to map the two.
- T_z and FOV are compatible with the same gestures, MT-$S(\kappa, C)$ and ST-$T(x, y)$. Since T_z is one of the most important controls, we wanted to keep it as simple as possible and thus preferred a single-touch gesture for it. Moving forward/backward seemed better matched with a vertical movement rather than a horizontal one, so we chose ST-$T(., y)$. For FOV, we chose MT-$S(\kappa, C)$.
- For ST-$T(x, .)$, we were left with T_x, R_y and O_y, the first two being more important than the third one. We decided to map ST-$T(x, .)$ to R_y so that single-touch interaction would support both *Move around* (with T_z and R_y) and *Look around* (with R_y).
- For MT-$T(., y)$, we were left with T_y and R_x. We chose the latter, as looking up/down was considered more important than controlling one's altitude.
- For MT-$T(x, .)$, we were left with T_x and R_y. Informal tests convinced us that the latter was preferable, considering our previous choice of R_x for MT-$T(., y)$.

Table 3. Compatibility between gesture flows (rows) and optical flows (columns): ○ incompatible, ● compatible, ◐ compatible under one of the conditions below. A dot in place of x or y indicates that this component is ignored. The gray cells correspond to the chosen mapping.

		T_x	T_y	T_z	R_x	R_y	FOV	O_y
ST	$T(x, .)$	●	○	●	○	◐(1)	●	◐(2)
	$T(., y)$	○	●	●	◐(3)	○	●	○
MT	$R(\alpha, C)$	○	○	○	◐(4)	◐(5)	○	●
MT	$S(\kappa, C)$	○	○	●	○	○	●	○
MT	$T(x, .)$	●	○	◐(6)	○	◐(1)	◐(6)	◐(2)
	$T(., y)$	○	●	◐(6)	◐(3)	○	◐(6)	○

(1) compatibility is inversely proportional to the vertical distance to the center of the screen
(2) the point of interest (P) must have been previously specified
(3) compatibility is inversely proportional to the horizontal distance to the center of the screen
(4) C must be in the middle of a vertical border of the screen, i.e. left or right
(5) C must be in the middle of a horizontal border of the screen, i.e. top or bottom
(6) contacts must be "close enough", i.e. within a certain radius, so they can be reduced to single-touch interaction

4 Implementation: The Move and Look Technique

A close look at Table 2 and the chosen mapping in Table 3 shows that ST-$T(x, y)$ corresponds to *Move around* while MT-$R(\alpha, C)$ corresponds to *Circle around*, MT-$S(\kappa, C)$ to *Scrutinize* and MT-$T(x, y)$ to *Look around*. As illustrated by Figure 5, each of our 4 high-level navigation tasks can thus be associated to an interaction state of the machine shown in Figure 3. This section details the implementation of the resulting navigation technique, which we called *Move&Look*.

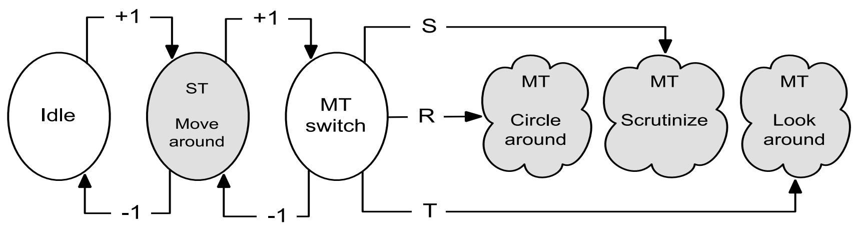


Fig. 5. The Move&Look technique, instanciated from Figure 3

4.1 Single-Touch Interaction: *Move Around*

When a single contact is detected, subsequent $T(x, y)$ finger movements are mapped to R_y, T_z camera movements to support the *move around* task. When touch is detected, a ray is casted in the 3D scene from the camera center through the contact point in the camera plane. The intersection with the scene (P) defines the point of interest, and the ray a path towards it. Progression along the path is controlled through finger movements in the following way:

- Lateral movements $(T(x, .))$ do not affect the camera position. Proximal finger movements translate the camera towards P and distal movements translate it backwards along the path $(T(., y))$.
- The distance between the initial contact point and the bottom of the display is mapped to the entire path length: the destination is reached when the finger reaches the bottom of the display.
- Distal finger movements past the initial contact point (i.e. above it) continue moving the camera backwards along the path. For consistency, the scale factor remains the same as when closing in on P.

Users can turn the camera left and right (R_y) through lateral finger movements. The camera orientation is computed so as to always keep the projection of P under the finger. R_y is computed either analytically [7] or numerically [23] to minimize the distance between the previous projection of P and the current finger position (we used the minimizer from ALGLIB to solve the different minimization problems).

4.2 Multi-touch Interaction Switch: RST Classifier

When multiple contacts are detected, their movement is analyzed to determine whether the state machine should switch to *circle around*, *scrutinize* or *look around*. The movement of the n contact points is interpreted as a rigid transformation combining a rotation $R(\alpha, C)$, an homogeneous scaling $S(\kappa, C)$, and a translation $T(x, y)$. The initial position of the contact points is noted r_i and their current position c_i. The R, S and T transformations correspond to the minimization of the cost function F defined by Equation 1. $T(x, y)$ is first computed from the centroids of the initial set of contact points (C_i) and the current one (C_c) according to equations 2, 3 and 4. The rotation angle α is the one that minimizes the cost function of Equation 5 and is computed using Equation 6 where $co = c_i - C_i$ and $ro = r_i - C_c$. The scale factor κ is similarly the one that minimizes the cost function of Equation 7 and is computed using Equation 8 where $cr = R^{-1}(\alpha).c$:

$$F(x,y,\alpha,\kappa) = \sum_{0\le i\le n} \|T(x,y)R(\alpha)S(\kappa)r_i - c_i\| \quad (1)$$

$$C_i = \frac{1}{n}\sum_{0\le i\le n} r_i \quad (2)$$

$$C_c = \frac{1}{n}\sum_{0\le i\le n} \mathrm{c_i} \quad (3)$$

$$T(x,y) = C_c - C_i \quad (4)$$

$$G(\alpha) = \sum_{0\le i\le n} \|R(\alpha)(r_i - C_i) - (c_i - C_c)\| \quad (5)$$

$$\alpha = \mathrm{atan}\,2\left(-\sum_{0\le i\le n} co_x ro_y - co_y ro_x\,, \sum_{0\le i\le n} co_y ro_y + co_x ro_x\right) \quad (6)$$

$$H(\kappa) = \sum_{0\le i\le n} \|S(\kappa)ro - R^{-1}(\alpha)co\| \quad (7)$$

$$\kappa = \frac{cr_x^2 + cr_y^2}{cr_x ro_x + cr_y ro_y} \quad (8)$$

The only rotations considered in these equations are those centered on the centroid of the contact points. Although we perceive it as an elementary rotation, moving one's index finger around one's thumb while keeping this one steady will thus be interpreted as a combination of a centroid-centered rotation and a translation. To tackle this problem, we weight all contact positions by the inverse of their traveled distance when computing C, the center of both the rotation $R(\alpha,C)$ and the homogeneous scaling $S(\kappa,C).T(x,y)$ is also adjusted by removing the displacement possibly introduced by $R(\alpha,C)$.

The x, y, C, α and κ parameters resulting from the above computations are used to determine the prominent gesture among *Rotate*, *Swipe* and *Pinch*. Our classifier considers one model $M(t)$ for each gesture type and returns the one that better fits the observations (highest R^2 value). The models map the initial configuration r_i to an estimated state $\widetilde{c_i}$ (Equation 9). The estimated error (the residual sum of squares) and the coefficient of determination R^2 are then computed using Equations 10, 11 and 12.

$$\widetilde{c_i} = M(t).r_i \quad (9)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (10)$$

$$SS_{res} = \sum_{0\le i\le n} \|c_i - \widetilde{c_i}\|^2 \quad (11)$$

$$SS_{tot} = \sum_{0\le i\le n} \|r_i - c_i\|^2 \quad (12)$$

Our classifying method is similar to the *GestureMatching* method used by Nacenta et al. [21], but instead of classifying the combined contribution of *Rotate*, *Swipe* and *Pinch* gestures, ours allows to classify the contribution of individual gesture types. Our method requires enough information to properly work. The classifier is thus enabled only when the summed distance covered by the contact points is beyond a given threshold. Based on preliminary tests, we found that a value of 10 pixels on a 90PPI display (around 2.8 mm) provides a good trade-off between latency and success rate, which is in agreement with other thresholds reported in the literature for similar applications [25].

4.3 Multi-touch Gestures: Circle Around, Look Around and Scrutinize

When the classifier detects a prominent *Rotate* gesture, the technique enters the *circle around* state of Figure 5 until all contacts are lost. $R(\alpha, C)$ provides the pivot point to rotate the scene and the angle of rotation (the center of the 3D rotation is computed from the projection of C in the 3D scene).

When a *Swipe* gesture is detected, the technique enters the *look-around* state until all contacts are lost. $T(x, y)$ is then used to rotate the camera (R_x, R_y) in a way similar to *move around*, but with two degrees of freedom instead of one.

When a *Pinch* gesture is detected, the technique enters the *scrutinize* state until all contacts are lost. The scale factor of $S(\kappa, C)$ is used to adjust the FOV of the camera. To ensure smooth camera movements, its look-at point remains fixed while contacts are moving. The FOV is restored to its initial value when all contacts are lost. This state further supports remote inspection by using $T(x, y)$ to rotate the camera (R_x, R_y), as in the *look-around* state.

5 Experiment

Our main motivation in this experiment was to assess our design choices by comparing Move&Look to other techniques from the literature (Screen-space [23], DabR [5] and Drag'n Go [20]) or available in commercial products (Virtual joysticks and the RealMyst technique described above), most of which have never been evaluated nor compared.

5.1 Task

Informal user testing with Move&Look suggested the technique was particularly efficient for interior designs mainly consisting of flat orthogonal surfaces. Encouraged by this, we wanted to assess the effectiveness of the technique in a more demanding environment. The task we chose consisted in collecting spheres placed inside boxes in an outdoor environment, and dropping them in a fountain at the center of the scene (Figure 6). To provide a fair comparison between multi-scale navigation techniques (Drag'n Go and Move&Look) and the others and focus on the evaluation of camera movements, the boxes were not positioned far away from each other but close to the central drop zone.

Participants had to find the boxes in the scene. For each box, they had to position the camera in front of its only open face to pick up the sphere it contained. A sphere would turn from red to green when the camera was close enough to indicate one could touch it to pick it up. Participants could carry only one sphere at a time, and it was automatically dropped once in the drop zone. Collision detection prevented participants from moving through objects, and a trial was considered as fully completed after all the spheres had been dropped. Participants were instructed to perform this as quickly as possible. They could ask the experimenter to reset the camera to its initial position or withdraw a trial if they felt unable to complete it. They were not encouraged to do so, however. The experimenter rather encouraged them to finish a trial if he felt they could succeed.

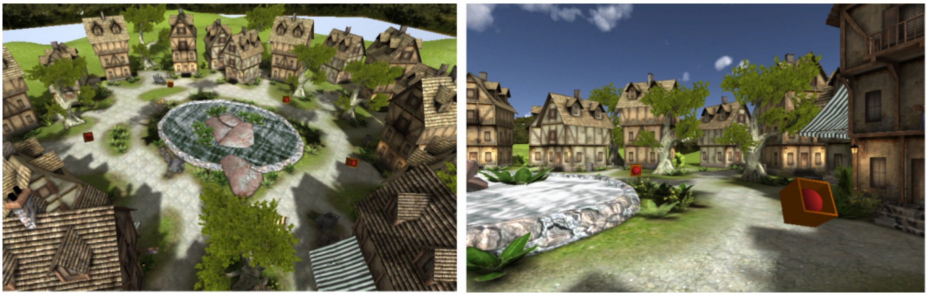


Fig. 6. Left: overview of the 3D environment used in the experiment. Right: detailed view showing a box containing a sphere to pick up and drop in the fountain.

5.2 Participants

Twelve unpaid male participants with a mean age of 35 (SD=14) served in the experiment. Five of them used a computer on a daily basis, played 3D video games and were familiar with touch-screens through mobile phones or tablets. Seven of them had a low experience with video games and were novice with touch interfaces.

5.3 Apparatus, Design and Procedure

Participants were seated in front of a 22” 3M multi-touch screen orientated at an angle of about 70° from a horizontal desk. The experiment was implemented using Unity 3.5[3]. A repeated measures within-subjects design was used. The independent variable was the interaction technique (TECH) with six levels: DabR, Screen-space, Drag’n Go, RealMyst (a custom implementation of the RealMyst technique), Virtual joysticks (a standard combination of two Unity virtual joysticks displayed at fixed positions) and Move&Look. A trial consisted in collecting 4 spheres and each technique was evaluated with 3 successive trials (TRIAL). In summary the experimental design was: 12 participants × 6 TECH × 3 TRIAL = 212 total trials.

The presentation order for TECH was counter-balanced across participants using a balanced Latin Square design. To favor expert usage and a fair comparison between techniques, each was first introduced by the experimenter with a demo and then a

[3] `http://unity3d.com/unity/whats-new/unity-3.5`

training session. Participants could also use a cheat sheet throughout the experiment. After each technique, participants filled a questionnaire inspired by the Nasa TLX test and at the end of the experiment, they were asked to rank the techniques and give additional feedback.

5.4 Results

The dependent variables were the completion time, the number of give-ups and camera resets, and the qualitative results.

Numbers of Give-Ups and Camera Resets — 25% of trials were aborted for Screen-space, 17% for RealMyst, 8% for Virtual joysticks and 0% for Drag'n Go, DabR and Move&Look. The camera was reseted in 66% of all trials for Screen-space, 14% for Virtual joysticks, 11% for RealMyst, 3% for DabR and Move&Look, and 0% for Drag'n Go.

Task Completion Time — Task completion time is defined as the time needed to successfully collect the four spheres and drop them in the fountain. Trials where participants gave up were removed for the analysis. Trials at least three standard deviations away from the mean for each TECH condition were considered as outliers and also removed. A repeated measures ANOVA showed a significant effect of TECH ($F_{5,55} = 9.9, p < 0.001$). Subsequent pairwise comparison showed significant differences ($p < 0.005$) between Drag'n Go and Screen-space, Drag'n Go and RealMyst, Move&Look and Screen-space, and Move&Look and RealMyst. No significant difference was found between Drag'n Go and Move&Look. Completion times were $97s$ for DabR, $184s$ for Screen-space, $60s$ for Drag'n Go, $117s$ for RealMyst, $111s$ for Virtual joysticks and $72s$ for Move&Look.

User Ranking and Questionnaire — The participants ranked the techniques in decreasing order of preference. Overall, Move&Look came first (10 participants ranked it first and 2 ranked it second) followed by Drag'n Go, DabR, Virtual joysticks, RealMyst and Screen-space. The participants who ranked Move&Look first explained it nicely complements Drag'n Go as it allows to control more degrees of freedom while keeping the navigation intuitive: it does not require focusing on the gestures to execute nor does it require planning a trajectory in the scene to reach a target. Screen-space was ranked last considering its lack of intuitiveness: in spite of the frequent use of the cheat sheet the participants did not understand how to effectively use the technique to navigate the way they wanted (we believe the important semantic distance explains this gap between users' intentions and the system's behaviours). These subjective results are in agreement with the quantitative results found for completion time and the numbers of give-ups up and camera resets.

After each technique, the participants answered questions related to the following six criteria on a 5 point Likert scale: mental demand, physical demand, performance, effort, frustration and satisfaction. The questions asked were similar to the ones available in the Nasa TLX test. We ran a Friedman analysis with Bonferroni-corrected Wilcoxon post-hoc analyses. This analysis shows significant differences between the techniques for all criteria, especially for the techniques at the bottom of the participants' ranking. Table 5 summarizes the significant differences that were found.

Table 4. Details of the post-hoc analysis for cases where one or more significant differences were found (●: significant difference, ○: non significant difference)

	Mental demand	Physical demand	Perfor-mance	Effort	Frustra-tion	Satisfac-tion
Move&Look-Screen-space	●	●	●	●	●	●
Move&Look-RealMyst	●	●	●	○	●	●
Drag'n Go-Screen-space	●	●	●	●	○	●
Drag'n Go-RealMyst	●	●	●	○	○	○
Drag'n Go-DabR	○	●	○	○	○	○
RealMyst-Virtual joysticks	○	●	○	○	○	○
RealMyst-DabR	○	○	○	○	●	○
Screen-space-DabR	●	○	○	○	●	●
Screen-space-Virtual joystick	●	●	○	○	○	●

User Feedback and Observations — During the experiment we encouraged the participants to "think aloud" and freely comment on the interaction techniques. Comments were overall in agreement with the user ranking.

Screen-space received the most negative critics. All participants repeatedly reported their frustration with this technique. They felt out of control and found the mappings between fingers and camera movements inconsistent. The finger movements corresponding to different screen-space controls can indeed be quite similar, as illustrated by the optical flows of Figure 4 (e.g. T_x and R_y). The output of the screen-space solver is also strongly influenced by the picked point in the 3D scene, and thus by the geometrical shape of the underlying objects. Lastly, the movements to execute in order to move forward (T_z) and to zoom (FOV) depend on whether the initial contact point is above or below the invisible horizon (in the former case, one has to move up, in the latter, one has to move down). All these reasons probably contribute to the fact that users were not able to anticipate camera motions. The comparison of Screen-space to other interaction techniques in 3D manipulation tasks corroborates these observations [18].

Participants found DabR, Virtual joysticks and RealMyst either too slow or too fast. We hypothesize this was caused by the use of transfer functions not specifically tuned for the particular 3D environment we used: long distances took too much time to travel while participants traveled too fast on short distances. Participants found the Virtual joysticks to be less fatiguing. We hypothesize this was due to the use of rate control, which reduces physical movements. Participants reported an important fatigue when using DabR and complained they had to pay attention to the number of fingers they used. They complained about the delay introduced by the time-based mode switch used by RealMyst and the fact that the traveling direction is not towards the selected point but along the T_z axis of the camera. Drag'n Go was particularly appreciated for its ability to quickly reach distant targets, but moving to a box while orienting the viewpoint in order to pick the sphere was found more difficult and required some planning. This was not reported as a problem with Move&Look.

6 Conclusion and Future Work

In this paper, we proposed an original methodology based on user-centered practices and optical flow analysis to address the problem of designing intuitive multi-touch navigation techniques for 3D environments. User-centered practices allow to define the navigation commands from the user's perspective while the optical flow analysis provides guidelines for defining intuitive multi-touch gestures to perform these commands. We instantiated this methodology for tasks articulated around the review of interior designs, which led to the design of a new interaction technique, Move&Look. The comparison of this technique to state of the art ones in a controlled experiment showed its overall superiority and revealed usability problems with the others. These results provide a first validation of the proposed design methodology. The methodology should be applied in other navigation contexts in order to further assess its effectiveness. The robustness of the proposed RST classifier should be formally evaluated, and it can certainly be improved. Even if participants did not complain about it, we observed them flattening their rotation gestures for the *circle around* command, probably because they unconsciously followed the corresponding optical flow. Our classifier could be modified to better take into account this oval shape, for example.

References

1. Ajaj, R., Vernier, F., Jacquemin, C.: Follow my finger navigation. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 228–231. Springer, Heidelberg (2009)
2. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D user interfaces: theory and practice. Addison-Wesley/Pearson Education (2003)
3. Britton, E.G., Lipscomb, J.S., Pique, M.E.: Making nested rotations convenient for the user. In: Proc. of SIGGRAPH 1978, pp. 222–227. ACM (1978)
4. Christie, M., Olivier, P.: Camera control in computer graphics: models, techniques and applications. In: ACM SIGGRAPH ASIA 2009 Courses, pp. 3:1–3:197. ACM (2009)
5. Edelmann, J., Schilling, A., Fleck, S.: The DabR - a multitouch system for intuitive 3D scene navigation. In: Proc. of 3DTV-CON, pp. 1–4 (2009)
6. Fu, C.W., Goh, W.B., Ng, J.A.: Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In: Proc. of CHI 2010. pp. 2213–2222. ACM (2010)
7. Gleicher, M., Witkin, A.: Through-the-lens camera control. In: Proc. of SIGGRAPH 1992, vol. 26, pp. 331–340 (1992)
8. Hachet, M., Decle, F., Knodel, S., Guitton, P.: Navidget for easy 3D camera positioning from 2D inputs. In: Proc. of 3DUI 2008, pp. 83–89. IEEE (2008)
9. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct manipulation interfaces. Human-Computer Interaction 1(4), 311–338 (1985)
10. Ingram, A., Wang, X., Ribarsky, W.: Towards the establishment of a framework for intuitive multi-touch interaction design. In: Proc. of AVI 2012, pp. 66–73. ACM (2012)
11. Ingram, J.N., Körding, K.P., Howard, I.S., Wolpert, D.M.: The statistics of natural hand movements. Experimental Brain Research 188(2), 223–236 (2008)
12. Jacob, R.J.K., Sibert, L.E., McFarlane, D.C., Mullen Jr., M.P.: Integrality and separability of input devices. ACM ToCHI 1(1), 3–26 (1994)

13. Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-based interaction: A framework for post-wimp interfaces. In: Proc. of CHI 2008, pp. 201–210. ACM (2008)
14. Kim, J.S., Gračanin, D., Matković, K., Quek, F.: iPhone/iPod Touch as input devices for navigation in immersive virtual environments. In: Proc. of VR 2009, pp. 261–262. IEEE (2009)
15. Kim, J.-S., Gračanin, D., Matković, K., Quek, F.: Finger walking in place (FWIP): A traveling technique in virtual environments. In: Butz, A., Fisher, B., Krüger, A., Olivier, P., Christie, M. (eds.) SG 2008. LNCS, vol. 5166, pp. 58–69. Springer, Heidelberg (2008)
16. Kruger, R., Carpendale, S., Scott, S.D., Tang, A.: Fluid integration of rotation and translation. In: Proc. of CHI 2005, pp. 601–610. ACM (2005)
17. Mackinlay, J.D., Card, S.K., Robertson, G.G.: Rapid controlled movement through a virtual 3D workspace. In: Proc. of SIGGRAPH 1990, pp. 171–176. ACM (1990)
18. Martinet, A., Casiez, G., Grisoni, L.: Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. IEEE TVCG 18(3), 369–380 (2012)
19. McCrae, J., Mordatch, I., Glueck, M., Khan, A.: Multiscale 3D navigation. In: Proc. of I3D 2009, pp. 7–14. ACM (2009)
20. Moerman, C., Marchal, D., Grisoni, L.: Drag'n Go: Simple and fast navigation in virtual environment. In: Proc. of 3DUI 2012, pp. 15–18. IEEE (2012)
21. Nacenta, M.A., Baudisch, P., Benko, H., Wilson, A.: Separability of spatial manipulations in multi-touch interfaces. In: Proc. of GI 2009. pp. 175–182. Canadian Information Processing Society (2009)
22. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man and Cybernetics 13(3), 257–266 (1983)
23. Reisman, J.L., Davidson, P.L., Han, J.Y.: A screen-space formulation for 2D and 3D direct manipulation. In: Proc. of UIST 2009, pp. 69–78. ACM (2009)
24. de la Rivière, J.B., Kervégant, C., Orvain, E., Dittlo, N.: CubTile: a multi-touch cubic interface. In: Proc. of VRST 2008, pp. 69–72. ACM (2008)
25. Zeleznik, R., Forsberg, A.: UniCam - 2D gestural camera controls for 3D environments. In: Proc. of I3D 1999, pp. 169–173. ACM (1999)

Truly Useful 3D Drawing System for Professional Designer by "Life-Sized and Operable" Feature and New Interaction

Shun'ichi Tano[1], Shinya Yamamoto[1], Junko Ichino[1], Tomonori Hashiyama[1], and Mitsuru Iwata[2]

[1] University of Electro-Communications, Graduate School of Information Systems, 1-5-1 Chifugaoka, Chofu-shi, Tokyo, Japan
[2] Tokyo Metropolitan College of Industrial Technology, 1-10-40 Higashioi, Shinagawa-ku, Tokyo, Japan
Tano@is.uec.ac.jp

Abstract. "Media" is an artifact that expands our creativity and intelligence. We have been studying the use of "Rich Media" to support creative and intelligent human activities. Specifically, for over ten years we have focused on the 3D space as one of "Rich Media" and developed many 3D sketch systems that support the design of 3D objects. However, their long-term evaluation has revealed that they are not used by designers in real fields. Even worse, they are treated as if they were just mere attractions in an amusement park. The fundamental problem is the lack of the indispensable function for 3D space. In this paper, we propose new design principles, "life-size and operability", which make the 3D sketch system truly valuable for the designer. The new 3D sketch system is designed on the basis of "life-size and operability", developed, and evaluated successfully.

Keywords: 3D Sketch, Life-size, Operability, Professional Designer, Mixed reality.

1 Introduction

"Media" is an artifact that expands our creativity and intelligence. The oldest media is words and numbers. The computer is now widely used as a media.

We have been studying a wide range of creativity-centered media to ensure that systems truly support creative and intelligent human activities. They range from those used by knowledge workers to those for car-exterior designers [1-9].

Specifically, for over ten years we have developed many 3D sketch systems that support the design of 3D objects, because the 3D sketch cannot be realized without the power of advanced information communication technology (ICT) [6-9]. We regarded the 3D sketch as the drastic extension of the traditional "pen and paper" media made possible by the power of ICT.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 37–55, 2013.

However, long-term evaluation has revealed that 3D sketches are not used by designers in real fields. Even worse, they are treated as if they were just mere attractions in an amusement park.

It shows that the rich media certainly fascinates the ordinary users but is ignored by the professional users in some cases. It may be the serious problem because there are many systems that blindly utilize the rich multimedia without the long-term user evaluation [18].

In this paper, we analyze the fundamental problems that prevent the systems from being used professionally. We point out that it is the lack of the indispensable function for 3D space. Then we propose a new design concept inspired by "mixed reality", that makes the 3D sketch system truly valuable for the designer, and exemplify the feasibility of the new design concept by describing our latest prototype system.

2 Related Works and Purpose

2.1 Related Works

Conventional research into 3D sketching can be categorized into two types. The first is generating 3D sketches from 2D sketches [15, 16]. The designer draws a 2D sketch, then the system converts it into a 3D sketch on the basis of certain assumptions, and finally the system displays it in a 3D space.

The second is drawing the 3D sketch directly in midair [10-14, 17]. The 3D lines are displayed as they are or as transformed smooth lines and converted into the model description in some systems [14, 17].

We have also developed a series of 3D sketch systems [6-9] in both categories mentioned above.

Although each system has its own strength and has been successfully evaluated by the designers, the common problem is that they are not utilized continuously by professional designers in daily design tasks over long periods of time. They are missing something that would make them indispensable for professional use.

2.2 Purpose of This Paper

In this paper, first the role of 3D space is categorized into two types: "draw in 3D space" and "view in 3D space".

Second, our four trial systems are briefly shown as examples that support all roles of 3D space.

Third, the long-term evaluation of the four trial systems is summarized. It shows that the fundamental problem is the lack of indispensable functions for 3D space.

Fourth, we propose new design principles "life-size and operability", which make the 3D sketch system truly valuable for the professional designer.

Finally, our latest 3D sketch system designed by the new design principle is explained in detail and evaluated by professional designers.

3 Our Trials for 3D Sketch System

We have developed a series of 3D sketch systems. The following subsection describes four typical systems [6-9] (called Systems 1, 2, 3 and 4).

3.1 Role of 3D Space

2D/3D space can be used in two ways as shown in Fig. 1. The first is the space where the designer draws the objects. To draw in 3D space means to draw objects in the 3D space directly, i.e., for designers to sketch them in midair in front of themselves. To draw in 2D space means to draw objects in the 2D space, i.e., using a pen and paper.

The second one is the space where the designer looks at the objects. To view in 3D space means designers look at objects in midair in front of them (stereovision). To view in 2D space means to look at objects as perspective 2D images on the 2D plane.

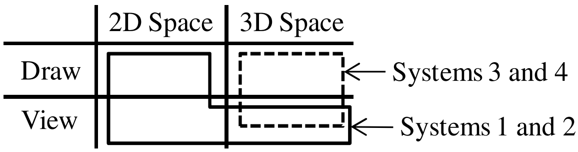


Fig. 1. "Draw in 2D/3D space" and "view in 2D/3D space"

Systems 1 and 2 support the 2D-draw and 2D/3D-view. The designer draws a 2D sketch and views it both in stereovision in the 3D space and as a perspective 2D image on the 2D plane. The reason the systems do not support 3D-draw is the difficulty of drawing in a 3D space directly.

Systems 3 and 4 have a mechanism to compensate for this difficulty, so they support the 3D-draw and 3D-view.

3.2 Our 3D System-1: "Godzilla"

System 1, called "Godzilla", aims to support creative design, specifically that of car-exterior designers.

3.2.1 Design Flow

Fig. 2 shows the typical design flow. First, the designer draws the concept image on the 2D pad (a tablet with an LCD) as shown in Fig. 3(a). The designer can grasp the sketch and hold it in midair, and it appears as a 3D image on the 3D pad (stereovision TV) as shown in Fig. 3(b). While holding and rotating the 3D-image, the designer can look at it from different viewpoints. When the designer grasps the image and puts it onto the 2D pad, it appears on the 2D pad as a 2D sketch. Note that our system displays a hand-drawn sketch all the time, even in the 3D space, and can automatically recognize the 3D shape of a 2D image and transform between the 2D and 3D sketches with different viewpoints while preserving the designer's pen touch.

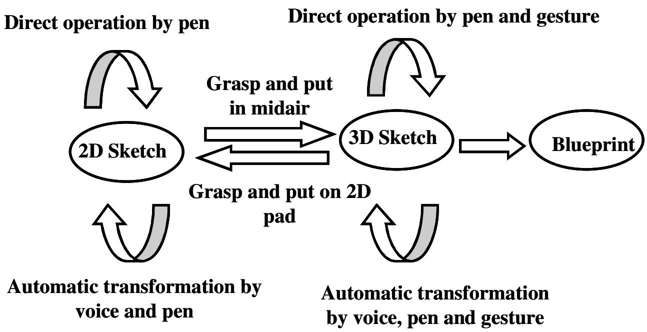
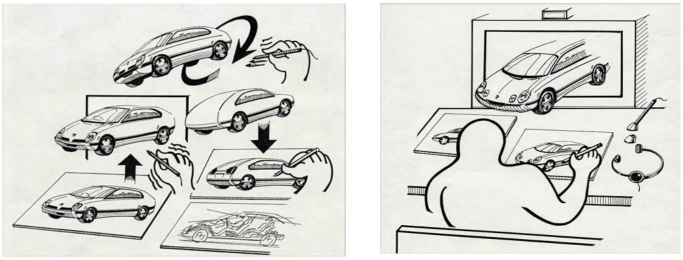


Fig. 2. Typical design flow



(a) Design Flow (b) Design Environment

Fig. 3. Image of design flow and environment

3.2.2 Prototype System

Fig. 4 is a photo of the Godzilla system. Fig. 5 shows examples of car design using Godzilla. Note that a 3D image is displayed in midair just in front of the 3D-pad. After we developed the Godzilla system, we took it to the design division of the Toyota automobile company for the initial evaluation.

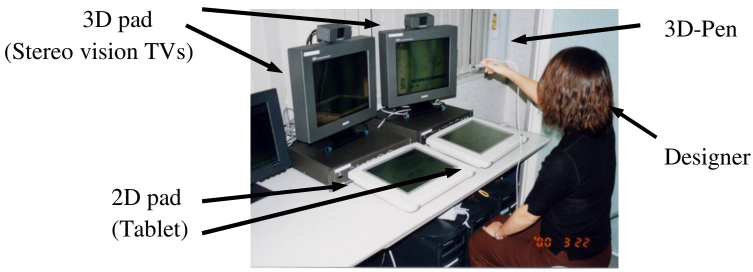


Fig. 4. Prototype system "Godzilla"

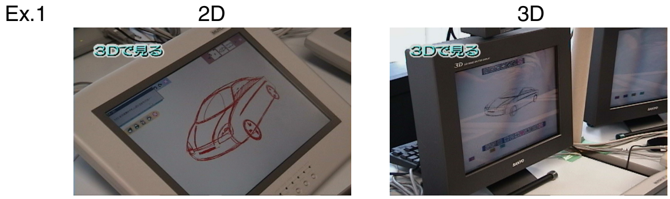


Fig. 5. Examples

3.3 Our 3D System-2: "Extended Godzilla"

Godzilla has problems: (1) The range of forms it can handle is restricted to car-like forms, and (2) separate 2D and 3D monitors feel unnatural.

3.3.1 Free Form Design Using a Combination of Seven Primitives

It is obviously impossible to recognize a 2D sketch of a 3D form without any knowledge of the sketched form since a 2D sketch cannot retain all of the shape information. To enable free-form design, we developed a design approach that enables the designer to draw primitive forms and then combine and modify them.

Our system has seven primitives: triangular pyramid, square pyramid, triangular prism, square prism, cone, cylinder, and sphere. A typical design flow is shown in Fig. 6. First, the designer sketches the primitive shapes, and the system recognizes them. The designer then combines and modifies the primitives or views and checks the shapes in the 3D space.

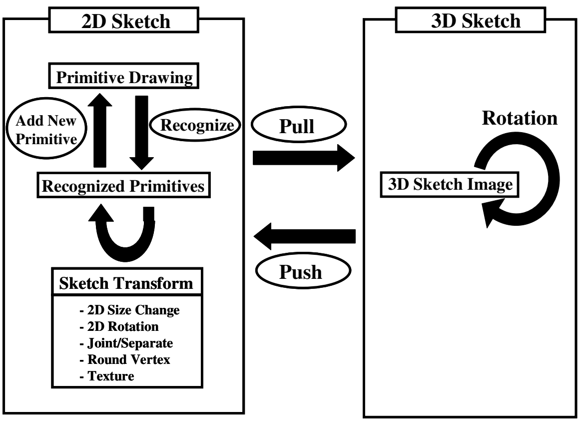


Fig. 6. Design flow in "Extended Godzilla" system

3.3.2 Natural Seamless 2D-3D Display

To provide a natural display, the display unit has to support the seamless 2D-3D transition. To meet these requirements, we used an LCD monitor with polarized light screens (a "micro-pole filter") and polarized glass (see Fig. 7).

Fig. 7. "Extended Godzilla" System

The only operation in 2D-3D space is "pull and push." As shown in Fig. 8, when the designer "pulls" an image in the 2D space, the image is gradually raised from the surface of the LCD. Conversely, when the designer "pushes" an image in the 3D space, the image gradually sinks into the LCD.

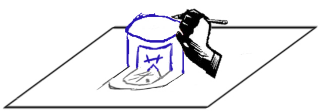


Fig. 8. "Pull and Push"

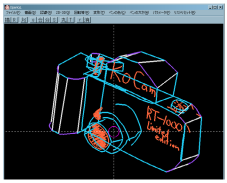


Fig. 9. Example

Representative sketches are shown in Fig. 9. Note that the images can be viewed from any angle since the sketches have a 3D structure.

3.4 Our 3D System 3: Rich Visual Feedback

Sketching in 3D space is difficult, particularly because the senses of depth and balance are poor. This indicates that a user interface is needed to compensate for the difficulties.

For the compensation, the interface needs to enhance the user's awareness of errors.

We found that using the metaphors of "shadow" and "hand mirror" effectively achieves this, as illustrated in Fig. 10. Note that the shadow and mirror are also 3D images.

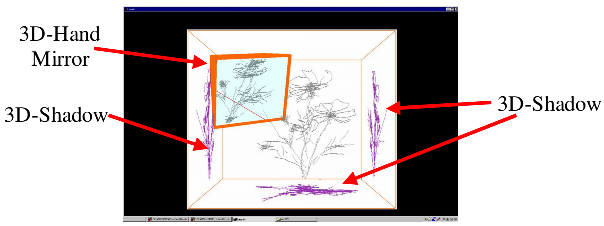


Fig. 10. Use of "Shadow" and "Hand Mirror"

3.5 Our 3D System-4: Force Feedback

System 4 tried to solve the same problem as System 3 by providing the designer with the virtual force field (virtual surface) in midair.

3.5.1 Approach

We devised four types of force field to cope with the problem. They are characterized in Fig. 11.

(a) Uniform force field: The uniform inertia is produced in the drawing area. The designers feel as if they were drawing in jelly.

(b) Automatic generation of surface: The artificial surface is automatically generated in accordance with the drawing. When many lines are judged on the same plain, the force field is automatically generated to shape the plain.

(c) User-defined surface: The user can indicate the needed surface by gesturing.

(d) Use of 3D rulers: We designed several virtual 3D rulers, which include the virtual sphere, the virtual rod with ditch (for drawing the straight line in 3D), the virtual 3D French curve, and so on. They are virtual in the sense that they are not real objects. The surface is artificially generated at the position of 3D sensor, so they are easily fixed in 3D.

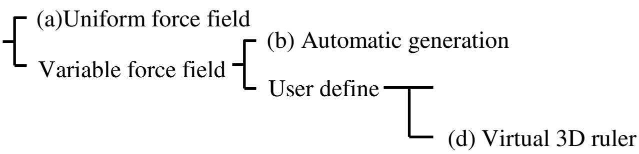
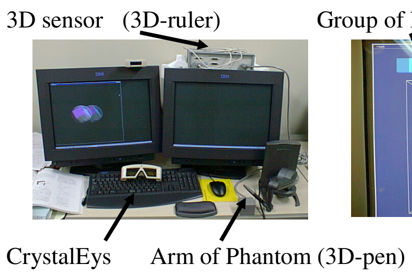


Fig. 11. Categories of four force fields

3.5.2 Prototype System

We have developed a prototype system (Fig. 12). The user has a 3D-sensor (for a virtual 3D-ruler) in his/her left hand and an arm of the Phantom (the force feedback device) in his/her right hand for drawing in 3D. Fig. 13 shows the automatic generation of the surface. Note that it is seen in double since the left-eye and right-eye images are displayed.



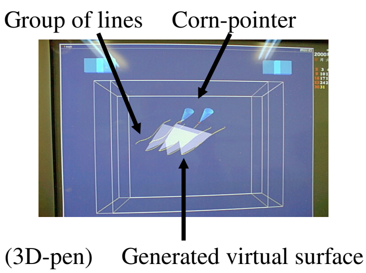


Fig. 12. Photo of prototype system **Fig. 13.** User interface

4 Long-Term Evaluation

4.1 Short-Term Evaluation Was Promising

For each system, we conducted a short-term user test that showed our user interface was promising. All designers welcomed the novel interaction.

4.2 Long-Term Evaluation: 3D Space Is Useless

However, they stop using the 3D space after a while. In the following subsection, the reason for this is analyzed in terms of "view in 3D" and "draw in 3D".

4.2.1 Analysis on "View in 3D"

While using Godzilla (System 1), the users draw a 2D sketch and lift it up into midair. Then the 3D sketch is displayed, fascinating the users. Sometimes the user is surprised at the function. Nevertheless, sooner or later, they notice that they do not need to look at it in the 3D space as a 3D sketch. They can look at it on the 2D-pad and can rotate it in a similar manner. The difference is whether they are displayed in 3D (stereoscopic) or in semi-3D on the 2D plane (perspective 2D image).

The same phenomena were also found while using Extended Godzilla (System 2). The users draw a 2D sketch and lift it up into midair gradually. Because the sketch seamlessly transitions from 2D to 3D, it surprises and fascinates all users. However, after a while, they bore of the seamless 2D-3D transition.

4.2.2 Analysis on Drawing in 3D

While using System 3, the designer can draw a beautiful 3D flower image directly as shown in Fig. 10 by utilizing the effect of visual feedback. Also, while using System 4, a designer can draw a image stably by utilizing the effect of the virtual surface in midair.

However, even though the designer is assisted by the visual feedback and the force feedback, 2D drawing is much easier than 3D drawing. Therefore, the designers tend to move to System 2 ("Extended Godzilla") as long as System 2 can handle the target image, so eventually they use 2D draw and 2D view as mentioned above.

4.3 Lack of Indispensability of 3D Space

Essentially, the 3D space is not for a designer in terms of either "view in 3D" and "draw in 3D".

The reason designers stop using the 3D space is that they can do their work without it. In other words, our systems do not provide the designers with an indispensable function that truly needs 3D space.

5 New Design Concept

As long as an indispensable function of 3D space is not found, it is pointless to develop a support system that uses 3D space. Therefore, we stopped our research into 3D space for a few years. Recently, we found indispensable functions and restarted the research activity. Here, the latest findings and the current prototype system are shown.

5.1 Indispensable Functions of 3D Space

We found two indispensable functions that need the 3D space. The first one is a life-sized 3D sketch. If the 3D sketch is life-sized, the user evaluates the size to compare their own body and the 3D sketch shown in the midair in front of the user.

If the “life-sized” nature is missing, the users cannot evaluate it on the basis of comparison with their body, so the necessity of 3D sketch is lost.

The second one is a 3D sketch that must be operable by the user. The user should be able to operate the 3D sketch, that is, touch, push, move, and so on. If the 3D sketch is operable, the user evaluates the ease of use by operating while stooping down, extending a hand, or twisting his/her body.

5.2 Design Concept “Life-Sized and Operable”

The 3D sketches must have a “life-sized and operable” nature. Since the life-sized 3D sketch can be evaluated by comparison with a user’s body, it needs to be displayed in a 3D space in front of the user. For example, the user can notice that the table of this kitchen is low or the emergency button is far from the operator’s chair.

Similarly, since users can evaluate the operational 3D sketch by moving their bodies while operating, it needs to be displayed in a 3D space in front of the users. For example, the user can notice when operating a lever that the warning lamp is hard to see or the tray of the copy machine is too low to remove the paper jam.

6 New 3D Sketch System with Mixed Reality

6.1 New Design Flow

Fig. 14 shows the new design process extended by the new principle. The design flow is explained briefly by using the copy machine design example (see also Fig. 15).

In the first step, the designer is thinking of a shape of a copy machine, asking him or herself, “What’s a smart design for a copy machine?” while drawing the idea in life-size and ubiquitously.

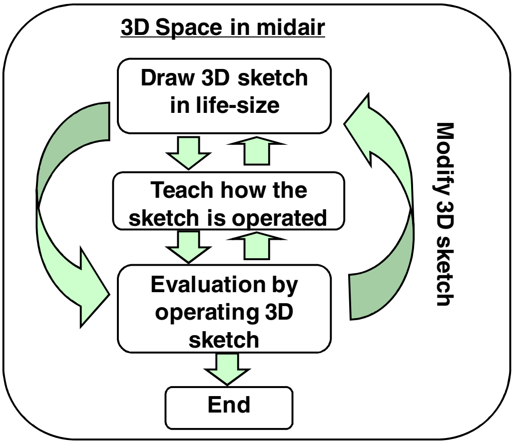


Fig. 14. New design process

Second, the designer is thinking of the operation of the sketch, thinking to him or herself, "The tray will move in this direction. If I push this button, the paper is ejected.", then formulating the operation rules by grasping and moving the sketch shown in 3D.

Third, the designer checks the usability by operating the sketch while sitting down, stooping down, extending an arm, and so on.

Then the designer may find that a button is hard to push because it is inconveniently located, the tray is hard to pull out because you have to get into an uncomfortable position, and so on. The designer simply erases the 3D sketch and redraws it.

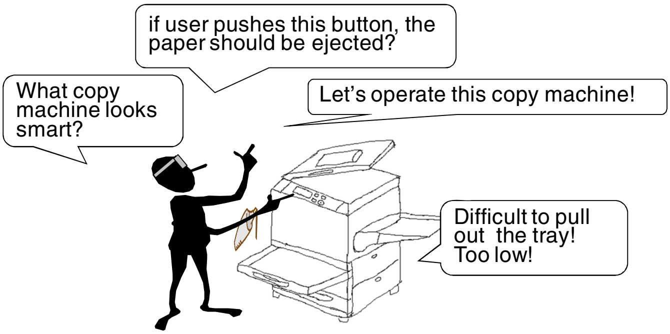


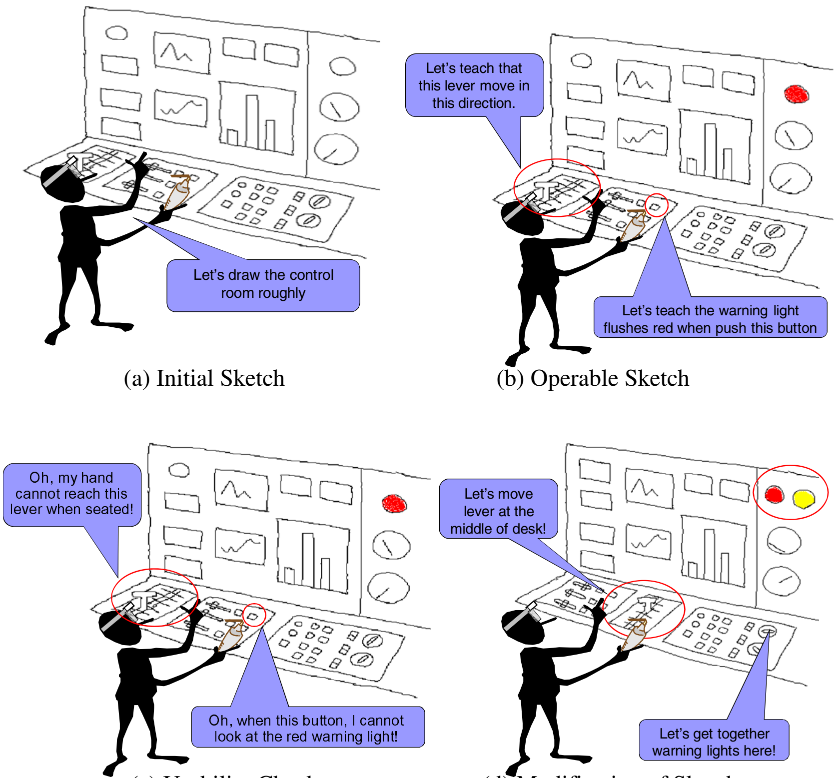
Fig. 15. Copy machine design example

6.2 Examples of Design Process

Fig. 16 shows our other application field, i.e., the control room design. It is very important to design a usable control room from the safety point of view. It is impossible to make a mock-up system due to the exorbitant costs. In a real design department, the designers will design it virtually by using the 3D-CAD system. During the design process, the designers cannot fully appreciate the size and operability. All they can do is to imagine the size and the operation scene in their mind.

By using our latest prototype system, the designer starts to draw the control room roughly (Fig.16 (a)). Then the designer teaches the system the operation of levers, buttons, and warning lights (Fig.16 (b)). The designer finds several problems (Fig.16 (c)) and fixes them by redrawing the sketch (Fig.16 (d)).

Note that all this is done by hand-drawn sketches. The scenario can be done in less than one hour.



(a) Initial Sketch (b) Operable Sketch

(c) Usability Check (d) Modification of Sketch

Fig. 16. Control room design example

6.3 Special User Interaction for the New Design Concept

The new design concept "life-size and operability" requires new interaction methods.

6.3.1 Interactions for "Life-Size"

Since it is difficult to draw a life-size image, our system has several unique interactions. Here three interactions are shown.

The color of the drawing line changes at the fixed length (50cm) as shown Fig. 17. The designer can recognize the length of the line while drawing the lines.

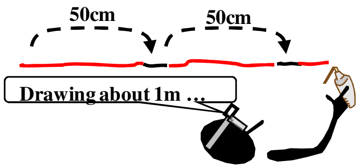


Fig. 17. Line length notification

Fig. 18 shows the lead line. It is a kind of auxiliary line. The designer is drawing a refrigerator (red box) between two real drawers. The lead line-A connects the designer's body (left hand) and the virtual image (refrigerator: red box). If the designer moves his/her left hand, the lead line-A moves too. The lead line-B connects two real objects. It is literally a static auxiliary line drawn in the real world. The lead line-C connects the designer's body (waist) and the real object (right drawer).

These lead lines help the designer grasp the 3D location of real and virtual objects.

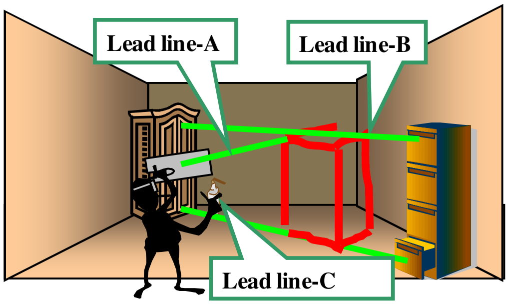


Fig. 18. Three types of lead line

Since the drawings are life-sized, they tend to be big. Therefore, the selection of the object is difficult. Our system has a few selection methods. Fig. 19 shows the projection based selection. The designer draw a closed line in front of him/her, the closed line is projected conically from the center of the designer's eye. The drawings included in the projected volume are selected. The closed line can be drawn in the different space. Then the merged volume is the selected zone as shown Fig. 19.

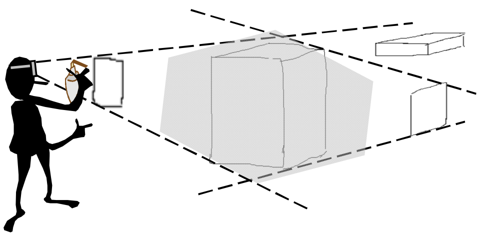


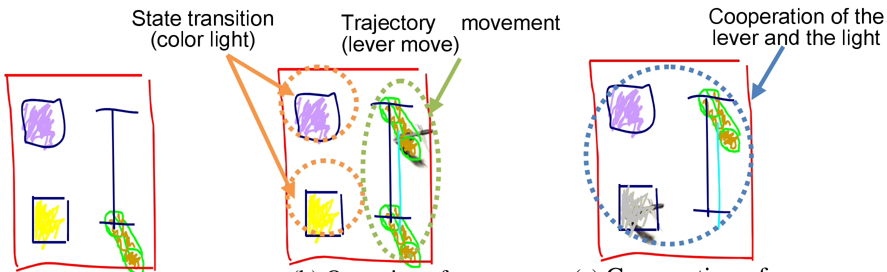
Fig. 19. Selection by projection

6.3.2 Interactions for "Operability"

The designer teaches the part of sketch how to respond to the users' operation (behavior) in three steps. An example (Fig.20) shows how to teach the operability of the control board, which consists of two color lights and one lever. The operability of the control board as follows;

- Lever moves vertically
- Light changes color in accordance with the lever's position

In the first step, the designer draws the overview of the control panel with two color lights and one lever (Fig.20 (a)). In the second step, the designer teaches the operation of single component, i.e., the state transition (ex. On/OFF switch, color light) and the trajectory movement (ex. lever (vertical move), dial (cursive move)). For example, in Fig.20 (b), the designer teaches the trajectory movement of the lever by selecting the sketch of lever and moving it vertically. In the last step, the designer teaches the cooperation of multiple components. In Fig.20 (c), the designer teaches the cooperation of the lever and the light by moving the lever at the terminal point and then switching the light to the preferred color.



(a) Initial Sketch (b) Operation of single component (c) Cooperation of multiple components

Fig. 20. New interactions for "operability"

6.4 Current Prototype System

The current prototype system consists of a see-through HMD, head-tracker, 3D-pen, and palette (command board) as shown in Fig. 21. The 3D position sense is implemented by combining the ultrasonic and magnetic sensor to handle a large 3D sketch, such as a control room. To promote the cooperative design, three or more HMDs (maximum 6) are connected.

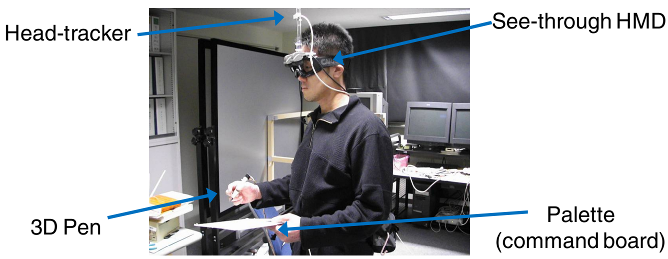
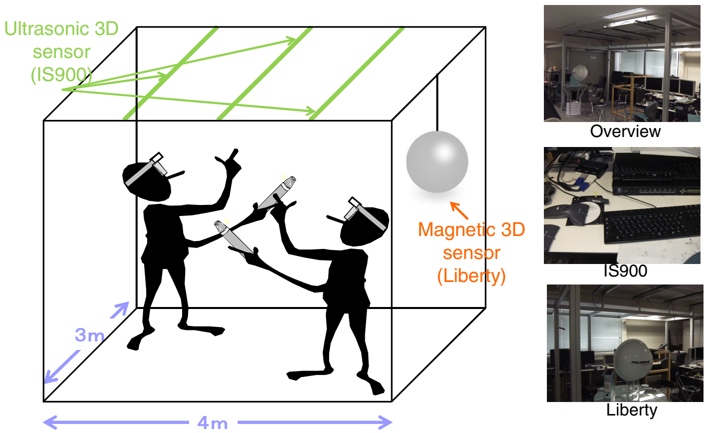


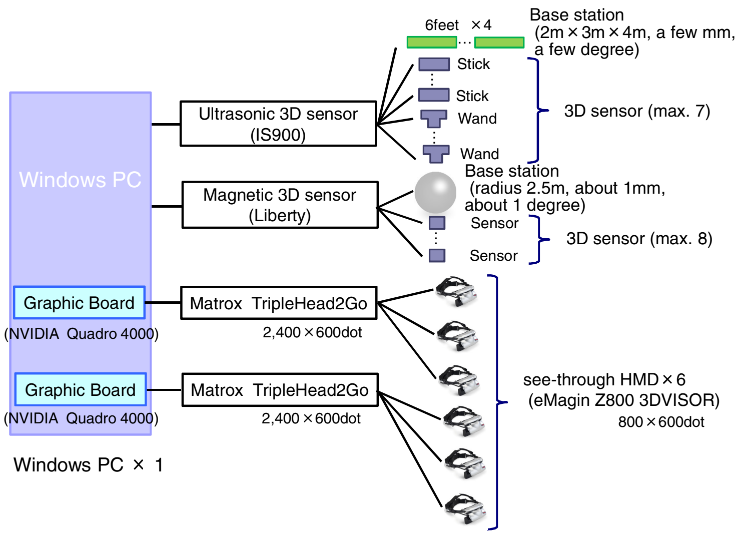
Fig. 21. Current prototype

The prototype system has 3m x 4m open space as the cooperative design space (Fig.22 (a)). Two types of 3D position sensor are equipped on the ceiling (ultrasonic sensor) and the edge of the wall (magnetic sensor) to provide the users with the spacious open area.

The current prototype system is implemented by single PC architecture shown in Fig.22 (b). The main PC has two graphic boards (Quadro 4000) which have 2400 x 600 dot screen each. The 2400 x 600 dot screen is divided into 3 800x600 dot screens by Matrox TripleHead2GO which are connected to the see-through 3D-HMDs (eMagin Z800). This simple architecture realized 6 users environment by the single computer.



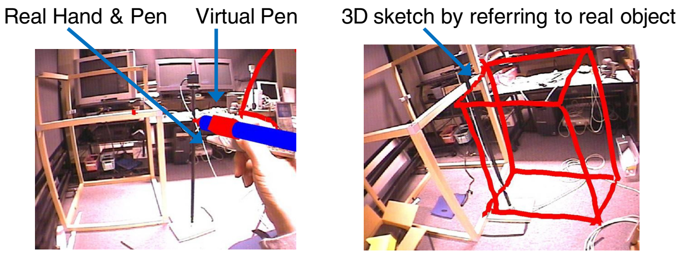
(a)Design Space (6 designers, 3m x 4m)



(b) System Architecture

Fig. 22. System overview

Fig. 23 shows examples of what the users view through the HMD. As shown in Fig. 23(a), a virtual pen is displayed on the user's real world pen. Fig. 23 (b) shows an example in which the designer draws a 3D sketch by referring to a real object's size.



(a) Virtual Pen (b) Object Drawing

Fig. 23. User's view through HMD

6.5 Preliminary Evaluation

We conducted a preliminary evaluation with the help of professional designers. Fig. 24 and Fig. 25 are the photos of the experiment.

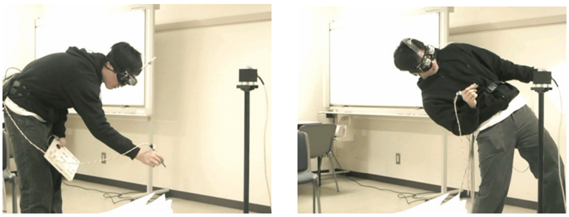


Fig. 24. Behavior of subject-A

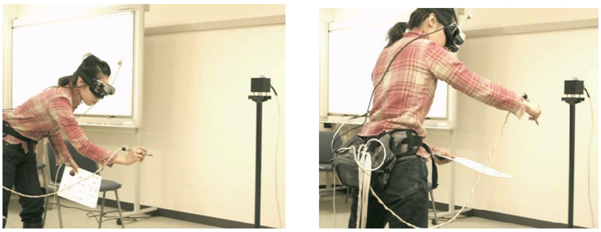


Fig. 25. Behavior of professional designer

The subjects were 5 students and 1 professional designer. All of the subjects took a training course that covered all interaction methods in 17 scenarios. The preliminary evaluation took place after they completed all the scenarios.

In the preliminary evaluation, all subject were given the design theme concerning to the new copy machine, then they began to design freely.

We videotaped the design process of 6 subjects and analyzed the specific behaviors.

Table 1 shows the frequency of the new interactions for "life-size", proposed in the section 6.3.1. All subjects have used all interactions and frequently used the selection by projection. The prototype system had several methods for the selection since it was difficult to select the part of the sketch that floated in the midair. For example, the prototype system had the direct selection method in which the user touched one line by one line and the touched lines were selected, the volumetric selection method in which the user defined the rectangular solid by drawing the diagonal line, and the lines in the volume were selected, and so on. Because of the "life size" nature, the selection by projection seemed to be preferred.

Table 1. Behavior of "life-size" interaction

	# of Line length notification	# of Lead line	# of Selection by projection	# of Volumetric selection
Subject-A	6	2	9	5
Subject-B	4	2	4	7
Subject-C	1	2	7	3
Subject-D	6	3	6	4
Subject-E	7	1	12	6
Designer	4	2	7	1
Average	4.7	2	7.5	4.3

Table 2 shows the behavior of the "operable" interaction. "# of operable sketch" does not mean the number of the operable sketches which remained in the final design but the number of the operable sketches which were drawn in all the design process (i.e., includes the try and error sketches). All subjects drew many operable sketches. Since one co-operable sketch consists of at least two single component, so the average number 5.5 means that at least 11 components are used in the co-operation. Surprisingly, almost all single components worked with other components. Also the number of operations 16.0 was so high that they seemed to focus on the usability evaluation.

Table 2. Behavior of "operable" interaction

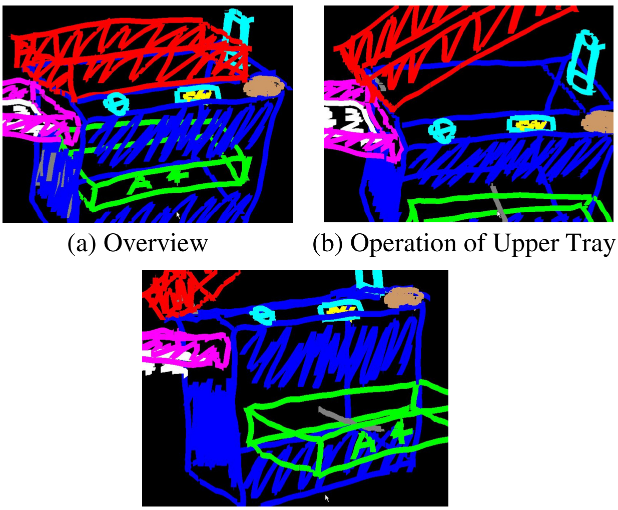
	# of Operable sketch (single component)	# of Co-operable sketch (multiple components)	# of User operation (behavior)
Subject-A	16	7	13
Subject-B	8	2	12
Subject-C	10	4	9
Subject-D	12	8	15
Subject-E	22	6	21
Designer	11	6	26
Average	13.1	5.5	16

Table 3 lists several statistics on the final design. They tend to use many colors that were used to show the different parts. The final design contained 6.8 operable components and 2.5 co-operable rules in average. They drew more operable sketch than we expected. Only half of the sketch components remained in the final design. It might be explained by the trial and error approach in designing of the interaction.

Table 3. Statistics of final design

	# of Colors	# of Operable sketch (single component)	# of Co-operable sketch (multiple components)
Subject-A	7	10	3
Subject-B	6	6	2
Subject-C	6	4	2
Subject-D	11	9	4
Subject-E	7	6	3
Designer	8	6	3
Average	7.5	6.8	2.5

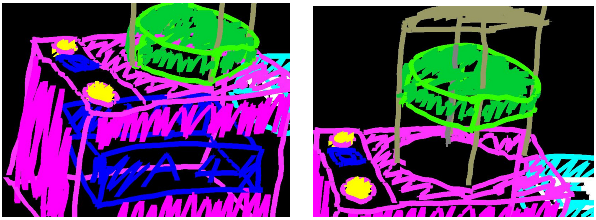
Fig. 26 and Fig. 27 are the results of the design. Note that the designs are real-sized and operable. Fig.27 is the design result when the designer was asked to design a novel copy machine such as "never before seen copy machine". They successfully designed a "novel operable" copy machine, which was a round copy machine usable from any direction. Note that the round tray is also operable.



(a) Overview (b) Operation of Upper Tray

(c) Operation of Paper Tray

Fig. 26. Design result-A

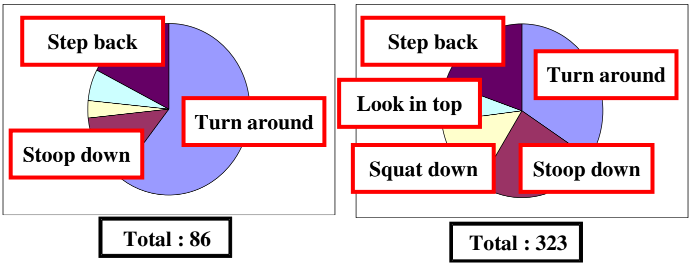


(a) Overview (b) Operation of Upper Tray

(c) Operation of Paper Tray

Fig. 27. Design result-B

Fig. 28 is the analysis of the professional designer's behavior while operating the 3D sketch. Fig. 28 implies that the designer willingly drew the life-seized sketch and frequently operated the 3D sketch to check the usability of the copy machine.



(a) Analysis-A (b) Analysis-B

Fig. 28. Analysis of professional designer's behavior

7 Conclusion

This research began from the serious realization that 3D sketch systems were not used by professional designer in real design fields. They are treated as if they were just mere attractions in an amusement park.

In this paper, we proposed a new design concept, "life-size and operability", which should make the 3D sketch system truly valuable for the designer.

According to the preliminary evaluation, the design concept seems promising. However, its validity has not yet been completely proven. We are now obtaining the data by applying it to real world design.

We hope our research is helpful to design the multimedia-based support system for professional use.

References

1. Dzulkhiflee, M., Tano, S., Iwata, M., Hashiyama, T.: Effectiveness of Annotating by Hand for non-Alphabetical Languages. In: CHI 2006, pp. 841–850 (2006)
2. Iwata, M., Sasaki, Y., Tano, S., Hashiyama, T., Ichino, J.: A Sketch Support System Based on Behavior of Designers. In: IEEE International Conference on Progress in Informatics and Computing-2010, pp. 1298–1304 (2010)
3. Ichino, J., Makita, T., Tano, S., Hashiyama, T.: Support for seamless linkage between less-detailed and more-detailed representations for comic design. In: CHI 2009, pp. 3979–3984 (2009)
4. Tano, Kamura, Iwata, Hashiyama: Digital Paper Concept for Reflective Writing by Seamless Traverse between Handwritten and Coded Information. In: HCI International 2005, E-book (2005)
5. Tano, S.: Quantitative Study on the Effectiveness of Pen-based Computing on Experiential and Reflective Cognitive Modes. Mobile Computing in Education, 46–49 (2009)
6. Tano, Kodera, Nakashima, Kawano, Nakanish, Hamagishi, Inoue, Watanabe, Okamoto, Kawagoe, Kaneko, Hotta, Tatsuoka: Godzilla: Seamless 2D and 3D Sketch Environment for Reflective and Creative Design Work. In: INTERACT 2003, pp. 131–138 (2003)
7. Tano, S., Komatsu, Y., Iwata, M.: Extended Godzilla: Free-form 3D-object Design by Sketching and Modifying Seven Primitives at Single 2D-3D Seamless Display. In: Masoodian, M., Jones, S., Rogers, B. (eds.) APCHI 2004. LNCS, vol. 3101, pp. 471–480. Springer, Heidelberg (2004)
8. Tano, Sugimoto: Natural Hand Writing in Unstable 3D space with Artificial Surface. In: CHI 2001, pp. 353–354 (2001)
9. Tano, S., Matsumoto, T., Iwata, M.: Quantitative Analysis of Human Behavior and Implied User Interface in 3D Sketching. In: Masoodian, M., Jones, S., Rogers, B. (eds.) APCHI 2004. LNCS, vol. 3101, pp. 481–490. Springer, Heidelberg (2004)
10. Wayne, et al.: Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry. In: Eurographics 2003, pp. 19–28 (2003)
11. Steven, et al.: Surface Drawing: Creating Organic 3D Shapes with the Hand and Tangible Tools. In: CHI 2001, pp. 261–268 (2001)
12. Gerols, et al.: Free Drawer: A Free-Form Sketching System on the Responsive Workbench. In: VRST 2001, pp. 167–174 (2001)
13. Daniel, et al.: A Fully Immersive 3D Artistic Medium and Interactive Experience. In: Proceedings 2001 ACM Symposium on Interactive 3D Graphics, pp. 85–93 (2001)
14. Tovi, et al.: Creating Principal 3D Curves with Digital Tape Drawing. In: CHI 2002, pp. 121–128 (2002)
15. Shin, H., Igarashi, T.: Magic canvas: interactive design of a 3-D scene prototype from freehand sketches. In: GI 2007, pp. 63–70 (2007)
16. Olsen, L., Samavati, F.F.: Stroke extraction and classification for mesh inflation. In: Proc. of SBIM 2010, pp. 9–16 (2010)
17. Perkunder, K., Israel, J.H., Alexa, M.: Shape modeling with sketched feature lines in immersive 3D environments. In: Proc. of SBIM 2010, pp. 127–134 (2010)
18. Tano, S., Yamamoto, M., Dzulkhiflee, M., Ichino, J., Hashiyama, T., Iwata, M.: Three Design Principles Learned through Developing a Series of 3D Sketch Systems: "Memory Capacity", "Cognitive Mode", and "Life-size and Operability". In: IEEE SMC 2012, pp. 880–887 (2012)

A One-Handed Multi-touch Method for 3D Rotations

Douglas Scheurich and Wolfgang Stuerzlinger

York University, Toronto, Canada
{dougsch,wolfgang}@cse.yorku.ca

Abstract. Rotating 3D objects is a difficult task. We present a new rotation technique based on collision-free "mating" to expedite 3D rotations. It is specifically designed for one-handed interaction on tablets or touchscreens. A user study found that our new technique decreased the time to rotate objects in 3D by more than 60% in situations where objects align. We found similar results when users translated and rotated objects in a 3D scene. Also, angle errors were 35% less with mating. In essence, our new rotation technique improves both the speed and accuracy of common 3D rotation tasks.

Keywords: 3D rotations, 3D user interfaces, multi-touch, tablets.

1 Introduction

The most common interaction tasks in three dimensional, 3D, virtual environments are navigation, object selection, and manipulation, such as translation and rotation. Object translation positions objects within the scene, whereas rotations orient objects. Despite being a fundamental task, there is no established standard for rotating 3D objects. One issue is that there is no "best" input device for 3D manipulation. In practice users use two-dimensional, 2D, pointing devices, including the mouse and touchscreens. 2D pointing devices offer good control of 2 degrees of freedom (DOF). However and in a 3D environment, control over 3 DOF is required for translations or rotations (yaw, pitch, and roll), or 6 DOF for both simultaneously. In many user interfaces this is handled through combinations of different widgets or touch gestures, i.e. through a combination of 2 DOF and 1 DOF controls. A mouse button is often assigned to control 2 DOF rotations. The third DOF is typically controlled via a modifier or the scroll wheel.

The computer-aided design program Solidworks recently introduced a simple form of object mating. There, clicking on a specific surface of an object followed by a click on another surface snaps these two together, so that the first surface "mates" onto the second. However, this simple mating technique may leave objects in positions where they interpenetrate.

Mating is a simple way to orient objects. Thus, we were surprised to discover that there was no documented work on mating methods that avoided or resolved

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 56–69, 2013.

collisions. This encouraged us to explore the idea of mating for rotating objects on a touchscreen interface while avoiding object interpenetration.

2 Previous Work

Relevant other work on 3D rotations uses either 2D or 3D input devices. For 2D devices we first discuss the mouse and then touchscreen methods.

With a mouse, Bade et al. [1] evaluated four different methods, Bell's [2] and Shoemake's [3] virtual trackballs, and two variants of the Two-Axis Valuator [4], in a user study. They found the Two-Axis Valuator to be best. Jacob et al. [8] investigated inspection tasks requiring rotations and found a similar result. However, the tasks in both of these studies required only 2DOF rotation control! Again, investigating only a subset of all 3D rotations, Partala [7] found virtual trackballs to be superior. Zhao et al. [5] recently investigated tasks that require full 3D rotation control. They did not identify significant differences between Bell's and Shoemake's trackballs and the Two-Axis Valuator.

Reisman et al. [9] presented a multi-touch method to control the position and rotation of 3D objects. The solver-based method tries to keep the object stable under the fingers. Yet, the result is not always predictable and rotations are often limited to 90 degrees in two of three directions (e.g. when a cube is facing the viewer). Rotations then require clutching. Martinet et al. [11] used this method in their 6 DOF manipulation system. Hancock et al.'s [10] technique permits 1 DOF rotation with a 2-finger "rotate" gesture and controls the other 2 DOF with a two-handed 2+1 finger gesture. Kin et al. [12] controls 2 DOF rotations with a single finger and the third DOF with a two handed gesture.

Bier [19] developed a system called Gargoyle3D based on the idea of "snap-dragging". This system uses a general-purpose gravity system combined with snapping in 3D to enable users to accurately position and orient objects in 3D. In the user interface Gargoyle3D offered more than 40 distinct commands to achieve these operations. This level of complexity makes it infeasible for tablet computing. Related work on SKETCH by Zeleznik et al. [20] requires at a minimum a three-button mouse and a keyboard for a modifier. While less complex than Bier's system, the sheer number of mouse and keyboard combinations again does not lend this idea to a simple or efficient implementation on a tablet.

Zhai et al. [13] and Boritz et al. [14] investigated 3D rotations as part of 6 DOF docking tasks. Hinckley et al. [6] compared 3D rotation techniques using 2D and 3D input devices and found that those based on 3D input devices were approximately a third faster. Recently Kratz et al. [15] proposed a system where the orientation of a user's hand controls 3D rotations. Their comparison with a virtual trackball found also a ~30% improvement. Liu et al. [21] proposed a system for full 6 DOF manipulation using four distinct one or two-finger touch gestures.

However, the authors note that their system is not well suited for fine control of model transformations.

3 A New Multi-touch 3D Rotation Technique

Tablets and tablet use is now a permanent addition to computing, considering the sustained explosive growth of the tablet industry in the last few years. Especially since the advent of smaller tablet form factors, such as the 7.9” iPad mini, one cannot assume that a tablet is going to be used with two hands simultaneously. Such devices are used almost anywhere by holding it in one hand while interacting with the system with the other. A great example is a video on the Autodesk website showcasing their Inventor Publisher Mobile Viewer[1] application for iOS and Android. Parts of these videos show an engineer working on a tablet. The engineer never once puts the tablet on a table or desk to work, and holds it with one hand while working with the other the entire time. Given this constraint, we noticed a lack of touch-based systems that require only a single hand for 3D manipulation. This also means that all previously presented, touch-based 3D rotation methods, which require two hands, are not ideally suited for tablet use.

We designed our new technique explicitly for one-handed use. We also wanted to avoid the unpredictable nature and the limitations of Reisman’s [9] approach. Moreover, we were inspired by a recent set of observations on 3D user interface design by Stuerzlinger et al. [16]. They point out that in the real world the vast majority of objects are aligned with planes or other objects. The reason is that (almost) all objects are in contact with others on our planet and “floating” objects are a rare exception. Tables usually stand on floors; pictures are attached to walls; light fixtures to the ceiling. Many such objects have only one free rotational DOF in their “normal” placement. In other words, truly random orientations are the exception in the real world. Consequently, we focus on user interfaces that are optimized for these pervasive cases and design our technique accordingly.

The idea of “mating” two surfaces fits this observation very well, except that naïve mating easily results in object interpenetration. Mating the seat of a chair “onto” the ground would put the backrest of the chair into the ground, which novices often find confusing [16]. This is demonstrated in Figure 1 below. Therefore, we enhance basic mating by always putting the chair into a position that avoids collisions, while keeping the seat plane parallel to the ground. As an added bonus mating also translates the object, which may lead to additional timesaving. Given that our enhanced form of mating always put objects into contact, we globally prevent objects from “floating” in our system. This limits the system to 2 DOF positioning, but also matches the capabilities of touchscreens better as fewer DOFs need to be controlled and simplifies the user interface. Objects can still assume any 3D rotation in our system, which is the main focus of our work here.

---

[1] `http://m.autodesk.com/mobile/servlet/product?siteID=17221380&id=17774143`. This website is accurate and available as of January 20th, 2013.

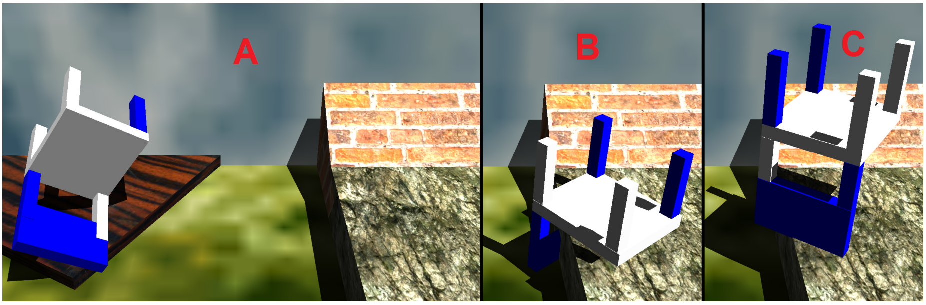


Fig. 1. The original orientation is shown in panel A. The user seeks to mate the seat of the chair from the left to the stone ledge on the right. Panel B illustrates the result of mating without preventing collisions. Such an operation leaves the chair in a confusing position, where the plane of the seat of the chair is parallel with the plane of the stone. However, the chair interpenetrates the floor. Panel C demonstrates mating with our collision resolution. The plane of the chair seat is still parallel with the stone surface. Yet, the chair has been "raised" so that it is in a collision free position.

3.1 Two Multi-touch Rotation Methods

As discussed above, previous research did not identify a clearly superior 3D rotation technique, especially one with a limited number of input controls suitable for deployment on a tablet. Hence, we base our multi-touch system on the Two-Axis Valuator to directly control 2 DOF rotations. To control the third DOF we use a different form of multi-touch gesture compared to previous work. We implemented two variations for this. One is designed for systems that permit only 3D rotations, the other for systems that support both positioning and rotation.

3.2 Rotation Only Method

Our first method targets 3D rotations only. It interprets a single finger drag as Two-Axis Valuator manipulation. This is usually done with the index finger. Dragging the finger left and right will rotate the object about the *y axis* while dragging the finger up and down will rotate about the *x axis* of the camera coordinate system. A two-finger touch rotates around the view direction (i.e. the *z axis* of the camera coordinate system, see Figure 3 below). Here we implement a new technique: if one finger stays in place and a second "scrolls" this is interpreted as a rotation. We found users preferred to accomplish this in one of two ways. When rotating objects in the middle of the screen, putting the index finger down and flicking the thumb left or right below it is a natural way to access this functionality. When operating on the right edge of the screen, placing the index finger down and scrolling with the middle finger seems most natural. Placing the middle finger down and scrolling with the index works well when working near the left edge of the screen. To summarize, a single finger drag rotates

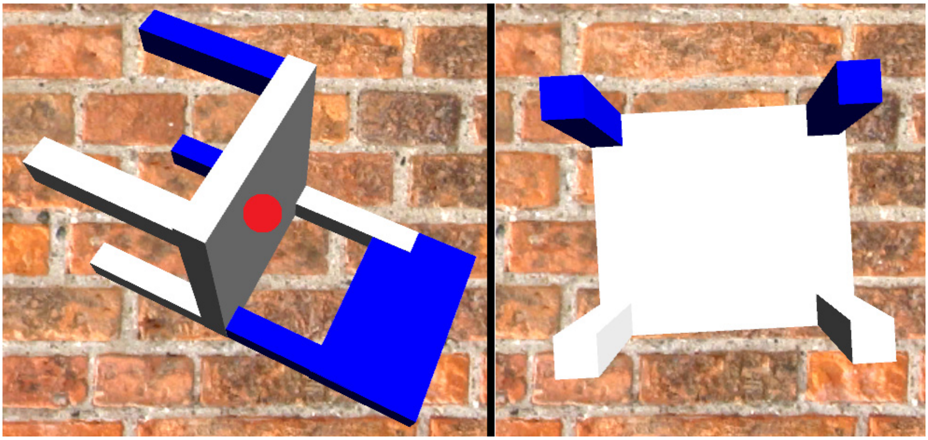


Fig. 2. Illustrating of mating for 3D rotations using method 1. In this image we can see the brick wall as a background object behind the chair. By double tapping on the seat of the chair, marked with the red dot, the object's surface that corresponds with the double tap location (i.e. the chair seat), will become aligned with the plane of the object behind it (i.e. the brick wall). Consequently and by double tapping the seat of the chair, the chair is automatically rotated so that the seat plane is parallel with the background surface. This figure also illustrates a trial in the first phase of the experiment. The user is able to manipulate the object on the left. The task is to match the pose of the left object with that of the right object.

about the *x* and *y axes*, while a single finger anchored in place (i.e. index finger) plus a finger dragging (i.e. the thumb) rotates the about the *z axis*. Double tapping a point on the object will use the enhanced mate functionality to mate the specified surface of the object with the plane behind it, as illustrated in Figure 2 below.

3.3 Rotation and Translation Method

Our second method targets both rotations and translations. As a result this method is more targeted at "real world" usage compared to the first method, since in an interactive environment some (or many) objects may be moveable. In the second method single finger movements control the (constrained) translation of an object along the surfaces of the scene using a variant of the technique presented by Oh et al. [17]. Basically, objects always snap to and/or slide on the surface beneath them. For example, imagine wanting to drag a block around a staircase. If you are dragging the block down the staircase, as you move your finger the block will slide across the surface of each step, and when it reaches the end of a step, it will snap from that step' surface onto the surface of the next step and continue to slide along that, etc.

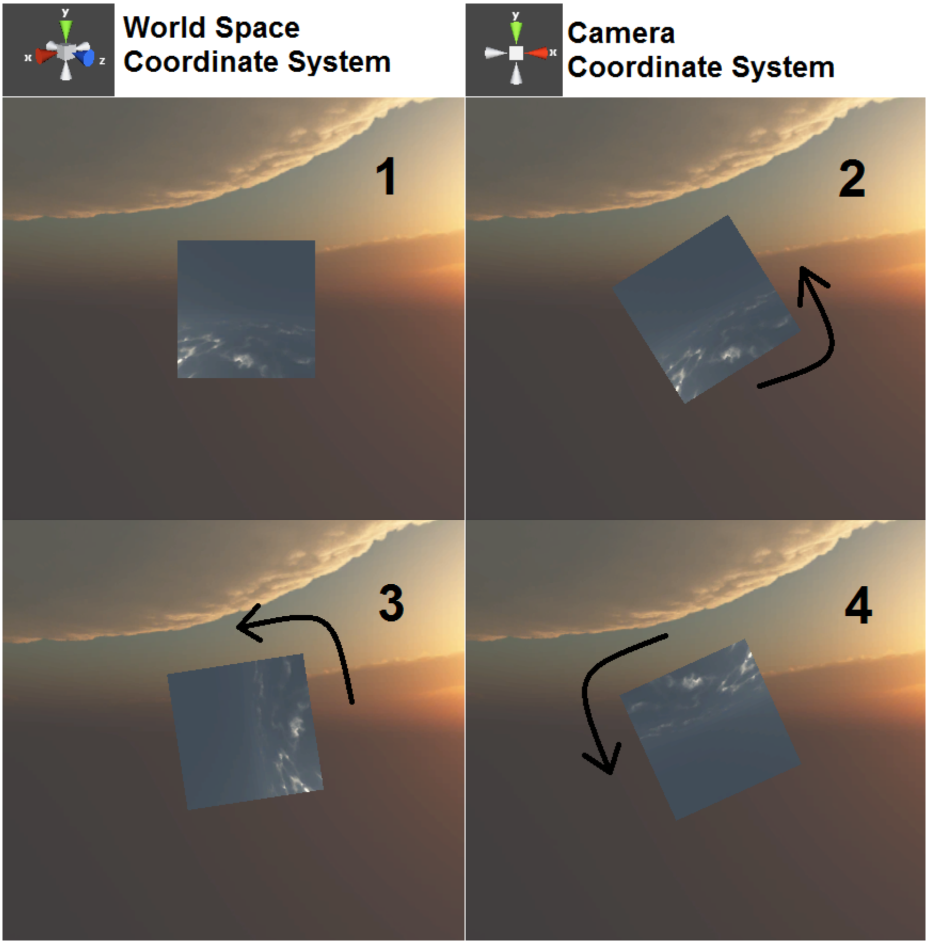


Fig. 3. For all frames the world space and camera coordinate systems are shown at the top of this illustration. The user is always looking down the *z-axis*. Frames 1 through 4 illustrate rotating a cube along the view direction (i.e. the *z-axis*). We can clearly see that only the camera coordinate system influences the rotation of the object.

Rotating objects in the second method is similar to that in method one described above, except there is one additional finger for each technique. Thus a two-finger drag gesture, typically used with two fingers side-by-side, controls the rotation through the Two-Axis Valuator. Moving two fingers left or right rotates the object about the *y axis* while moving them up or down rotates the object about the *x axis* of the camera coordinate system. With two-fingers in place, moving a third finger rotates the object around the view direction (i.e. the *z axis* of the camera coordinate system). Users commonly use the middle and index finger in place while moving the thumb below them, especially when rotating objects in the middle of the screen. When space is constrained at the edge of the screen, a different possibility is to use the middle, index, and to move the ring finger to rotate objects. A single finger tap on a surface of an object followed by a finger tap elsewhere in the scene mates the two surfaces. This is illustrated in Figures 4 and 5 below.

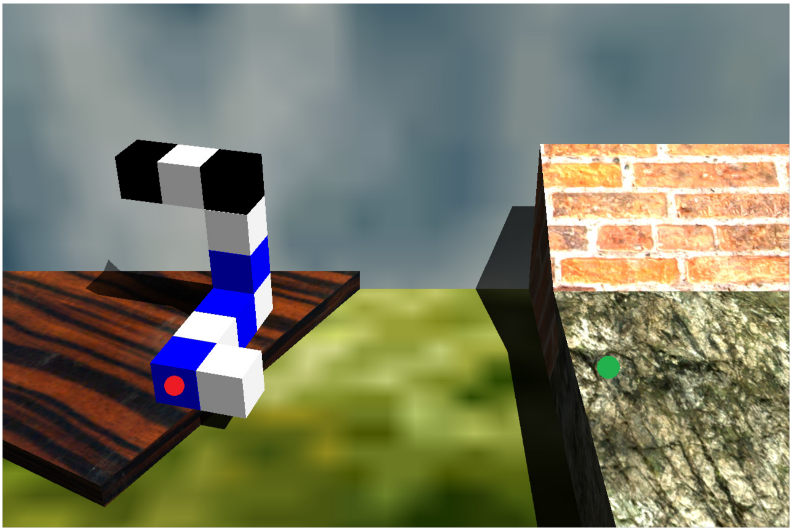


Fig. 4. The initial scene before performing a mating operation. By taping the Shepard-Metzler object at the location indicated by the red dot, and then taping elsewhere in the scene, say the green dot on the right hand side, the object is simultaneously rotated and translated (i.e. mated) onto the new surface. The final position is illustrated in Figure 5 below.

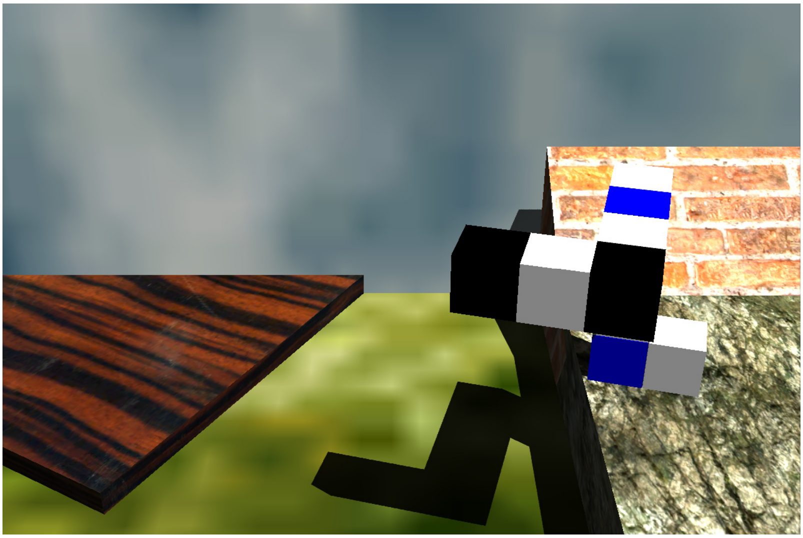


Fig. 5. Final position of the Shepard-Metzler object after the mating operation

If mating results in a collision, the object is pushed away from the target surface until the object is only in contact. This was illustrated in Figure 1 above. The reason for this is that in the real world objects very rarely interpenetrate and we designed our system to conform to this. If in a 3D world the user wants to place a chair on the ground, it is unlikely their goal is to have parts of the chair above the ground and parts of the chair disappearing into the floor. The effect of pushing objects away from the surface is relative to the user and the object. For example and if an object is mated onto a wall and interpenetrates it, the object will be pushed towards the user so that the object remains visible and inside the scene. The alternative of pushing it away from the user and the object becoming potentially invisible, is much more confusing. This effect is illustrated in Figures 6 and 7 below.

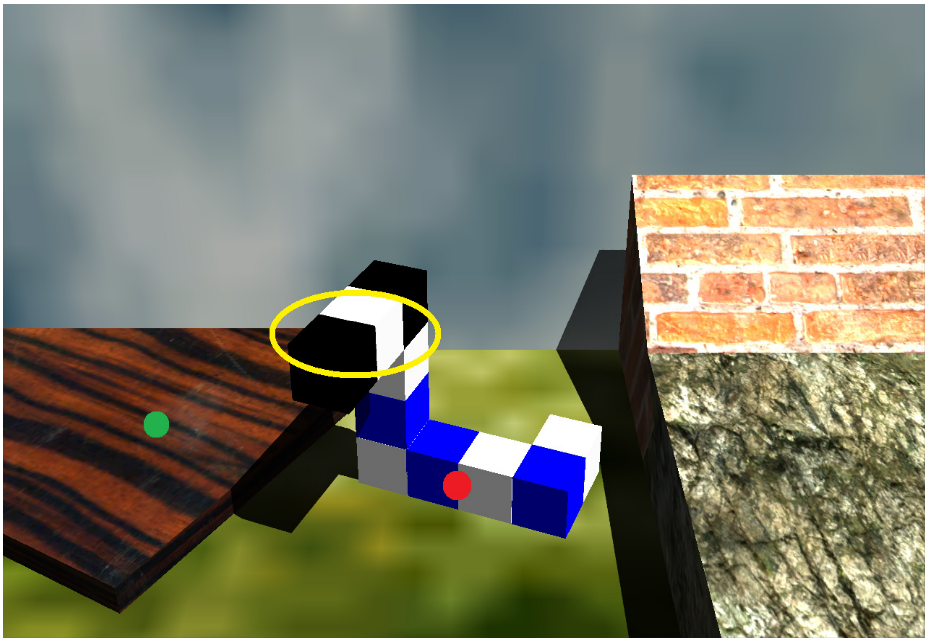


Fig. 6. In this example the Shepard-Metzler object is mated with the wooden floor to the left. If the user taps the object at the red dot and then taps the floor at the green dot, a naïve mating operation would result in the circled yellow part of the object disappearing into the wooden floor.

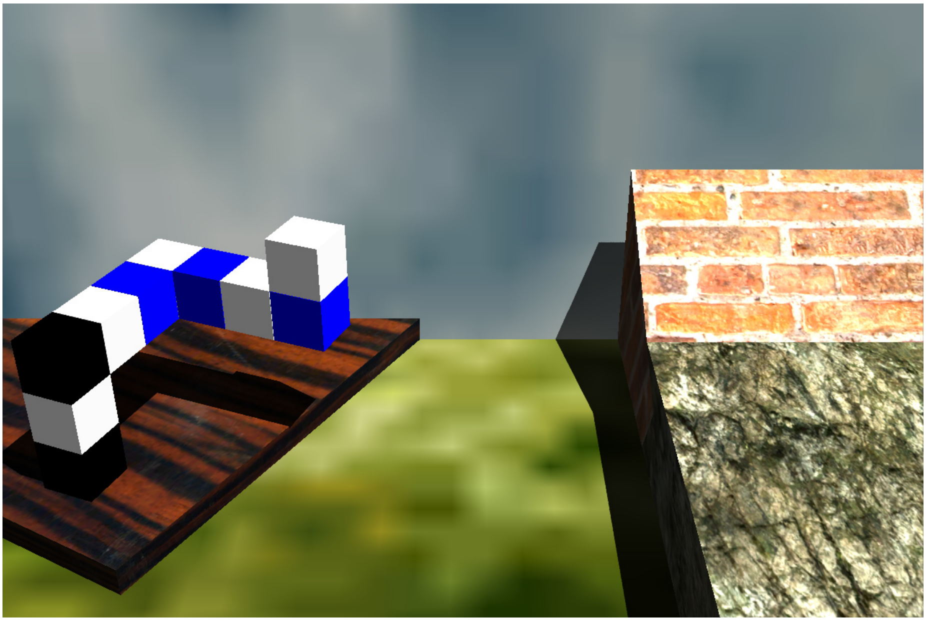


Fig. 7. The final position of the object in the scenario of Figure 6. The object has been pushed upwards, so that no parts interpenetrate the floor.

The movement from the initial position to the final position is animated to help users understand the result and to prevent them from having to visually reacquire the object position and orientation. This is especially helpful if an object is being moved over large distances over the scene. Also, before the objected is moved, our system pre-determines the final position and checks for collisions. If a collision is found, our system pushes the object away from the surface until there are no collisions, and this becomes the new final position of the object. Since the rotation and translation of the object is animated, if the object ends up in an unintended position or location, say as the result of a tap at the wrong location, the user will be able to see this visually and make appropriate corrections.

Moreover, a recently mated object is temporarily constrained. The system then rotates the object in the local coordinate system of the target surface. This enables users to mate an object onto any surface and then to quickly adjust the remaining DOF using either a two- or three-finger rotation. This permits users to quickly snap an object, for example a coffee cup, onto a table and then to turn the cup on the table to correctly orient the cup. The initial mating operation will ensure that the bottom of the cup is in complete contact with the surface, ensuring that it is not floating, leaving the user with the task of controlling only a single DOF to correctly orient the cup. This is substantially easier than having to always control all 3 DOF.

4 Methodology

4.1 Participants

Twelve paid volunteer participants were recruited from the local university campus and city. The age of the 6 male and 6 female participants ranged from 19 to 35 years (*mean* = 26.17, *SD* = 4.47). All had never participated in a 3D study before. All were right handed and preferred to use the tablet with their right hand. 3D video game usage was between 0 to 4 hours per week (*mean* = 1.92, *SD* = 1.62). Potential participants were asked in advance of the study if they had previously participated in a 3D study from our lab (even with another examiner) and we excluded people who indicated they had performed a user study involving rotating or translating 3D objects.

4.2 Apparatus

We conducted the study on an 8" Gadmei T883-3D tablet with Android 4.0.3. A widescreen monitor on a desktop computer was used in the second half of the study (i.e. during method 2) to display target scenes. The reasoning for this is explained below in the methodology section. We built a variety of commonly recognizable 3D objects, as well as several inspired by the Shepard-Metzler test [18]. After a pilot study we decided on a car, chair, dog, and one Shepard-Metzler object as shown in Figure 8 below. Colorings were introduced to disambiguate poses as, for example, a view onto the bottom of a chair with legs with identical colors does not reveal the full 3D rotation of the whole object at a glance. The white parts of objects highlighted in different colors for feedback. This reasoning for this is explained in more detail below.

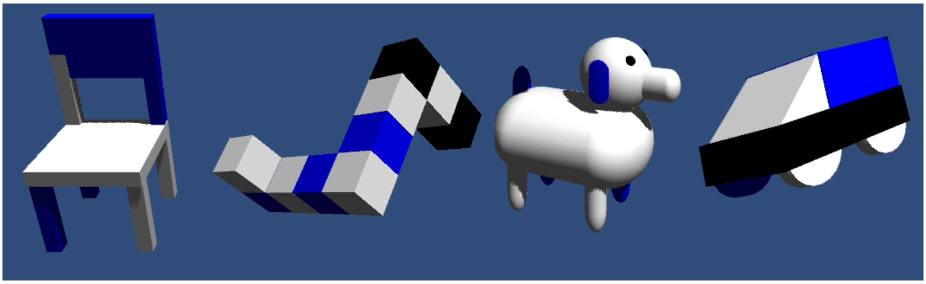


Fig. 8. Objects selected for use in both studies

4.3 Procedure

The study was conducted in a quiet room with the investigator, the sitting participant, and at most one other student present. The software was first configured to the participant's handedness. This affected the first part of the study where the rotating object was on the side of the user's preference and the target displayed on the other one.

Then the study was explained along with a system demonstration. All participants acknowledged they understood how the system worked and had no questions.

Overall we used a 2x2 within subject design with 2 phases. The first phase targeted only 3D rotations, whereas the second investigated rotations with constrained translations. The conditions were mating enabled or not, and surface aligned target orientations or not. These conditions were counterbalanced over all trials. In each set of 48 trials, subjects were asked to rotate 4 models 6 times with either mating enabled or not. Targets were aligned three of these six times, while the others had random target orientations. The order of each of these 24 trial blocks was determined using a Fisher-Yates shuffle. To determine the starting pose of the rotatable object we used two randomly shuffled copies of a list of 12 difference angles: 15, 30, 45, … , 165, 180 degrees. Each copy of the list matches to targets being aligned or not. To compute the final starting 3D orientation we first defined a rotation axis by generating a random point on a unit sphere. The user rotatable object was then rotated "back" from the target orientation about this axis by the angle chosen above. The participants' task was to rotate the object to within a quaternion angle of 10 degrees from the target orientation. Users could not abort trials.

Figure 2 above illustrates a trial in the first phase. Participants were presented with 2 objects. The target object on the left (or right) could not be manipulated and was the shown in the desired orientation. The object on the other side was rotatable. When the user touched the rotatable object the white segments turned magenta to indicate selection. When the object was within the 10 degree limit the white segments turned green to signal the user that the object had been rotated within an acceptable accuracy tolerance (the quaternion angle of 10 degrees mentioned above). A button then appeared in the upper corner of the screen on the side of the un-moveable object. Users pressed this button to move to the next trial, where it disappeared until the next object was correctly aligned. In phase one the focus was only on rotating objects. Thus, it was feasible to display both the target orientation and the user controlled object on the tablet screen side-by-side.

The task in the second phase shown in Figures 4 and 5 was to both translate and rotate the object into the target pose, a 5 DOF task. Said target pose was displayed on a second, desktop monitor in this phase. Since users performed tasks involving translations and rotations, which requires more space for interaction and scene display, we were unable to display both an interactive scene and the target scene on the same device. Hence the target scene was displayed on the monitor, giving the user a larger space to translate objects on the tablet. For example, the tablet would display the scene from Figure 4, while the desktop monitor would display the scene from Figure 5. The participants' task was then to orient the object from his scene to match the scene displayed on the desktop.

Here, white parts of the object highlighted in cyan when within 10 degrees of the correct orientation, in yellow when within 1/50th of size of the scene, and green when close to the correct pose. When the user was close enough, a button appeared in the upper right corner of the tablet screen which when pressed went onto the next trial. When pressed, the scene on the computer monitor changed. Thus users were not able to see the next trial in advance of completing the current trial.

5 Results

We found that using our mating system decreased rotation times substantially. According to a repeated measures ANOVA and in the 3D rotation task investigated in phase 1, there were significant effects of mating on completion time ($F_{1,11} = 23.06$, $p < .001$) and target alignment ($F_{1,11} = 100.92$, $p < .0001$). Both mating and aligned targets were significantly faster. There was also a significant interaction between the conditions. A Tukey-Kramer posthoc test shows that aligned scenarios with mating were approximately 65% faster than all other combinations. Figure 9 below illustrates average completion times. The results for the error angles show a significant effect for mating ($F_{1,11} = 93.83$, $p < .001$) and also confirm that aligned targets were positioned significantly more accurately ($F_{1,11} = 63.75$, $p < .001$).

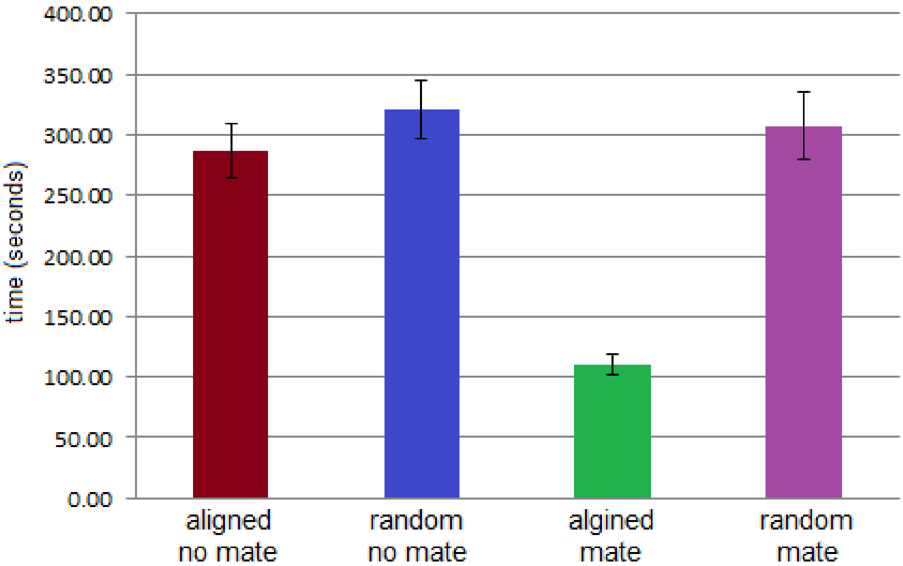


Fig. 9. Mean completion times for study 1 with standard error shown

In the 3D translation and rotation task in phase 2 there was again a significant effect of mating ($F_{1,11} = 37.68$, $p < .0001$) and aligned targets ($F_{1,11} = 61.7$, $p < .0001$) on task completion time. There was also a significant interaction between them and Tukey-Kramer identifies that aligned scenarios with mating were completed approximately 64% faster than all other combinations. Figure 10 below illustrates the average completion time per participant. In this phase, there was a significant effect on error angles for both mating ($F_{1,11} = 108.75$, $p < .0001$) and aligned targets ($F_{1,11} = 24.71$, $p < .0005$), as well as a significant interaction between them. Tukey-Kramer reveals that aligned objects were posed in the mating condition approximately 35% more accurately in terms of error angles compared to all other combinations.

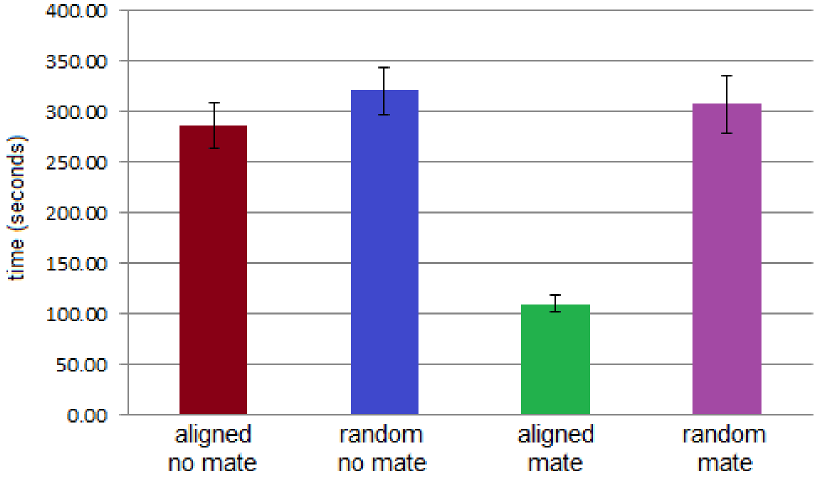


Fig. 10. Mean completion times for study 2 with standard error shown

5.1 Discussion

Our enhanced mating-based 3D rotation technique substantially decreases the time required to match aligned target orientations by 64% or more, while significantly improving accuracy. Given that most objects are aligned to others in many real world scenarios, this is substantial and exceeds all improvements found in previous work. Participants generally found either the chair or the dog the easiest object to rotate. Unanimously, the Shepard-Metzler model was judged the most difficult. Participants found the mating interface simple to use and all perceived it as faster. Although our original design for the flick gesture was targeted at thumb flicking, about half the participants preferred to use the ring finger instead for this technique.

Although this is not currently implemented in our system, we can easily adapt our method to enable mating even when there are no other surfaces present. For this we would simply add a (potentially invisible) back plane surface that is orthogonal to the view direction and proceed with mating to said surface. Optionally, this surface may be partially transparent to offer visual guidance to the user. One limitation of our current implementation is that we do not support mating objects to two (or more) surfaces simultaneously. This is an area we plan to investigate in future work.

5.2 Conclusion

We presented a new multi-touch 3D rotation technique based on mating to accelerate common tasks. It is targeted at one-handed touchscreen use, especially on tablets. Our user study revealed that the new technique improves manipulation times by more than 60% for common 3D rotation tasks. Rotation accuracy is significantly improved as well. In the future, we plan to investigate the performance of this mating technique for full 6 DOF manipulation tasks.

References

1. Bade, R., Ritter, F., Preim, B.: Usability comparison of mouse-based interaction techniques for predictable 3D rotation. In: Butz, A., Fisher, B., Krüger, A., Olivier, P. (eds.) SG 2005. LNCS, vol. 3638, pp. 138–150. Springer, Heidelberg (2005)
2. Bell, G.: Bell's Trackball (1988), http://members.tripod.com/professor_tom/index.html
3. Shoemake, K., Arcball: A user interface for specifying three-dimensional orientation using a mouse. In: Graphics Interface 1992, pp. 151–156 (1992)
4. Chen, M., Mountford, S.J., Sellen, A.: A study in interactive 3-D rotation using 2-D control devices. In: SIGGRAPH 1988, pp. 121–129 (1988)
5. Zhao, Y.J., Shuralyov, D., Stuerzlinger, W.: Comparison of multiple 3d rotation methods. In: VECIMS 2011, pp. 13–17 (2011)
6. Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., Kassell, N.: Usability analysis of 3D rotation techniques. In: UIST 1997, pp. 1–10 (1997)
7. Partala, T.: Controlling a single 3D object: Viewpoint metaphors, speed and subjective satisfaction. In: INTERACT 1999, pp. 486–493 (1999)
8. Jacob, I., Oliver, J.: Evaluation of techniques for specifying 3D rotations with a 2D input device. In: HCI 1995, pp. 63–76 (1995)
9. Reisman, J., Davidson, P., Han, Y.: A screen-space formulation for 2D and 3D direct manipulation. In: UIST 2009, pp. 69–78 (2009)
10. Hancock, M., ten Cate, T., Carpendale, S.: Sticky tools: Full 6DOF force-based interaction for multi-touch tables. In: ITS 2009, pp. 133–140 (2009)
11. Martinet, A., Casiez, G., Grisoni, L.: The effect of DOF separation in 3D manipulation tasks with multitouch displays. In: VRST 2010, pp. 111–118 (2010)
12. Kin, K., Miller, T., Bollensdorff, B., De Rose, T., Hartmann, B., Agrawala, M.: Eden: A professional multitouch tool for constructing virtual organic environments. In: CHI 2011, pp. 1343–1352 (2011)
13. Zhai, S., Milgram, P.: Quantifying coordination in multiple DOF movement and its application to valuating 6 DOF input systems. In: CHI 1998, pp. 320–327 (1998)
14. Boritz, J., Booth, K.S.: A study of interactive 6 DOF docking in a computerized virtual environment. In: VRAIS 1998, pp. 139–146 (1998)
15. Kratz, S., Rohs, M., Guse, D., Müller, J., Bailly, G., Nischt, M.: PalmSpace: Continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In: AVI 2012, pp. 181–188 (2012)
16. Stuerzlinger, W., Wingrave, C.: The value of constraints for 3D user interfaces. In: Virtual Realities: Dagstuhl Seminar 2011, pp. 203–224 (2011)
17. Oh, J.-Y., Stuerzlinger, W.: Moving Objects with 2D Input Devices in CAD Systems and Desktop Virtual Environments. In: Graphics Interface, pp. 195–202 (2005)
18. Shepard, R., Metzler, J.: Mental rotation of three dimensional objects. Science 71(972), 701–703 (1971)
19. Bier, E.A.: Snap-Dragging in Three Dimensions. In: Proceedings of the 1990 Symposium on Interactive 3D Graphics, pp. 193–204. ACM, New York (1990)
20. Zeleznik, R.C., Herndom, K.P., Hughes, J.F.: SKETCH: an interface for sketching 3D scenes. In: ACM SIGGRAPH 2007, Article 19. ACM, New York (2007)
21. Liu, J., Au, O.K.-C., Fu, H., Tai, C.-L.: Two-finger gestures for 6dof manipulation of 3d objects. Comp. Graph. Forum 31(7pt1), 2047–2055 (2012)

HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments

Weidong Huang[1], Leila Alem[1], and Franco Tecchia[2]

[1] CSIRO, Australia
[2] PERCRO - Scuola Superiore Sant'Anna, Italy
{Tony.Huang,Leila.Alem}@csiro.au, f.tecchia@sssup.it

Abstract. A collaboration scenario involving a remote helper guiding in real time a local worker in performing a task on physical objects is common in a wide range of industries including health, mining and manufacturing. An established ICT approach to supporting this type of collaboration is to provide a shared visual space and some form of remote gesture. The shared space and remote gesture are generally presented in a 2D video form. Recent research in tele-presence has indicated that technologies that support co-presence and immersion not only improve the process of collaboration but also improve spatial awareness of the remote participant. We therefore propose a novel approach to developing a 3D system based on a 3D shared space and 3D hand gestures. A proof of concept system for remote guidance called HandsIn3D has been developed. This system uses a head tracked stereoscopic HMD that allows the helper to be immersed in the virtual 3D space of the worker's workspace. The system captures in 3D the hands of the helper and fuses the hands into the shared workspace. This paper introduces HandsIn3D and presents a user study to demonstrate the feasibility of our approach.

Keywords: remote collaboration; co-presence, mixed reality, hand gesture, shared visual space.

1 Introduction

It is quite common nowadays for two or more geographically distributed collaborators to work together in order to perform actions on physical objects in the real world. For example, one remote expert might be assisting an onsite maintenance operator in repairing a piece of equipment. Such collaboration scenarios are highly asymmetrical: the onsite operator is co-located with the machine being manipulated or fixed but does not have the required expertise to do the job, while the remote expert does not have physical access to the machine but knows how to trouble shoot and fix it. This type of collaboration scenarios is common in many domains such as manufacturing, education, tele-health and mining.

When co-located, collaborators share common ground, thus being able to constantly use hands gestures to clarify and ground their messages while communicating with each other verbally. However, when collaborators are geographically distributed, such

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 70–77, 2013.

common ground no longer exists, resulting in them not being able to communicate the same way as they do when co-located. Prior research has shown that providing shared visual spaces and supporting remote gesture can help to build common ground [2, 3]. A shared visual space is one where collaborators can see the same objects at roughly the same time. As a result, a number of remote guiding systems have been reported in the literature to achieve these two goals. While how remote gesture is supported may differ from system to system (e.g., [1, 4]), the shared visual space is generally provided in the 2D format of either video feeds or projection on surfaces.

A recent study on a remote guidance system by Huang and Alem [5] indicated that with 2D shared spaces, helpers had difficulties in perceiving spatial relation of objects. Helpers also had a relatively lower sense of co-presence [6]. Spatial understanding is critical for helpers to make right judgements on objects and guide workers accordingly, while co-presence has been shown to be associated with user experience and task performance [7]. Therefore, these are two important factors and should be addressed properly.

Research has shown that immersive virtual environments (IVEs) help improve spatial understanding [9]. Further, IVEs also bring other benefits [8], such as higher sense of co-presence, improved spatial awareness, more accurate cognitive transfer between simulation and reality and better task performance. Although they have been shown useful in supporting general tele-collaboration in which all collaborators work within the same virtual environment, we wonder if IVEs still help in the context of remote guidance.

We therefore propose a new approach that provides 3D shared visual spaces. A prototype system called HandsIn3D has been developed for the purpose of the proof of concept (see [10] for more details). This system uses a head tracked stereoscopic HMD that allows the helper to be immersed and perform guidance in the virtual 3D space of the worker's workspace. In the remainder of this paper, we introduce HandsIn3D and present a user study of it.

2 HandsIn3D

HandsIn3D is currently running on a single PC. It has two logical sections: the worker space and the helper space (see Figure 1). The worker performs a physical task at the worker space, while the helper provides guidance to the worker at the helper space. The left image shows the layout of the worker space. A user sits at the desk performing a task on physical objects (for example, assembly of toy blocks). A 3D camera is mounted overhead to capture the workspace in front of the user including the hands of the user and objects. A LCD non-stereoscopic monitor is placed on the desk to display the 3D view of the workspace augmented by the guiding information.

The right image of Figure 1 shows the layout of the helper space. In this space, there is a 3D camera that captures the hands of the helper. The helper wears a stereoscopic HMD and sits in front of an optical head tracker. The HMD allows a realistic virtual immersion in the 3D space captured by the camera placed in the worker space, while the tracker tracks the HMD position and orientation.

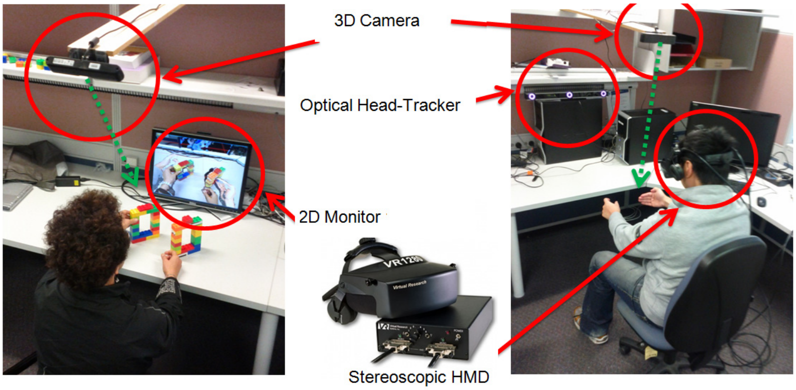


Fig. 1. Worker space (left) and helper space (right)

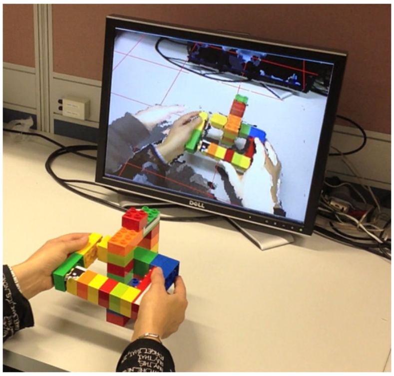


Fig. 2. The *shared virtual interaction space* is shown in the LCD monitor. The 3D meshes captured by two cameras are co-located and fused together. Four hands can be spotted in the virtual scene: 2 from the worker and 2 from the helper [10].

The system functions as follows. On the worker side, the worker talks to the helper, looks at the visual aids on the screen, picks up and performs actions on the objects. On the helper side, the helper wears the HMD, looks into the virtual space (and looks around the space if necessary), talks to the worker and guides him by performing hand gestures such as pointing to an object and forming a shape with two hands. During the process of interaction, the camera on the worker side captures the workspace in front of the worker, while the camera on the helper side captures the hands of the helper. The acquired 3D scene data from both sides are fused in real time to form a single common workspace which we call *shared virtual interaction space*. This space is displayed in HMD. The image in Figure 2 is provided to help understand what is presented to the helper during the task execution: by fusing together the 3D meshes acquired by the two cameras, an augmented view of the workspace where the hands of the helper are co-located in the worker space is synthetically created, as shown in the LCD monitor. On the other hand, being presented with this augmented view, the worker can easily mimic the movements of the helper hands and perform the Lego assembly task accordingly.

The main features of the system therefore include the following:

- Users can speak to each other.
- The helper can see the workspace of the worker via the shared virtual interaction space.
- The helper can perform hand gestures.
- The worker can see the hand gestures of the helper on the screen.
- The two hands of the worker are freed for manipulation of physical objects.

In addition, the shared virtual interaction space also implements additional features to improve the sense of 3D immersion for the helper. These features include 1) the objects and hands cast shadows in the space; 2) the HMD is tracked which allows the helper to see the space from different angles.

3 A User Study

The user study was conducted to evaluate our 3D gesture based interaction paradigm. We were particularly interested in how helpers felt about the 3D user interface.

3.1 Method

Fourteen participants who had no prior knowledge of the system were recruited. Upon their agreement to participate, they were randomly grouped as pairs to perform a collaborative task. In this study, we used the assembly of Lego toy blocks as our experimental task, which is considered as representative of real world physical tasks and has been used for the similar studies. During the task, the worker was asked to assemble the Lego toys into a reasonably complex model under the instruction of the helper. The helper was instructed that he could provide verbal and gestural instructions to the worker at any time. The worker, on the other hand, had no idea about what steps were needed to complete the task.

In order to give users a better appreciation of our new 3D interface in relation to different design options, following the assembly task, the pair was given the opportunity to explore and experience different levels of immersion: 1) no stereoscopic vision, no head tracking and no hands shadow (2D interface), 2) stereoscopic vision, no head tracking and no hands shadow, 3) stereoscopic vision, head tracking and no hands shadow, and 4) stereoscopic vision, head tracking and hands shadow (full 3D interface). This last feature is one that was implemented in HandsIn3D and that participants used in the guiding task at the start of the trial.

Participants were required to complete worker and helper questionnaires after the assembly tasks. These questionnaires asked participants to rate a set of usability metrics, answer some open questions and share their experience of using the system. The usability measures include both commonly used ones and those specific to the system. For more details, see the Results subsection.

3.2 Procedure

The study was conducted a pair after another in a meeting room and was observed by an experimenter. The helper space and worker space were separated in a reasonable distance by a dividing board so that two participants could not see each other. Upon arrival, they were randomly assigned helper and worker roles and informed about the procedure of the study. The helper interface and the worker interface were introduced. They were also given the chance to get familiar with the system and try out the equipment. Then the helper was taken to an office room where he/she was shown a model that needed to be constructed. The helper was given time to think about and plan how to do it and remember the steps.

Then the helper went back to the experimental room and put the HMD on and the experiment started. After the assembly task, the pair of participants was asked to experience the different interface features listed in the last subsection in an informal style. The switch between the interface features was controlled by the experimenter. During the process, the participants were told which feature the system was using. They could play with the toy blocks and talk to each other about the assembly steps. But they were not allowed to comment and share how they felt about the system and the features. This was to ensure that their later responses to the questionnaires were not affected by each other's partner and were of their own. After going through all four features, each participant was asked to fill the helper or worker questionnaire for the role played. Then the participants switched roles and the above process was repeated again. Note that this time the model to be constructed was different but with a similar level of complexity.

After finishing the assembly tasks and questionnaires, participants were debriefed about the purposes of the study, followed by a semi-structured interview. They were encouraged to share their experiences, comments on the system, ask questions and suggest improvements. The whole session took about one hour on average.

3.3 Results and Discussion

Observations. It was observed at the beginning, some participants were very shy about wearing a HMD, resulting in very few head movements. Participants needed prompting and encouragement in order to start moving their head around and change their field of view. This indicates that users may need to take some time to get used to system, as one user commented: "It took me about 10 seconds to adapt to the 3D viewpoints. But after that everything is fine."

Apart from this, all pairs of the participants were able to complete their assigned tasks without apparent difficulties. Their communications seemed smooth. Both helpers and workers looked comfortable performing tasks with the system. More specifically, workers were able to follow the verbal instructions from helpers and understand what they were asked to do by looking at the visual aids shown on the screen. Helpers were able to guide workers through the task process with the HMD worn on their head and using hand gestures.

Usability Ratings. Fourteen participants filled two questionnaires each: the helper questionnaire and the worker questionnaire. We had 28 responses in total. A set of usability measures were rated based on a scale of 1 to 7 with 1 being strongly negative, 7 being strongly positive and 4 being neutral; the higher the rating is, the better the usability. The average ratings are illustrated in Figure 3. Note that 1) helpers had two extra items to rate: perception of spatial relationship between objects and sense of immersion; 2) due to the space limitations, we only report here the ratings of the full 3D system.

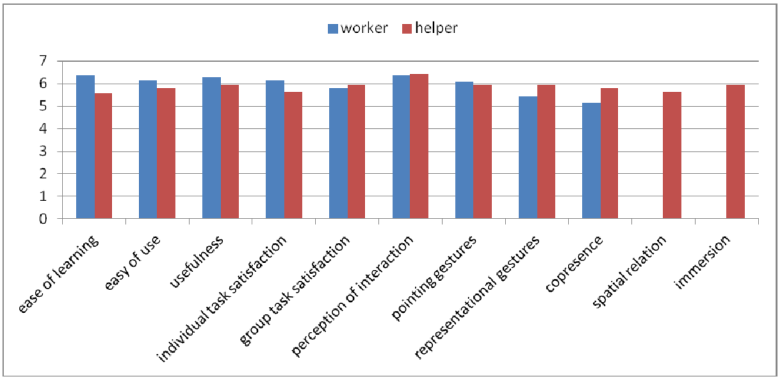


Fig. 3. Average user ratings

As can be seen from Figure 3, despite the slight variations between helpers and workers and across usability items, all items were rated greater than 4, indicating that participants were generally positive about the system. Further, helpers rated the system relatively low for its learning and usage compared to workers. While the system made workers more satisfied with their own individual performance, helpers were more satisfied with the overall group performance. In addition, while helpers gave the same rating for being able to perform both pointing and representational gestures, workers seemed to perceive pointing gestures more easily than representational gestures.

In regard to co-presence, both ratings were over 5, which were higher than those reported by Huang and Alem [6] (just above 4). This indicates that our 3D system offered a higher sense of "being together" for participants. When compared to workers, helpers reported a relatively higher sense of co-presence. Helpers also had positive ratings for perception of object spatial relation and sense of immersion. All these indicate that our 3D design approach did work as expected.

User Experiences. Based on user responses to the open questions and user interviews, participants were generally positive about the system, as one participant stated that "it is very impressive and a great experience to use this system."

More specifically, participants appreciated the feature that workers are able to see and helpers are able to perform hand gestures. A helper commented that "he (the worker) knew exactly what I meant by 'here', 'this one' and 'that one'." A number of workers simply commented that "(hand gestures were) easy to understand and follow."

Consistent with the usability ratings, the 3D interface has boosted a strong sense of co-presence and immersion for helpers. It was commented that the system had given participants a feeling of being in front of the remote workspace and co-presenting with the remote objects and their remote partner. Comments from helpers include "I feel I was right there at the screen and really wanted to grab the objects." and "I can feel that her hand was in my way, or my hand was in her way. So in this sense, I felt we were in the same space." A few workers also commented that seeing both hands in the field and using words like "this" and "that" during the conversation made a strong visual impression and physical presence.

User comments also provide further evidence that the 3D interface improved perception of spatial relation and participants appreciated that. For example, "You can see the difference between 2 objects with the same base but different heights in 3D." "3D helped to see how the final shape looked. With 2D, I had to ask the worker to show me the piece in another angle." "It gives the depth of the objects, so remote guiding could be easier in some cases."

Participants generally liked the idea of having shadows of hand and objects, commenting that it would be easier to point and gesture in the remote workspace as hand shadows could provide them with an indication of hand location in relation to the remote objects. However, there were mixed responses when participants were asked whether the shadow feature actually helped. For example, "Yes, it helps. It makes a good stimulation effect. So I can do better communication with my partner."

"The shadow helps me feel that the view is in 3D. But I think I can still understand without the shadow." "No, there are some real shadows in grey color. The black shadow is artificial and a little bit annoying." "The shadow could sometime cover objects and I think this could potentially lead to something wrong (maybe a transparent shadow)." "Yes, (shadow helps) for pointing only, but not much on rotating etc."

In regard to head tracking, participants commented that it enables helpers to see more of the way that blocks are connected without needing their partner to rotate them and that it makes workers aware of what helpers are looking at.

Further, in comparison with the 2D interface, participants commented that 2D is fine with simple tasks, but 3D offers much richer user experience and is more preferable and useful for complex procedures when a high level of user engagement is required. For example, "3D is more realistic as I can see all angles." "In 2D, it seems like playing a game. When changing my viewpoints into 3D, I got a feeling of going back to real world." "(3D) helps more when I need to give instruction related to space concept." "3D interface makes it easy to match the screen and the physical objects." "3D feels real. 2D interface is enough for simple tasks but 3D interface helps more when the task gets more complicated."

Although the main purpose was to test the usability and usefulness of our 3D concept for remote guidance on physical tasks, user comments also gave some hints for further improvements and more features. These include 1) use a lighter and more easily adjustable helmet; 2) increase image quality and resolutions; 3) differentiate worker and helper hands by color and make them transparent; 4) provide a more dynamic and more immersive environment for helpers to interact with, such as when the helper moves closer to the objects, they become bigger; 5) enable helpers to have access to manuals while guiding; 6) make shadows grey and transparent.

4 Concluding Remarks

Our user study has showed that the proposed 3D immersive interface is helpful for improving users' perception of spatial relation and their sense of co-presence and that the system is generally useful and usable, particularly more so for complex tasks.

The study also points out some future research and development directions. We plan to advance the prototype into a close-to-production system so that we can test it in a more realistic setting. For example, separate the two sides of the system and connect them through the internet, instead of hosting them by the same PC. We also plan to compare HandsIn3D with its 2D versions through rigorously controlled studies so that we can have more quantitative and objective information about the benefits of immersive virtual environments in supporting remote guidance.

References

1. Alem, L., Tecchia, F., Huang, W.: HandsOnVideo: Towards a gesture based mobile AR system for remote collaboration. In: Recent Trends of Mobile Collaborative Augmented Reality, pp. 127–138 (2011)
2. Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., Kramer, A.D.I.: Gestures over video streams to support remote collaboration on physical tasks. Human-Computer Interaction 19, 273–309 (2004)
3. Gergle, D., Kraut, R.E., Fussell, S.R.: Using Visual Information for Grounding and Awareness in Collaborative Tasks. Human-Computer Interaction 28, 1–39 (2013)
4. Gurevich, P., Lanir, J., Cohen, B., Stone, R.: TeleAdvisor: a versatile augmented reality tool for remote assistance. In: CHI 2011, pp. 619–622 (2011)
5. Huang, W., Alem, L.: Gesturing in the Air: Supporting Full Mobility in Remote Collaboration on Physical Tasks. Journal of Universal Computer Science (2013)
6. Huang, W., Alem, L.: Supporting Hand Gestures in Mobile Remote Collaboration: A Usability Evaluation. In: Proceedings of the 25th BCS Conference on Human Computer Interaction (2011)
7. Kraut, R.E., Gergle, D., Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence. In: CSCW 2002, pp. 31–40 (2002)
8. Mortensen, J., Vinayagamoorthy, V., Slater, M., Steed, A., Lok, B., Whitton, M.C.: Collaboration in tele-immersive environments. In: EGVE 2002, pp. 93–101 (2002)
9. Schuchardt, P., Bowman, D.A.: The benefits of immersion for spatial understanding of complex underground cave systems. In: VRST 2007, pp. 121–124 (2007)
10. Tecchia, F., Alem, L., Huang, W.: 3D helping hands: A gesture based MR system for remote collaboration. In: VRCAI 2012, pp. 323–328 (2012)

MotionBender: A Gesture-Based Interaction Technique for Editing Motion Paths

Frederic Raber[1], Quan Nguyen[1], and Michael Kipp[2]

[1] DFKI GmbH, 66123 Saarbrücken
[2] Augsburg University of Applied Sciences, An der Hochschule 1, 86161 Augsburg
{Frederic.raber,quan.nguyen}@dfki.de,
michael.kipp@hs-augsburg.de

Abstract. Precision tasks in 3D like object manipulation or character animation call for new gestural interfaces that utilize many input degrees of freedom. We present MotionBender, a sensor-based interaction technique for post-editing the motion of e.g. the hands in character animation data. For the visualization of motion we use *motion paths*, often used for showing e.g. the movement of the hand through space over time, and allow the user to directly "bend" the 3D motion path with his/her hands and twist it into the right shape. In a comparative evaluation with a mouse-based interface we found that subjects using our technique were significantly faster. Moreover, with our technique, subject movement was more coordinated, i.e. movement was done in all three dimensions in parallel, and the participants preferred our technique in a post-experiment questionnaire. We also found a gender effect: male users both like the gesture interaction better and achieve better performance.

Keywords: coordination, character animation, motion trajectory, Kinect, 3D user interfaces.

1 Motivation

Character animation is required in many fields, from movie production and games to anthropomorphic interfaces. However, animation is a complex craft that requires high expertise. Current animation approaches can be divided into two categories. In *interactive methods* animation is created on-the-fly. The animator either moves a virtual model using his/her limbs or a controller. The model's motion is recorded during this animation process. In *keyframe methods* the animator defines several decisive "key" frames, using a 3D modeling tool. In-between frames are automatically computed by interpolation. This is the standard method in the industry. Neff et al. [5] presented an interactive system where this animation is done via the mouse. The movement of the mouse is mapped to respective joint movements by correlation maps. Kipp/Nguyen [3] explored the possibilities of a multitouch-surface to interactively animate the arm and hand of a puppet. There are several articles [4][2] which describe keyframe methods, using a motion trajectory for visualizing and editing the animation of each joint. In these solutions the trajectory is modified with conventional input devices like a mouse which have few degrees of freedom.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 78–85, 2013.

We propose a novel interaction technique called MotionBender (Figure 1) for editing motion paths by combining the possibilities of direct manipulation using a vision-based motion sensor (Kinect) and a simple button for switching between edit modes. The modes differentiate between selecting where to "grab" the path and moving these parts to "bend" the path into a new shape.

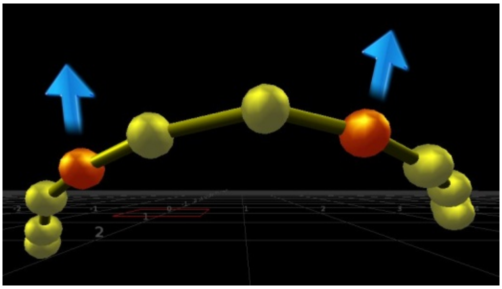


Fig. 1. In character animation, the movement of a joint, like the hand/wrist is visualized as a motion path. With our technique the user can modify this path by grabbing it (two red spheres) and bending it. The goals of our editing technique are:

- **Effective**: Suitable and useful for post-processing character animation data.
- **Efficient**: At least as fast as the current industry-standard techniques
- **Intuitive**: Can be directly used without too much instruction.
- **Joy of use**: Users should enjoy working with this technique because they feel "in control" in the sense of a *Natural User Interface* (NUI)

For such a gesture-based interface, it is important that the user makes use of all degrees of freedom that such an interface offers. If s/he uses the system in a sequential way, one dimension at a time, the benefit against conventional input systems is decreased. We used a coordination measure in our evaluation to measure how coordinated the user's movement is.

2 Interaction Technique

2.1 Overview

Our system is called *MotionBender*. It was created with a system like Min et al.'s [4] in mind, providing the possibility for a sensor-based interactive editing of the motion paths. Like in Min et al.'s work, every joint (e.g. the wrist of one hand) has a motion path, describing the movement of this joint during the animation, as shown in Figure 1. The spheres represent the positions of a single joint at successive time points. The user selects two of the spheres on the curve (marked red) and moves them to another location in the virtual world. The surrounding parts of the motion path are automatically adjusted. After the user has finished, MotionBender recomputes the animation of the joint according to the new motion path. Figure 2 shows a user using MotionBender in a typical working situation. The corresponding video can be found online[1].

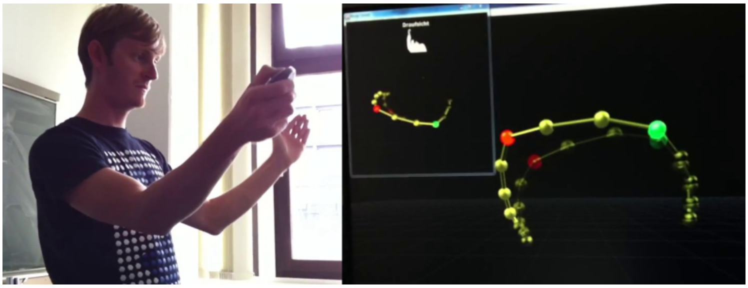


Fig. 2. User working with the MotionBender interface

2.2 Workflow

We focus on the editing of a single motion path. The manipulation is done with a Kinect sensor and a two-button device in one hand. There are two modes: *grab* and *bend*. The user stands in front of the computer. In grab mode, the user selects two points on the motion path (marked red in Figure 1) where the path is "grabbed". This is done by clicking and holding the first button and moving the hands apart and together on a horizontal axis in front of the user. In bend mode, activated by clicking and holding the second button, the two selected spheres and their neighboring spheres follow the user's hand movements in 3D space (indicated by the blue arrows in Figure 1). Figure 3 shows how neighboring points (yellow) are moved in relation to the grabbed points (red).

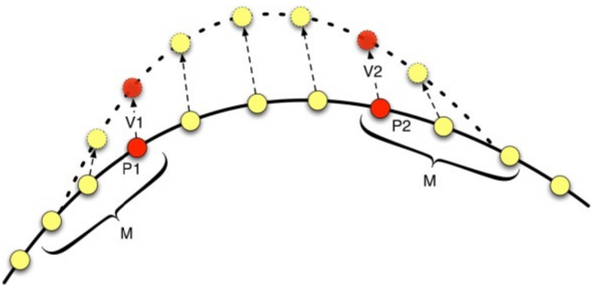


Fig. 3. Effect of the neighbor function

To compute the change for the neighboring points, we differentiate between three types of points. Let us assume that points P1 and P2 have been moved by vectors V1 and V2.

[1] `http://tiny.cc/MotionBender`

- Points between P1 and P2 are moved by a vector which is a weighted interpolation between V1 and V2, where the weight corresponds on the distance to the points.
- All other points, if they are less than M points away, are moved into the same direction as the nearest point moved by the user, but with a decreased distance. Formally, for every point P which is nearer to P1, its movement vector V will be calculated as V = k * V1, where k decreases linearly with increasing distance between P and P1.
- All other points are not moved.

3 Experiment

3.1 Methods

Participants

One of the main goals of our work was to create a tool which can also be used by novices in character animation or similar fields. Therefore, we decided not to take experts for our evaluation, but lay people. We tested 21 participants. Two were taken out due to aborted runs. We thus analyzed 19 people (12 female, 7 male) of age 18-56 years (average 24). Two of the subjects are left-handed but use the mouse with their right hand. Participants were not paid.

Control Condition (Mouse Interface)

When trying to design the control condition, we started out with a conventional one-mouse interface, using a graphical widget to allow control over the various degrees of freedom. However, as Owen's work on bimanual curve manipulation [6] suggests, this would lead to an unfair comparison, as bimanual input is more efficient than unimanual. Since the Kinect also has more DOF, we had to extend this to make a fairer comparison. To allow the simultaneous usage of two hands, we designed an interface with two mice, one for each "grab point". As for the control mapping of each mouse, we made the x/y mouse motion control the x/y motion on the screen (frontal plane), whereas the scroll wheel would move along the depth axis (z axis). One could argue that the scroll wheel is usually used for up/down motion (y axis). However we deemed the mapping from x/y mouse motion to x/y motion in 3D space even more common, and since the mouse is usually pointing toward the screen the scroll wheel acts like physical metaphor of rolling with a wheel into the screen. In user studies, participants did not negatively comment on the scroll wheel mapping.

There are 12 trials, divided into two subsets A and B of 6 items each. Each subset was randomized in a controlled manner (e.g. reversed) to obtain new sets A' and B'. Each participant worked on either A or A’ for one condition, and B or B’ for the other condition. To avoid training effects on one condition, the first half of the subjects started using the condition “Kinect” and continued with the “Mouse” condition, and vice versa for the second half of the subjects.

Apparatus

For the experiment the user had to stand in front of a 22 inch screen in a distance of about two meters. For the two-mouse interface, the subject was sitting directly in front of the screen. The complete surface of the office desk was available for him to act upon with the mice (Figure 4). We are aware of the fact that the difference in the user-screen distance could give a slight advantage for the mouse condition, based on the assumption that for precision tasks it is better to be closer.

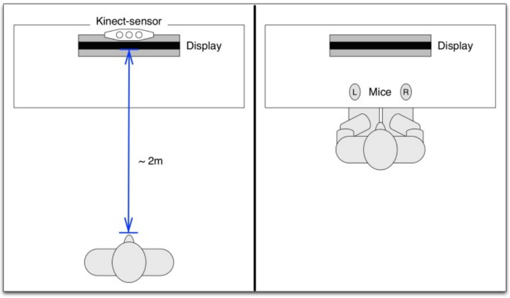


Fig. 4. The Kinect setup (left) and the two-mouse interface (right)

Task

The task was to match a given trajectory to a target trajectory (Figure 2). Each subject uses both interface versions and has to match six different trajectories for each interface. A trial is done if either the distance between given and target trajectory is below a certain threshold (successful match) or the maximum edit time (120 sec) has expired. In either case, the user directly continues with the next trial.

Procedure

The test procedure for all test candidates was as follows:

- At the start of the experiment the subject reads the written instructions.
- After the subject finishes reading, the experimenter gives a demo of the interfaces and gives hints on solution strategies.
- After a short question and answer session, the subject can practice on two sample trials, which will not be part of the main evaluation phase later. The sample trials and their order are always the same.
- The main evaluation begins, where the subject is presented 6 trials to be matched in the first interface condition (Kinect or Mouse).
- Phases 2-4 are repeated for the second condition (Mouse or Kinect).
- The experiment ends with a short two-page questionnaire, where the subject expresses his/her experiences with both interfaces in a multiple-choice form on the first page, and gives free feedback about problems/issues on the second page.

Measures

For our analysis, we computed three measures.

1. **Completion time:** Average time needed to complete a trial.
2. **Mean squared error** of all points which have not been matched yet.

3. **Coordination**: For comparing the movements regarding coordination, we use the measure defined by Kipp/Nguyen [3]. Coordination measures whether movement is performed in all three dimensions in parallel (high value) or sequentialized, i.e. performed along one dimension at a time (low value). Alternative coordination measures [1][7] are based on a fixed optimal solution path. For our problem, the optimal solution may be counterintuitive. The various selection possibilities for selecting grab points further complicate the matter. Therefore, such optimal-path metrics may yield misleading results. Kipp/Nguyen's measure is independent of optimal paths, taking only the raw movement vectors into account.

3.2 Results

Completion Time

Figure 5 shows the average completion times over all participants and motion path samples. A paired t-test proved that subjects are significantly faster with the Kinect (M=79.36; SD=23.9) compared with the mouse (M=83.71; SD=19.02): t(18)=3.89, p<0.01. To test for gender, we analyzed males and females separately. While the performance for the mouse was statistically equal, we found that, for the Kinect, male participants were much faster (M=65.24) than female ones (M=87.59) which we found to be highly significant (t(17), p-value<0.001).

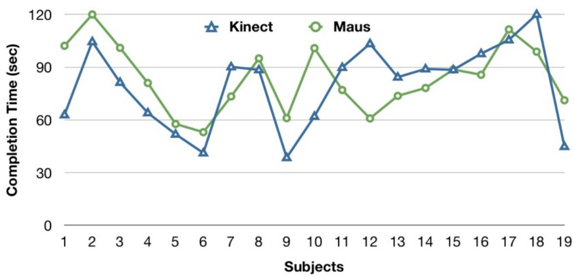


Fig. 5. Average completion times over all subjects

Coordination

Using the above mentioned coordination measure [3] we compared the performance for the left and right hand separately over all participants and devices (see Figure 6). In both cases the Kinect interface encourages significantly higher coordination: For the mouse device the coordination values are almost zero (left hand 0.009, right hand 0.007), whereas the Kinect interface reaches a value of 0.29 for the left hand and 0.26 for the right hand.

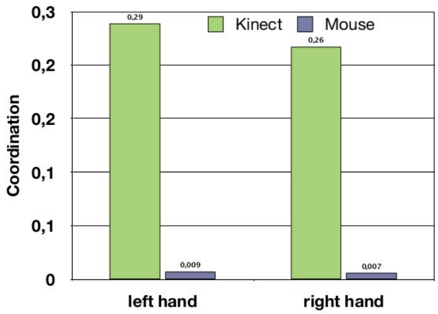


Fig. 6. In terms of coordination Kinect (green) significantly outperforms mouse (blue)

Questionnaire

In a post-experiment questionnaire we asked the participants questions which discriminated between Kinect and mouse on a 5-point differential scale where the middle position was labeled "both or equal". This questionnaire contained 17 questions regarding usability and user experience. For analysis, we transformed this to an interval between -2 for mouse and +2 for Kinect. We summed up ratings for each question over all subjects and computed a chi square statistic, comparing the ratings against the expected neutral value of zero. Only three questions reached significance, all three in favor of the Kinect interface.

- Which interface was **easier** to use? (M=.21; $chi^2(2, n=19) = 11.47; p<0.005$)
- Which device was more **fun** to use? (M=.84; $chi^2 (2, n=19) = 8.58; p<0.05$)
- Which device do you **prefer**? (M=.47; $chi^2 (2, n=19) = 11.47; p<0.005$)

4 Discussion

The results show that our MotionBender interface is objectively more efficient and subjectively better liked. Performance was significantly better, movement was clearly more coordinated and participants preferred it over the mouse interface and enjoyed it more. We also found statistical evidence that male users seem to profit more from our gesture-based technique than female users. Due to the low number of participants this may be an artifact and must be validated in future studies, which might also reveal a possible cause of that effect. However, time improvements were relatively small, compared to the overall task time. Our experiment design might have favored the mouse interface due to the closer user-screen distance. However, even if this was the case it only strengthens the results.

The Kinect was more coordinated than mouse. For mouse, users mostly acted separately, so they stopped moving the mouse when adjusting the depth position by rotating the scroll wheel. Some found this a limiting factor, others liked the fact that no unintentional movement along the z-axis was possible. The left hand was more coordinated than the right one in the mouse condition. We often observed that people are acting serially in matching the two selected points. They concentrated on the first (left), neglecting the second one initially. On the sensor-based interface, there is less of a difference. This might also be a result of the reduced cognitive difficulty with the direct-manipulation concept. Some problems emerged because subjects had problems to visually understand the 3D scene. Some suggested that a 3D stereoscopic view could reduce problems. We had two views on the scene, a top and a front view. Some users were confused by these two different views, and temporarily lost orientation.

In one interesting case, a subject had general problems with precise hand movements. She found the Kinect interface much easier to use because due to the whole-body motion some focus was taken away from the precision of her hand movements.

Conclusion

We presented an interaction technique called MotionBender, suitable for Character Animation. We used the concept of motion paths for the visualization of the animation and provided an editing method for post processing the animation by grabbing and bending the motion path. For this, we used a combination of a motion sensor (Kinect) and a simple button controller (wireless mouse) as input devices. The technique was intended to be efficient, intuitive and enjoyable. We validated our technique against a mouse-based interface and found MotionBender to be faster, easier to use and more enjoyable. Although that a comparison of our technique against a single-mouse interface subjectively appears to be unfair, that question should be checked in a future work. We also found that our technique encouraged highly coordinated movement. A gender effect indicates that this technique particularly appeals to male users but this finding needs further study. Future work must examine how this technique could fit into the larger character animation workflow and whether experts would prefer this technique over the accepted industry standards.

References

1. Balakrishnan, R., Hinckley, K.: Symmetric bimanual interaction. In: Proceedings of CHI, pp. 33–40 (2000)
2. Kim, M., Hyun, K., Kim, J., Lee, J.: Synchronized multi-character motion editing. ACM Trans. Graph. 28(3), 79:1–79:9 (2009)
3. Kipp, M., Nguyen, Q.: Multitouch puppetry: creating coordinated 3D motion for an articulated arm. In: ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010, pp. 147–156. ACM (2010)
4. Min, J., Chen, Y.-L., Chai, J.: Interactive generation of human animation with deformable motion models. ACM Trans. Graph. 29(1), 9:1–9:12 (2009)
5. Neff, M., Kim, Y.: Interactive editing of motion style using drives and correlations. In: Proceedings of SCA, pp. 103–112. ACM (2009)
6. Owen, R., Kurtenbach, G., Fitzmaurice, G., Baudel, T., Buxton, B.: When it gets more difficult, use both hands: exploring bimanual curve manipulation. In: Proceedings of Graphics Interface 2005, GI 2005, pp. 17–24 (2005)
7. Zhai, S., Milgram, P.: Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In: Proceedings of CHI, pp. 320–327. ACM Press/Addison-Wesley Publishing Co. (1998)

RelicPad: A Hands-On, Mobile Approach to Collaborative Exploration of Virtual Museum Artifacts

Steven Neale[1], Winyu Chinthammit[1], Christopher Lueg[2], and Paddy Nixon[1]

[1] HIT Lab AU
[2] School of Computing and Information Systems,
University of Tasmania, Australia
{Steven.Neale,Winyu.Chinthammit,Christopher.Lueg,
Paddy.Nixon}@utas.edu.au

Abstract. In an ideal world, physical museum artefacts could be touched, handled, examined and passed between interested viewers by hand. Unfortunately, this is not always possible – artefacts may be too fragile to handle or pass around, or groups of people with mutual interests in objects may not be in the same location. This can be problematic when attempting to explain or make sense of the physical properties of artefacts.

To address these problems, we propose that direct manipulation of 3D content based on real-world interaction metaphors can help collaborators (both co and remotely located) to construct personal and mutual physical and spatial awareness of artefacts, while networked communication and collaboration allow for ideas and knowledge to be exchanged and shared.

We present our interpretations from two studies of *RelicPad*, a tablet-based application that allows users to manually manipulate museum artefacts and to 'point out' areas of interest to each other using 3D annotations, facilitating a mutual awareness of spatial properties and referencing during discussion.

Keywords: Museum artefacts, remote collaboration, tablet interfaces, 3D interaction techniques, virtual reality.

1 Introduction

Handling physical museum artefacts is acknowledged as being a powerful experience. As objects with meaning, artefacts help to create "strong personal connections" to pasts and people across time and culture, prompting thought, memory and understanding [1]. This makes it particularly useful to have artefacts at hand when discussing them or explaining them to other people. However, in situations where handling artefacts is not possible (objects being enclosed in display cases, too fragile or valuable to be handled, or located in a different part of the world), the vital spatial referencing and physical understanding that handling artefacts provides is lost.

The Queen Victoria Museum & Art Gallery (QVMAG) [2] in Launceston, Tasmania, who provided resources for our research, described to us an example scenario of museum curators considering the purchase of an artefact from an overseas museum.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 86–103, 2013.

With time, money, and effort at stake, it is hugely advantageous for curators to be able to understand the physical properties of artefacts before committing to a decision, even if they are not able to handle them in person. Other scenarios within the problem domain include the discussion of artefacts amongst online communities (digital museum visitors), between curators in different museums around the world, long distance lectures, and presentations.

This paper describes an initial design and evaluations of *RelicPad*, a tablet-based application that addresses this problem by facilitating physical, reality-based interaction with and collaborative discussion of virtual (3D representations of) museum artefacts. In collaboration with the QVMAG, a piece of nineteenth century scrimshaw (scrollwork, engravings, and carvings done in bone or ivory [3]) has been rendered in 3D for exploration and discussion using *RelicPad*. During a discussion, users:

- Manipulate the virtual museum artefact by using their hands to manually interact with the tablet.
- Leave 3D markers (referred to in this paper as 'interest points') on different parts of the virtual museum artefact.

These features allow users to manually manipulate a virtual museum artefact and to 'point' to specific areas of interest, despite not being able to see the actions and gestures of others. This allows users to construct their own awareness of an artefact's physical and spatial properties and to communicate spatial references for others.

We first describe our motivation for the research, and present some related work. Next, we describe the application and provide an overview of its functionality. We then introduce and present the results of our two studies, and offer a discussion of our findings. Finally, we draw conclusions based on our interpretations of the discussion and finish by outlining our future work.

2 Motivation

While replicating the experience of physically handling an artefact is impossible, our research presents an alternative approach to making sense of artefacts through exploration and collaborative discussion, based on 3D representations of their physical properties. As well as the loss of the tactile and physical understanding and sensation that comes from handling objects individually, it is also problematic for collaborators to convey gestural clues about the spatial relationships between themselves and the objects around them without mutual access to objects [4], making it difficult for individuals in different locations to maintain focus on them.

Our research is aimed at providing a usable application for a number of different museum user groups, including researchers, curators, educators, and visitors (both physical and online). Each of these groups has their own sets of motivations for engaging with artefacts - overseeing and caring for collections, organising educational programs, public service and community outreach, authenticating, evaluating and categorizing artefacts, presenting information to the public, or receiving information from the museum institution [5] [6] [7]. The unifying thread between users in these groups is that they all have the desire to engage with artefacts, and that they all have opinions, ideas, or knowledge to share, drawing on everything from culture and personal experience to specialized training or knowledge. We believe that this makes a usable tool for exploration and collaborative discussion of virtual museum artefacts an exciting prospect for users across the whole spectrum of museum user groups.

It has been suggested that task environments have an outer boundary of what is visible to the person(s) carrying out the task, known as the 'horizon of observation' [8]. Naturally, how an interactive system makes use of this boundary, particularly for collaboration, is going to have "consequences for the process of acquiring knowledge" [8]. However, there are many situations where the horizon of observation around an artefact is too limited to be able to fully explore or understand it. Artefacts themselves are often enclosed and cannot be seen or touched from all sides; if passed around between large groups (of museum visitors, for example), only smaller subgroups of people will have good access to it at any given time; online visitors may only have access to still images via museum websites; and in remote-collaborative situations, even if one collaborator has access to the artefact, the rest of the collaborators will be limited to what is described to them by voice, shown to them as still images, or in the case of videoconferencing held up in front of a camera.

RelicPad expands this 'horizon of observation' in ways not normally possible with existing technologies, supporting real-time exploration of a virtual museum artefact and collaboration with others, and allowing users to easily focus, communicate and discuss ideas and theories about artefacts in a 3D context. As a mobile, tablet-based application, *RelicPad* is also adaptable to numerous collaborative scenarios, from passing one or more tablets around a large group of museum visitors to remote collaboration by way of real-time networked interaction. Putting the task 'in the hands' using the mobile context makes the 'horizon of observation' or 'window' to the artefact applicable to different groups of people working across varied contexts of use.

3 Related Work

As outlined in the introduction, handling physical museum artefacts is a powerful experience, linking the handler to individuals and periods of time that they would otherwise have no tangible contact with [9]. This is largely down to the sensory qualities of artefacts – they are "a 3D experience", more tactile than a still image or a recording, and can be turned over and viewed from all sides [9], providing handlers with "significant insight and understanding of the physical and material aspects of objects" [1]. While simultaneously replicating all of the sensory qualities associated with handling artefacts is not a realistic aim, tablets provide a method of displaying a high-quality 3D rendering of a museum artefact, and of using different interaction techniques to view it from all sides.

3D scanning of museum artefacts is now "a practical reality" [10], providing a wealth of 3D museum content for exploration and discussion. The challenge is interacting with this content in a way that better represents the manual exploration that users expect from traditional object handling. Tangible interactions using devices such as tablets and touchscreens might better represent certain interaction concepts than a mouse is able to [11]. But while there has been plenty of research into interaction techniques for the manipulation of 3D virtual objects, their application to mobile devices such as smartphones and tablets remains a scarcely explored area.

Most interaction techniques developed “for stationary computers” are “not applicable” to such devices [12], which seems a missed opportunity considering that the rapidly improved display capabilities of smartphones and tablets have made the delivery of rich, interactive 3D content very achievable. The University of Virginia Art Museum’s (UVaM) *Interactive iPad Museum Catalog* [13] allows users to view a number of pre-selected artifacts from the museum’s collection as high-quality 3D visualizations, but interaction itself is limited to dragging the finger on a touchscreen to rotate around a single axis, and scaling. Explorations into manipulating 3D content on mobile devices have experimented with concepts such as ‘tilting’ devices using either computer-vision techniques [12] or built-in sensors such as gyroscopes, compasses and accelerometers [14], which have proven to be a promising alternative to more traditional 2D interaction or touch techniques for 3D content rotation tasks.

But regardless of whether touch or tilt techniques are being used, we believe that the techniques used to manipulate virtual representations of objects should be based “on the real world” [15]. When the user is already skilled at performing the actions that underpin the basic operation of the system, the “mental effort required” for that operation can be significantly reduced [15], and so a strong representational metaphor of object handling, requiring minimal thought and leaving users free to focus their attention on the physical nuances of the object (and collaborative discussion) rather than how to manipulate it, should be considered as part of the interaction technique.

As important as the manipulation of the virtual museum artefact are the interactions for supporting and organizing discussion between collaborators, particularly remotely. Physical objects can play an important role in collaboration. As well as using them to “complete their own activities”, collaborators often use objects to “coordinate [these activities] in real-time with the conduct of others” [16], and it has been suggested that collaboration “relies upon [collaborators’] mundane abilities to develop and sustain mutually compatible, even reciprocal, perspectives” of their environment and the objects within it [16]. Naturally, interaction with a virtual museum artefact needs to support similar processes of understanding and referencing between collaborators, who may well be connected remotely and unable to see each other.

Common tools for collaboration and discussion (such as *Skype* or *Windows Live Messenger*) allow users to share and exchange files, view each others’ screens, send instant messages, and communicate in real-time using both voice and video. Existing research into remote collaboration systems has shown tags, metadata and annotations to be useful in mutually focusing attention for 2D data (e.g. text, images and video). In museum informatics, distributed systems for sharing information about artefacts and collections are a well-researched area in their own right, with various web [17] and multi-media [18] based approaches that encourage and facilitate the distribution of that museum data across as wide a museum community as possible.

Visual media in such systems are generally 2D, and interactive content (beyond basic text, images and video) is limited. 2D interactions such as clicking images and following hyperlinks reflect this and there are few examples of these principles being used in 3D contexts, leaving these technologies with a very fixed horizon of observation that can only be pushed so far. However, we now have such a wealth of available technologies for displaying and interacting with 3D content that interactive systems for remote collaboration could be making far better use of the visual channels available to them, particularly where the spatial referencing of 3D objects is concerned.

Annotations are a common feature of collaborative technologies, and increasing interactivity has seen annotations move from being solely about "managing data and metadata" to becoming "critical" resources in "supporting communicative practice" [19]. The *Vannotea* system used annotations as 'metadata stores' to enable "the collaborative indexing, browsing, annotation and discussion of [video] content between multiple groups at remote locations" [20], while the *Kinected Conference* sees annotations used to convey users' whereabouts in 3D, using video depth, audio cues and face-tracking algorithms to assign dynamic and interactive context tags to remote collaborators in a videoconference [21].

As well as creative and interactive uses of annotations for spatial referencing in remote collaboration, there are also examples of previous attempts to expand the 'horizon of observation' around 3D virtual content. *Lighthouse*, a remote-collaborative system for troubleshooting printer problems, used synchronized 3D representations of printers visible to both the customer and the troubleshooter, allowing troubleshooters to manipulate a shared pointer to highlight problems with the printer for the customer [22]. However, even systems like *Lighthouse* still use a mouse and keyboard (2D interfaces) as the input methods for interactions with 3D content, and so there is room to explore how collaboration, discussion and annotation can be integrated with 3D object manipulation to expand the horizon of observation around virtual museum artefacts for exploration and discussion by multiple (remote) collaborators.

4 Description of *RelicPad*

Research suggests that the actions generated by physical manipulation of tangible interfaces help to "draw up previous knowledge" and "generate important motoric representations to support other forms of representation" [11]. Based on the idea that manually manipulating virtual artefacts and marking interest points supports the exploration and discussion of objects (just as physically handling an object helps to build context) and supported by our review of the problem area and related work, our approach was to develop an initial prototype for *RelicPad* (described in this section) and then to identify, explore and evaluate design issues through a series of user studies (sections 5 & 6).

4.1 Overview of *RelicPad* Features

As well as the application to user groups from a variety of museum contexts outlined in our motivations, the clear benefits of being able to physically manipulate the interface influenced our decision to use a mobile, tablet-based platform as the interface technology for our research. Tablets also have a unified input and output space; the interaction and the resulting action happen in the same place (usually the hands), and so the observation viewpoint is the same as when handling a physical object.

RelicPad aims to provide a digital alternative to two key physical interactions – physically moving an object around in the hands, and pointing at different areas of objects. Taking into account the underlying aim of enabling users to share information about these 3D artefacts, *RelicPad* can be broken down into three fundamental elements that underpin the application:

- Manipulation (rotation and scaling) of the virtual museum artefact in 3D.
- Real-time marking of interest points in 3D space.
- Interactive conversation history.

These three elements offer digital representations of physical object handling, pointing, and organizing topics of discussion, enabling users to explore and discuss virtual museum artefacts in collaborative scenarios. Users can manipulate (rotate and scale) the virtual artefact using their hands to build up an understanding of it in 3D, can mark interest points to show each other where (and in what context) something interests them, and can refer to a conversation history to remind themselves of how conclusions were made and revisit earlier topics of interest. An example of how these elements come together in the interface can be seen in Figure 1:

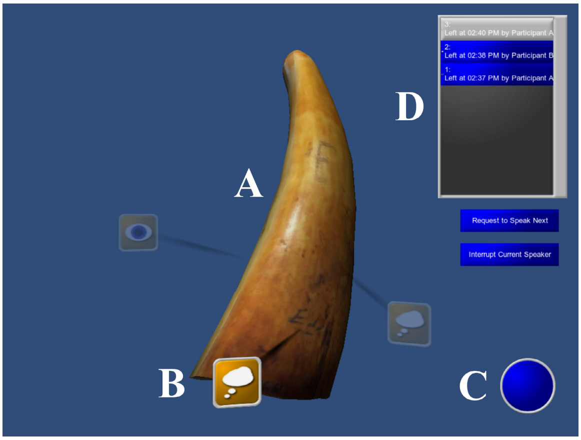


Fig. 1. RelicPad interface; 3D representation of artefact (A), interest point (B), rotation button or 'clutch' (C), and interactive conversation history (D)

Manipulation of the Virtual Museum Artefact in 3D

In order to expand the 'horizon of observation' around virtual museum artefacts in ways that current applications do not, it is important to ground these digital interactions in reality-based movements that are more akin to handling physical objects. It is largely the 3D experience of handling physical objects – turning them over, looking inside them, viewing them from all angles – that prompts thought and understanding, helping the handler to make sense of the object [1]. Finding an appropriate interaction metaphor that gives users freedom and control over the manipulation (rotation and scaling) of the virtual museum artefact is important.

Our research has looked at two interaction metaphors for manually manipulating 3D objects using a tablet as the interaction device – tilt and touch (Figure 2):

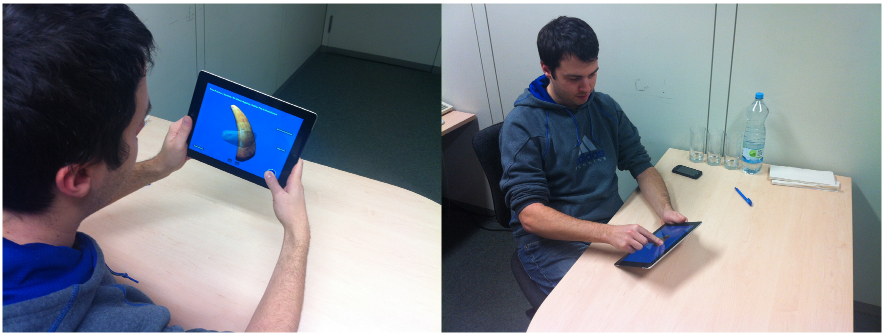


Fig. 2. Two interaction metaphors – tilt (left) and touch (right)

The tilt metaphor works by using the orientation of the tablet, given by the built-in accelerometer and gyroscope, to calculate the orientation of the 3D content, and represents the idea of using the hands simultaneously to grasp a physical object and rotate, twist, and turn it around. Visually, it carries a strong representation of moving a physical object with both hands. In contrast, the 'touch' metaphor works by orientating the 3D content according to the position of the user's finger(s) on the device's screen, representing something more akin to holding a physical object (in this case the tablet) in one hand, and using the other to rotate, twist or turn the object *within* that hand. Although more visually abstract than the tilt, this is a more familiar and traditional tablet interaction technique.

Section 6 describes the second study undertaken as part of this research, a comparison between four different techniques for rotating the 3D artefact (two techniques for each of the two interaction metaphors described above) and three different techniques for scaling the 3D artefact, and outlines each of these techniques in more detail.

Marking of Interest Points

With physical objects, people point at them to provide each other with a clear frame of reference and to clarify areas of interest. Pointing is a clear visual gesture that draws attention to something that is deemed to be interesting, and marking interest points with *RelicPad* supports discussion of virtual museum artefacts in the same way. Users attach interest points to virtual museum artefacts that other users can see for themselves in 3D space, in much the same way as people use pointing to guide others to what they want them to see during discussions about physical objects.

Marking interest points using *RelicPad* is achieved by tapping the screen on the area of the virtual museum artefact where the interest point is to be placed. As well as being a very familiar gesture to users of mobiles, tablets and touchscreens, the 'tap' gesture also resembles pointing in the physical and visual sense, making use of a single extended finger. Tapping the virtual museum artefact brings up a box menu with a choice of three possible 'context' icons (see *Interactive Conversation History* subsection below) plus a fourth 'cancel' icon in case the user decides not to leave the point after all. Tapping one of these icons leaves an interest point with the selected context in the desired location.

Above the choice of context icons is a text-field, allowing users to assign a keyword(s) to the interest point if they wish. Interest points themselves are offset slightly to the left or the right of the screen (depending on which side of the middle of the screen the user decided to leave the interest point) and linked to the virtual museum artefact by a line, so that when initially left they are not obscuring it.

Interactive Conversation History

The interest points left during a discussion come together to form an interactive 'conversation history'. An important classification tool for this conversation history is the context that can be used to define an interest when it is being marked. There are three context choices available, representing aesthetics (something about the way the object looks that the user finds interesting), geometry (something about the shape of the object that the user finds interesting), or meaning (something that the user feels provides clues about the cultural significance or idea behind the object).

In addition to (optional) user provided keywords, the name of the user who left each interest point and the time at which it was left, the classification of context (identifiable by the look of the interest point) allows for a historical record of the activity in discussions, conversations and collaborations to be kept, organized according to the interest points users have created. For collaborative use, this provides "a persistent record of interaction and collaboration" that can be easily referred back to [23], enabling users to revisit earlier interest points and "remind [themselves] of the process by which they reached previous interim conclusions" [19].

RelicPad's conversation history consists of a scrollable menu in the top-right hand corner of the screen that stores all of the interest points in a discussion, adding each new point at the top of the list. As well as the point's context icon, each interest point record in the conversation history also contains the keyword given (if any), the name of the user that left the point, and the time at which it was left. This interactive but relatively traditional two-dimensional list leaves the conversation history in the periphery of the user's attention until it is needed – users can focus on their exploration and discussion of the artefact and on the physical interactions used to operate other *RelicPad* features, but can refer back to earlier points of discussion when needed.

5 First Study - Communication and Collaboration Using *RelicPad*

Our first study aimed to get a feel for *RelicPad* as a user experience, and give an early indication of how useful it could be in context. The idea was to find out if participants could, by connecting to each other remotely using *RelicPad*, discuss an artefact together, share ideas about it, and draw each other's attention to its various features, in order to arrive at a conclusion about what the artefact is.

5.1 Experiment

We invited 22 participants, not selected based on any specific criteria, to take part in 11 collaboration sessions (2 'remote' collaborators discussing the virtual museum artefact with each other in each session). Rather than specifically recruit participants from any of the previously described museum user groups, we opted for participants with no prior knowledge of the artefact so that as opposed to relying on any pre-existing knowledge of what the artefact might be, they would use interest points and spatial referencing to collaboratively generate and share ideas and understanding. This scenario of 'exploring from scratch' is applicable to any of the museum user groups, even curators and researchers who regardless of knowledge and experience will still be presented with new and exciting objects to make sense of from time to time.

Participants were each given an iPad running *RelicPad* with our 3D model of the QVMAG's scrimshaw piece loaded as the virtual museum artefact. Prior to the experiment, we identified (in collaboration with QVMAG curators) 7 key features of the scrimshaw piece that we expected participants might be able to identify together. During the sessions, participants were encouraged to talk to each other, share ideas, and mark interest points in order to form conclusions about what the artefact is, where it came from, and the story behind it. We anticipated that our subsequent data collection would shed some light on how they had used *RelicPad's* features to do this.

Each experiment lasted for approximately 30 minutes – 10 minutes introduction and explanation, 10 minutes for the collaboration session, and 10 minutes to complete the subsequent questionnaire. Participants were seated opposite each other at adjacent desks, with a large screen obscuring their views of each other – they were able to hear each other as they would during a phone call or videoconference, but were unable to see each other's movements, gestures, or interactions with the tablet. The investigator was seated to the side of the participants, with a clear view of both of them for making observations.

Data was recorded as a sequential analysis of activity and interaction between the participants, accompanied by questionnaires. The sequential analysis consisted of three elements - an audio recording of each discussion session, a log of interactions recorded by the iPads and stored in an online database (recording when an interest point was left, as well as the context and any keywords that were given to it), and handwritten observations made by the investigator of notable exchanges between the participants during each session. These three elements were time synchronized to paint a picture of how and why participants came to different conclusions using *RelicPad*, and how *RelicPad* features were used in this process.

The questionnaires, given to each participant after the discussion was over, consisted of two sets of ten questions adapted from the System Usability Scale (SUS), a widely accepted, "simple, ten-item scale giving a global view of subjective assessments of usability" [24], the answers from which can be used to generate a single number representing "a composite measure of the overall usability of the system being studied" [24]. The first ten questions focused on the five basic factors of usability (effectiveness, efficiency, safety, utility and learnability) while the second ten questions focused on five core components of user engagement – identity, adaptivity, narrative, immersion and flow [25]. The remaining questions were a mixture of closed and open-ended questions seeking additional feedback on how users felt about different *RelicPad* features, what they felt they had learned from the collaboration, and their overall impressions of *RelicPad* as an experience.

During each session, participants were each given an iPad running a different version of the prototype. One participant was always using the tangible 'tilting' interaction method for rotating the virtual museum artefact, while the other participant was always using a more traditional 4-button 'directional pad' for rotating it. The tilt rotation was rate-controlled (the virtual museum artefact rotates faster according to how far the tablet is tilted), with a rotation button or 'clutch' held to initiate and released to cease rotation. The 4 directional buttons were used to rotate the artefact up or down around the x-axis, or left and right around the y-axis.

Neither participant knew that their collaborator was using a different technique to rotate the virtual museum artefact. The idea behind this was to evaluate whether or not there was any difference in how usable or engaging participants found *RelicPad* based on whether or not they used a physical, tangible rotation method, or a more traditional 2D interaction technique.

5.2 Results

Observations and audio recordings showed that during the 11 collaboration sessions, pairs of participants were able to arrive on average at 4.45 of the previously identified 7 conclusions about the scrimshaw. Certain things were particularly well noticed by participants, including faded lettering on the back of the scrimshaw (a focal point of the discussion in 10 of the 11 sessions) and the cracks in and discoloring of the scrimshaw (focal points of the discussion in 9 of the 11 sessions).

The combination of observations and activity logs gave an indication of how interest points are used during collaboration. On average, 12 interest points are left during each of the 11 sessions. Participants on average will specifically refer to 48% of these interest points, whether that be to tell their collaborator that they have just left (or are about to leave) the interest point, to tell them exactly what part of the artefact it is being attached to, or to give them directions to help them find out where it is. Participants also on average ask their collaborator for clarification or an explanation of 13% of interest points; asking where an interest point is, what it was supposed to be attached to, or whether they are looking at the correct one.

Questionnaire responses revealed a positive response to *RelicPad*. Although 8 out of the 22 participants unfortunately did not give a definitive answer to the question, 12 of the 14 participants who did respond described *RelicPad* as a positive experience. When asked to provide additional comments about the experience, 13 out of the 22 participants commented on the application being good for remote collaboration, conversation, and positional referencing, with 10 participants specifically referring to its suitability for interacting with artefacts in museum or educational contexts.

The first ten and second ten sets of questions were used to calculate single numbers for usability and user engagement, as described in the previous section. In both cases the mean scores were good – 70.6 for usability and 73.07 for engagement. Participants were asked to rate *RelicPad* interactions based on how frequently they felt they used them (1 being very infrequently and 5 being very often). Results showed that on average participants felt that they had marked interest points very frequently (4.05), often added keywords to them (3.09), and occasionally chosen to change the current interest point (2.77) (to refer back to something from earlier in the discussion).

Participants were asked to rate how positive or negative they felt (5 being very positive) about the three fundamental elements that, as previously mentioned, combine to create the overall *RelicPad* experience – rotation of the virtual museum artefact in 3D, real-time marking of interest points, and the interactive conversation history. On average, the marking of interest points was the most popular feature among the participants (3.95), while rotation of the virtual museum artefact (3.82) and the interactive conversation history (3.32) were also both well received.

6 Second Study – Techniques for Manipulating Virtual Museum Artefacts

Separating the results of the first study according to which rotation technique was used, we found that there was no significant statistical difference between the tilt and directional button techniques, and conclusions about which technique was more usable or affected the experience differently to the other could not be drawn. Two-tailed independent t-tests proved this to be the case for basic usability ($t=-.981$, df 15.6, $P=.342$), user engagement ($t=-.612$, df 20, $P=.547$), ease of rotation ($t=-1.64$, df 20, $P=0.116$) and overall impressions of rotation ($t=-2.01$, df 20, $P=0.058$).

However, ‘touch’ and ‘drag’ interaction techniques (well established interactions with tablet devices) and a direct (one-to-one) mapping of tilt to rotation (as opposed to rate-controlled) were both suggested as ways of improving the rotation technique. This was interpreted as an indication that while the rate-controlled tilt was not significantly worse than a more traditional 2D interaction, it was not significantly better, leaving room for further exploration of interaction techniques and metaphors that might better represent the physical exploration associated with object handling.

Our second study compared four rotation techniques for manipulation of a 3D virtual object – two making use of the ‘tilt’ metaphor and two making use of the ‘touch’ metaphor (both metaphors are outlined in section 4). The two ‘tilt’ techniques were a rate-controlled tilt (the angle at which the device is tilted defines the speed at which the virtual object rotates) and a direct-mapping tilt (the virtual object’s rotation follows the angle at which the device is tilted exactly). The two ‘touch’ rotation techniques were a ‘virtual trackball’ implementation (enclosing the virtual object in a sphere which is dragged with a single finger in order to rotate, described as a ‘virtual sphere’ in [26]) and a ‘multi-touch’ approach (dragging with a single finger to rotate the artefact on the x and y axes, and rotating two fingers clockwise or counter-clockwise in order to rotate the artefact on the z axis).

Three techniques for scaling the virtual museum artefact were also compared – a ‘plus & minus buttons’ approach (holding one of two buttons to increase or decrease the scale of the artifact), a ‘slider bar’ approach (continuous dragging between the two end points on a slider bar widget to increase or decrease the scale), and a ‘multi-touch’ approach (increasing or decreasing the distance between two fingers, also known as ‘pinching’ and ‘spreading’, to increase or decrease the scale of the artefact).

6.1 Experiment

12 participants took part in a set of ‘object-matching’ trials, using the different rotation and scaling techniques to match a virtual museum artefact (the scrimshaw piece from the first study) with a semi-transparent target orientation of the same artefact. The rotation techniques were organized according to a balanced Latin-square design (to minimize biases caused by practice or fatigue), and for each technique there were 18 trials – 6 with each of the three scaling techniques. Of those 6 trials there were 2 simple, 2 medium and 2 complex rotation difficulties. Difficulty was defined as being whether matching the virtual artefacts required rotation on one (simple), two (medium), or all three axes. This makes a total of 72 trials per participant.

For each trial, participants’ speed (time taken to complete the trial), rotation error (difference between virtual and target artefacts in degrees) and scale error (difference in size of the virtual and target artefacts represented as vectors) were recorded. Participants were asked to think about each trial in terms of both speed and accuracy and move onto the next trial as soon as they were happy. However, to keep things moving and prevent the experiment from taking an unreasonable amount of time, participants were asked to move onto the next trial after around 90 seconds.

In between each rotation technique (every 18 trials) participants were asked to answer some questions about that technique, and after all of the trials had been completed to answer questions looking back on all of the rotation and scaling techniques together. These questions asked participants to rate different aspects of the various techniques numerically, or to provide a few short sentences on the techniques.

The design of the experiment was based on a number of similar object-matching experiments from past research into techniques and technologies for 3D object manipulation, from early explorations with desktop-based VR systems [26] [27] [28], to more recent approaches to manipulating 3D content with mobile devices [12] [14] and touch displays [29].

6.2 Results

Table 1 shows means and standard deviations (SD) for how quickly and how accurately participants were able to complete the trials, on average, using each of the four rotation techniques – Rate-Controlled (RC) Tilt, Tilt with Direct Mapping (DM), the Virtual Trackball (VT) touch technique, and Multi-Touch. Touch techniques performed better than tilt techniques, being both quicker and more accurate to use:

Table 1. Speed and accuracy of the four compared rotation techniques

Averages	Tilt (RC)	Tilt (DM)	Touch (VT)	Multi-Touch
Time Taken (seconds)	58.03 (SD 13.32)	51.75 (SD 14.59)	49.72 (SD 18.06)	47.99 (SD 10.84)
Rotation Error (degrees)	12.22 (SD 9.92)	11.81 (SD 14.07)	6.38 (SD 3.91)	7.47 (SD 4.49)

Analysis of variance showed the effects of rotation technique on both the time taken to complete the trials (F = 3.966, df 3, 68, P = 0.011) and on the average rotation error (F = 4.278, df 3, 68, P = 0.008) to be statistically significant.

A comparison of how quickly and accurately trials could be completed, on average, using each of the scaling techniques is shown in table 2 (means and standard deviations (SD)). Here, the touch technique performed better than its 2D counterparts, proving to be both the quickest and the most accurate technique:

Table 2. Speed and accuracy of the three compared scaling techniques

Averages	+/- Buttons	Slider Bar	Multi-Touch
Time Taken (seconds)	56.08 (SD 9.73)	50.91 (SD 7.34)	48.62 (SD 10.39)
Scaling Error (size difference)	7.03 (SD 3.93)	6.63 (SD 4.07)	4.65 (SD 1.70)

Analysis of variance shows the effects of scaling technique on the time taken to complete the trials to be of statistical significance (F = 3.857, df 2, 69, P = 0.026), but the effects of scaling technique on the average scaling error were minimal.

Comparing all combinations of rotation and scaling techniques together supports these results, with the combination of Multi-Touch rotation and Multi-Touch scaling resulting in the fastest average trial completion time at 41.61s (see Figure 3), the smallest average scaling error (3.93), and the second-smallest average rotation error (6.10 degrees). The worst combination was Tilt (Rate-Controlled) rotation with Plus & Minus Buttons for scaling, which gave the slowest average task completion time (64.65), largest average rotation error (19.69), and largest average scaling error (8.66). However, analysis of variance showed that apart from the Tilt (Rate-Controlled) rotation with Plus & Minus Buttons for scaling being by far the poorest combination of techniques, the differences between the rest of the combinations were minimal, and not of statistical significance.

The questionnaires asked participants to rank the four rotation techniques in order from 1 (best) to 4 (worst) in relation to a number of different criteria, including: ease of rotation; perceived accuracy of rotation; perceived speed of trial completion; enjoyment; understanding of the movement of the artefact in 3D; perceived control over the artefact; and favourite technique. On average, Multi-Touch recorded the lowest (best) average (between 1.42 and 1.75) for all of the criteria, while Tilt

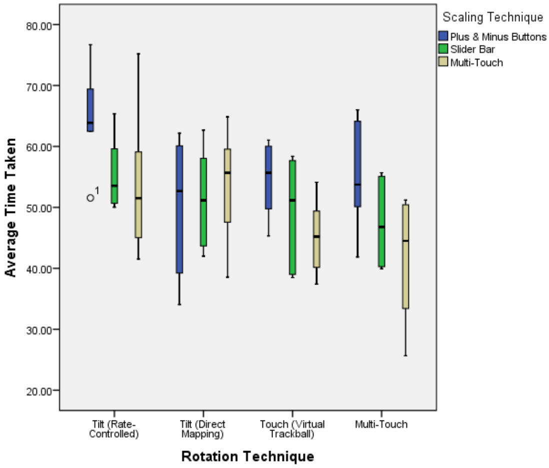


Fig. 3. Graph showing average time taken to complete trials (in seconds) for each rotation technique, separated by scaling technique

(Rate-Controlled) recorded the highest (worst) average (between 2.92 and 3.33) for all of the criteria. Participants were also asked to rank the three rotation techniques in order from 1 (best) to 3 (worst) in relation to the same criteria – Multi-Touch scaling also recorded the lowest (best) average (between 1.42 and 1.92) for all of the criteria.

7 Discussion

Based on the results of the two studies, we have come to our own interpretations of how users have received our application, how applicable it is to the described problem domain, and which interaction metaphors give users the most control over the manipulation of virtual museum artefacts. We found that participants used interest points to share spatial references, that they could collaboratively construct solid ideas of what the artefact was, and that they felt positive about being able to rotate the virtual artefact in 3D, suggesting that *RelicPad* can facilitate exploration and collaborative discussion of virtual museum artefacts. We also found that touch interaction techniques leave users feeling in control of the manipulation of the virtual museum artefact.

In the first study, marking interest points proved to be the most popular and commonly used feature of *RelicPad*. Almost half of the interest points left were directly referred to by participants verbally, usually to explain what the interest point refers to as it is being left ("that there, it's a parasol"; "and you can see in this section here that it's lighter"; "down below there, that marking"; "it's broken, on your left side – I'll point it out for you"). This indicates that participants used the interest points for the purpose we intended – 'pointing out' interesting areas or features of the artefact to assist each other in maintaining a shared spatial reference of the topic of discussion.

Importantly, participants noticed and discussed most of the things that were identified prior to the experiment as providing the history, story and significance of the scrimshaw piece. This indicates that participants not only enjoyed marking interest points, but that they used them for the intended purpose and that this bore results in terms of the collaboration. This is especially interesting given that to most participants it was not obvious from the start that the artefact was scrimshaw; it was through discussion and theorizing with each other that they were able to come to conclusions about what it was or wasn't. It did appear, however, that there was a tendency to notice visual stimuli, however difficult to see, but not always to fully establish what they represented. The lettering on the back of the scrimshaw was discussed at length in 10 out of the 11 sessions, but in only 3 of those 10 was it agreed that they could be the initials of the creator (or the recipient) of the scrimshaw.

The first study compared the rate-controlled tilt technique for rotating the virtual museum artefact against a directional button widget. With the initial design, the tilt was seen as being a strong visual representation of moving a museum artefact with the hands, but the difference in rotation method used seemed to have little impact on the usability of the application or what participants were able to made of the experience of interacting with the virtual artefact. The second study looked to explore alternative interaction metaphors for manipulating the virtual artefact, and to see what happened when scaling was brought into the picture as well as rotation.

Object matching trials comparing combinations of the different rotation and scaling techniques showed that touch techniques, particularly 'multi-touch' techniques, performed significantly better and were more popular than tilt techniques. Virtual objects could be rotated and scaled more quickly using touch techniques, and also with more accuracy. Participants also reported enjoying using them more, having a better understanding of how the virtual object moves in 3D using them, and crucially having more control over the virtual museum artefact.

A number of factors may contribute to this – tablets have weight, and physically rotating them for long periods of time was tiring for some participants. The issue of viewpoint may also play a part here – with touch techniques the observation viewpoint is fixed while the hands manipulate the virtual objects (more akin to kind of viewing angles associated with physical object handling), while with the tilt it could be thought of as being the opposite, with the hands moving the observation viewpoint as opposed to the virtual object itself.

We interpret this, coupled with the ability to accurately 'fine-tune' the position of the virtual artefact, as an indication that while the tilt metaphor carries a stronger visual representation of moving a physical object with the hands, the touch metaphor gives the user more control. By leveraging interactions styles that are familiar to many users, touch techniques and the control they provide leave users free to focus their attention on the content and the collaboration itself, rather than the techniques required to interact with the virtual artefact. As such, we believe that the touch metaphor provides a more reality-based representation of natural and relatively thoughtless interaction with the hands for exploring the spatial properties of (virtual) artefacts.

8 Conclusions

Our interpretation of results shows that using interest points to draw attention to notable features of a virtual museum artefact helps users to understand and clarify areas of particular interest, and that over the course of a collaborative discussion these interest points are used to drive conversations and exchange ideas and theories. Participants in our study appeared to relish the ability to use the touch-tapping interaction to 'point' at artefacts and to highlight areas of interest for each other.

Comparing different metaphors for the manipulation of virtual museum artefacts suggested that established tablet interaction techniques such as multi-touch are the most efficient and enjoyable for controlling the manipulation of virtual museum artifacts on a tablet, leveraging skills that users are already familiar with to comfortably control the 3D experience and freeing them to focus their attention on other aspects of the system, such as marking and discussing points of interest.

We interpret this as an indication that for interaction techniques to feel 'reality-based', familiarity and control (moving an object in the hands comes naturally, as do touch techniques when using tablets) are perhaps more important than a strong visual representation (tilting the tablet looks more like handling an object, but it isn't the first thing people think to do with a tablet). This suggests an interesting compromise between how techniques look and how they feel when designing interactions.

As well as being well received as a positive experience by many participants, feedback showed that many of those participants saw the suitability of the application in museum contexts, specifically commenting that the application was good for remote collaboration and spatial referencing. *RelicPad* was able to expand each participant's horizon of observation in relation to both the virtual museum artefact itself and also to the focal points of each other's attention, and was used by participants to good effect. With the right interaction metaphors in place, it offers a promising solution to the scenarios presented in the introduction and motivation sections, and the example scenario provided by staff at the QVMAG.

8.1 Future Work

The experiments highlighted a number of interesting directions for the design of *RelicPad*. Many felt that the interest points themselves could have been more interactive – navigated to by touch, editable (renaming and deletion), more easily re-identifiable (either by keyword or by participant-based colour coding), and linked together to form a conversation thread. Combining the elements of interactivity from the conversation history into the interest points themselves may encourage collaborators to go beyond what they see of artefacts and to discuss further what those things might represent, giving spatial reference not only to areas of interest but to topics of interest as well.

In future experiments we also plan to evaluate whether the different museum user groups (or combinations of these groups) outlined earlier in the paper have different levels of success than others in coming to conclusions about artefacts whilst using the system. We will also compare the outcomes of collaborations made using *RelicPad* against those of face-to-face, co-located collaborations with the physical artefact to hand, and of existing remote-collaborative chat and video-conferencing alternatives such as *Skype*.

Acknowledgements. Thanks go to Andrew Johnson and Jon Addison from the QVMAG for their enthusiasm and for permission to use one of their artefacts during the experiment, and to the Human-Computer Interaction research group at the University of Konstanz, Germany for help and resources during the second study. Thanks also to Bruce Andrews of the HIT Lab AU for his help with programming and development.

References

1. Boyes, A., Cousens, C. In: the Hand - The Effect of Museum Handling Sessions on Student Learning. In: Networks, Art Design Media Higher Education Academy Subject Centre, University of Brighton (2012)
2. Queen Victoria Museum & Art Gallery, http://www.qvmag.tas.gov.au/qvmag/
3. Wikipedia. Scrimshaw, http://en.wikipedia.org/wiki/Scrimshaw
4. Everitt, K.M., Klemmer, S.R., Lee, R., Landay, J.A.: Two Worlds Apart: Bridging the Gap Between Physical and Virtual Media for Distributed Design Collaboration. In: Conference on Human Factors in Computing Systems, pp. 553–560. ACM Press, New York (2003)
5. Bureau of Labor Statistics, Curators, Museum Technicians, and Conservators, United States Department of Labor (2012), http://www.bls.gov/ooh/Education-Training-and-Library/Curators-and-museum-technicians.htm
6. National Museum of Australia, Research and Museology, National Museum of Australia, http://www.nma.gov.au/history/research
7. Simon, N.: What's a Virtual Visitor Worth?, Museum 2.0 (2008), http://museumtwo.blogspot.de/2008/12/whats-virtual-visitor-worth.html
8. Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge (1995)
9. Mastoris, S.: Learning Through Objects. Group for Education in Museums, http://www.gem.org.uk/res/advice/ball/res_lto.php
10. Smithsonian Institute, MCI Imaging Studio, Smithsonian Museum Conservation Institute (2012), http://www.si.edu/MCIImagingStudio/3DScanning
11. Manches, A., Price, S.: Designing Learning Representations Around Physical Manipulation: Hands and Objects. In: 10th International Conference on Interaction Design and Children, pp. 81–89. ACM Press, New York (2011)
12. Henrysson, A., Marshall, J., Billinghurst, M.: Experiments in 3D Interaction for Mobile Phone AR. In: 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia, pp. 187–194. ACM Press, New York (2007)
13. Put 3D Objects at your Visitors' Fingertips: UVaM on the iPad, Idea.org (2011), http://www.idea.org/blog/2011/11/29/put-3d-objects-at-your-visitors-fingertips-uvam-on-the-ipad/
14. Kratz, S., Rohs, M., Guse, D., Müller, M., Bailly, G., Nischt, M.: PalmSpace: Continuous Around-Device Gestures vs. Multitouch for 3D Rotation Tasks on Mobile Devices. In: International Working Conference on Advanced Visual Interfaces, pp. 181–188. ACM Press, New York (2012)

15. Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-Based Interaction: Unifying the New Generation of Interaction Styles. In: Conference on Human Factors in Computing Systems, pp. 2465–2470. ACM Press, New York (2007)
16. Hindmarsh, J., Fraser, M., Heath, C., Benford, S., Greenhalgh, C.: Object-Focused Interaction in Collaborative Virtual Environments. ACM Transactions on Computer-Human Interaction 7, 477–509 (2000)
17. Goodman, C., Schmitz, P., Sheppard, D., Clark, C.: Architecting CollectionSpace: A Web-Based Collections Management and Information System for 21st Century Museum Professionals. In: Museums and the Web, Archives & Museum Informatics, Toronto (2010)
18. Oomen, J., Baltussen, L.B., van Erp, M.: Sharing Cultural Heritage the Linked Open Data Way: Why You Should Sign Up. In: Museums and the Web, Archives & Museum Informatics, Toronto (2012)
19. Fraser, M., Hindmarsh, J., Best, K., Heath, C., Biegel, G., Greenhalgh, C., Reeves, S.: Remote Collaboration Over Video Data: Towards Real-Time E-Social Science. Computer Supported Cooperative Work 15(4), 257–279 (2006)
20. Schroeter, R., Hunter, J., Kosovic, D.: Vannotea – A Collaborative Video Indexing, Annotation and Discussion System for Broadband Networks. In: Knoweldge Markup and Semantic Annotation Workshop, 2nd International Conference on Knowledge Capture. ACM Press, New York (2003)
21. DeVincenzi, A., Yao, L., Ishii, H., Raskar, R.: Kinected Conference: Augmenting Video Imaging with Calibrated Depth and Audio. In: Conference on Computer-Supported Cooperative Work, pp. 621–624. ACM Press, New York (2011)
22. O'Neill, J., Castellani, S., Roulland, F., Hairon, N., Juliano, C., Dai, L.: From Ethnographic Study to Mixed Reality: A Remote Collaborative Troubleshooting System. In: Conference on Computer-Supported Collaborative Work, pp. 225–234. ACM Press, New York (2011)
23. Stahl, G., Koschmann, T., Suthers, D.D.: Computer Supported Collaborative Learning. Cambridge University Press, Cambridge (2006)
24. Brooke, J.: SUS - A Quick and Dirty Usability Scale. Usability Evaluation in Industry, pp.189–194 (1996)
25. Benyon, D.: Designing Interactive Systems – A Comprehensive Guide to HCI and Interaction Design. Pearson, Harlow (2010)
26. Chen, M., Mountford, S.J., Sellen, A.: A Study in Interactive 3-D Rotation Using 2-D Control Devices. Computer Graphics 22(4), 121–129 (1988)
27. Hinckley, K., Tullio, J., Pausch, R., Proffitt, D., Kassell, N.: Usability Analysis of 3D Rotation Techniques. In: 10th Annual ACM Symposium on User Interface Software and Technology, pp. 1–10. ACM Press (1997)
28. Poupyrev, I., Weghorst, S., Fels, S.: Non-Isomorphic 3D Rotational Techniques. In: Conference on Human Factors in Computing Systems, pp. 540–547. ACM Press (2000)
29. Martinet, A., Casiez, G., Grisoni, L.: The Effect of DOF Separation in 3D Manipulation Tasks with Multi-Touch Displays. In: 17th ACM Symposium on Virtual Reality Software and Technology, pp. 111–118. ACM Press (2010)

Funneling and Saltation Effects for Tactile Interaction with “Detached” Out of the Body Virtual Objects

Jaedong Lee, Sangyong Lee, and Gerard J. Kim*

Digital Experience Laboratory,
Korea University, Seoul, Korea
{jdlee,xyleez,gjkim}@korea.ac.kr

Abstract. In a previous work, we confirmed the existing effects of “Out of the Body” tactile illusion for virtual and augmented objects through funneling and saltation. However, it required a virtual imagery to be attached to the user for directly extending one’s body. This paper aims at investigating similar phantom tactile sensations exist when the virtual object is visually detached from the user’s body. Two usability experiments were conducted to verify the hypothesized phantom tactile effects: one for funneling and the other, saltation. Our results have shown that in addition to the perception of the phantom sensations with the “detached” visual feedback, the interaction experience was significantly enriched (vs. when without explicit visual feedback). We also discovered for the first time that for funneling, phantom sensations can be elicited without any visual feedback at all. The findings can be applied to the tactile interaction design using minimal number of actuators on a variety of media platforms including the mobile, holography and augmented reality.

Keywords: Phantom sensation, Illusory feedback, Funneling, Saltation, Vibro-tactile feedback, Multimodal feedback.

1 Introduction

Tactile feedback has become almost indispensable in improving interaction experience. Realization of tactile feedback by using vibration devices is one inexpensive and practical method. However, due to its size and mechanics, a single vibrator scheme is most often employed and it is only able to convey simplistic on-off type of events. Instead, as a way to improve the tactile experience, a more advanced form of vibro-tactile feedback most often involves an array of vibrators [6, 7] that brings about mechanical and cost complications and a constraint that a relatively significant area of the body has to be in full contact with the array.

One way that researchers have considered to overcome this problem is to create illusory (or pseudo) tactile feedback and combining it with the corresponding visual or auditory feedback [19]. For example, in a previous work, we confirmed the existing effects of “Out of the Body” tactile illusion for virtual and augmented objects through

* Corresponding author.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 104–121, 2013.

saltation and funneling. However, it required a virtual imagery to be attached to the user for directly extending and connecting one's body. In other words, this is undesirable because the user must have both hands/fingers in contact with the virtual object (see Figure 1). Thus, this paper aims at reproducing the same "Out of the Body" phantom sensation effect with "floating" (i.e. no virtual extension) dynamic virtual object detached from the body (see Figure 3). In particular, we consider two phantom tactile sensation phenomena, namely funneling [4] and saltation [14].

Funneling (Saltation) refers to the illusory tactile sensation occurring away from the actual places of timed (simultaneous) vibratory stimulations. The intended location of the phantom sensation can be changed by modulating the intensity (Funneling) or inter-stimulus time interval (Saltation). Funneling and saltation have often been applied to reduce the number of tactile actuators in tactile interaction design [18, 35]. Recently, researchers have discovered such phantom sensations can be extended to the "Out of the Body" [27] and for "Out of the Body" virtual objects [24] (see Figure 1), thus making it possible to generate phantom tactile sensations as if coming from an external object (both real and virtual). Possible applications of such a phenomenon are shown in Figure 2.

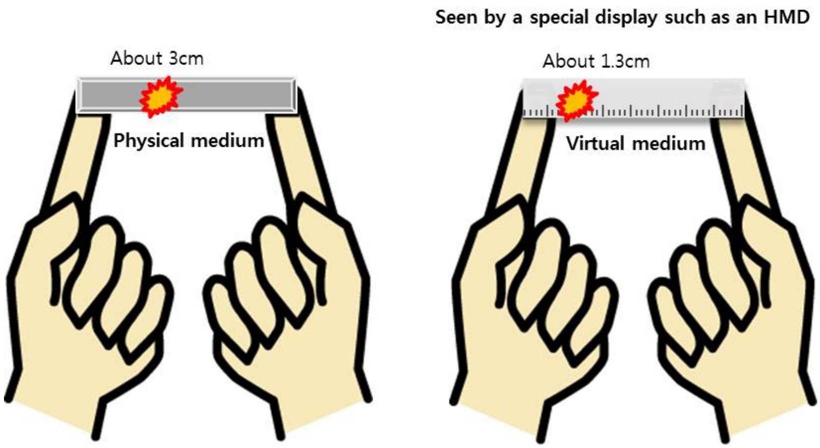


Fig. 1. The concept of "Out of the Body" tactile experience from a hand-held physical or virtual medium. A medium for extending and connecting the body parts is required, virtual or real. The phantom sensation is more evident with the physical medium (e.g. felt at 3cm from the left) than the virtual (felt at 1.3cm with the same stimulation) [24].

It has been found that the extent or the controllability of the effect is diminished when extended to "Out of the Body" and even more so when a virtual object is used as the medium extending one's body [24, 27]. Consequently, we seek and experimentally investigate the possible synergistic effects by associating it with "dynamic" visual feedback to improve the tactile experience and controllability, possibly even without the medium (real or virtual) that connects the body parts (i.e. tactile interaction with virtual objects completely detached from the body). If validated, such a phenomenon can be applied to tactically interacting with holographic objects hovering in the air (Figure 2(b)).

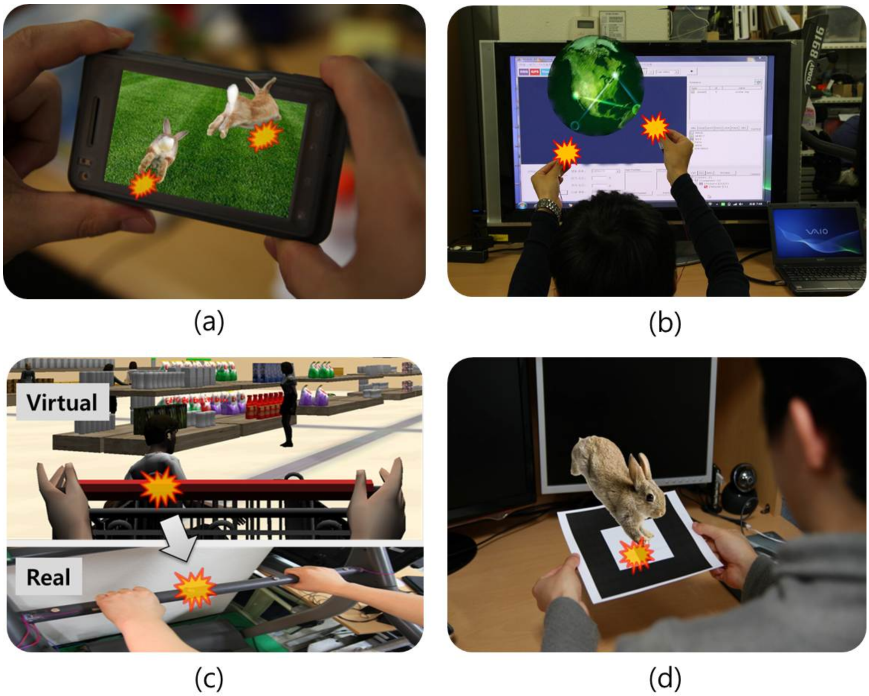


Fig. 2. Possible applications of the "Out of the Body" phantom tactile sensation: Two handed/fingered interaction and feeling tactile sensations as if coming from the middle of the (a) mobile device, (b) hovering holographic-virtual imagery, (c) indirectly from a virtual object in a monitor and (d) an augmented marker (e.g. seen through a head mounted display)

This paper is organized as follows. In the next section, we first review previous research literatures related to phantom tactile sensation, multi-sensory integration and their application to practical interaction design. Then, we describe the two validation experiments and report the results. Finally, we conclude the paper with a discussion and directions for future research.

2 Related Work

Funneling and saltation are the two major perceptual illusion techniques for vibro-tactile feedback. Funneling refers to stimulating the skin at two (or more) different locations simultaneously with different amplitudes and eliciting phantom sensations in the space between [1, 4]. Several researchers have applied this phenomenon to human interfaces [3, 19, 28, 31, 32]. For instance, Hoggan et al. experimented with using three vibrators on a mobile device to emulate a tactile progress indicator [18]. Tan et al. applied saltation to implement a tactile chair using a 3 x 3 tactile array for a pattern recognition application [38].

Miyazaki has discovered the saltation could be extended to body-worn (e.g. hand-held) objects and to create "Out of the Body" tactile experience [27]. Furthermore, Lee et al. has confirmed the same phenomenon existed for virtual objects but with reduced effects and less precise controllability [24]. Other researchers have investigated different ways and effects to apply phantom sensation by employing different stimulation interpolation methods [1], varying the values of ISI's [10, 15, 16], relative vibration amplitudes [1, 31], stimulation duration and frequency [10, 37], inter-stimulation distances [32], and even applying saltation to non-continuous skin (e.g. from the right arm to the left, fingertip to fingertip) [12, 40]. Note that in our previous work, only minimal static visual feedback was used, namely the virtual "ruler" representing only the medium bridging the two body points (rather than the actual visual representation of the tactile event) at which the actual vibratory stimulations were given, to recreate the original "Out of the Body" phenomenon. No detailed studies have been reported regarding phantom tactile feedback with dynamic or detached visual feedback. Also note that aforementioned works [3, 18, 19, 28, 31, 32, 38] that have applied tactile phantom sensations to human interfaces did it so directly to the human skin and had not investigated the use of the "Out of the Body" phenomenon nor the issue of thereby minimizing the number of tactile actuators.

Interestingly, Flach et al. [13] and Kilgard [21] have found that the phantom sensation was much influenced by the subject's focus of attention, anticipation and/or the line of sight. This strongly hints the possibility of further synergistic effects with more apparent visual effects associated with the intended phantom sensation.

In fact, the synergistic sensory integration is not new. It is generally accepted that multi-sensory feedback is additively helpful to interactive task performance [23]. This is only true provided when the respective modality feedback is consistent in its content and timing with one another [30]. Many synergistic multimodal interaction systems have been devised and studied employing gestures [5], voice [11], proprioception [26], speech/audio [17], and force feedback [34]. Aside from just improving task performance, multisensory interactions can also modify user perception, as illustrated by the famous McGurk effect. The McGurk effect is a perceptual phenomenon in which vision alters speech perception [25]. Simple visual tricks can easily alter the body image that is created by the proprioceptive sense [33]. Although the best known cross modal effects are those of vision influencing other modalities, visual perception can be altered by other modalities as well [39].

3 Experiment I: Effects of Funneling with "Detached" Visual Object

3.1 Purpose and Hypothesis

In the first experiment, we have compared the tactile experience of funneling for a virtual object, between (1) when it was associated with a (dynamic) visual presentation detached from the body and (2) when no visual presentation is given at all (as a reference). In a usual application setting, virtual objects will normally be rendered without a part that visually extends or connects user's body parts (Figure 2). Our interest is first to assess whether funneling elicits phantom sensation, its extent and effects to the overall interaction experience.

It is well expected that no dislocated phantom tactile experience will be elicited without any visual feedback (Figure 4, bottom right). We still test for it as a base case. Also it has been shown through prior research [27] that singular vibration (e.g. with a single vibrator) cannot create any localized phantom sensation for "virtual" object external to the body either. To reiterate, we are interested in and hypothesize the existence and quality of the phantom tactile sensation of the virtual objects even when it does not directly attached to the user body. We also expect that results for the detached "Out of the Body" virtual objects with the hypothesis that results will mostly extend to detached "Out of the Body" physical objects

3.2 Experimental Design and Set Up

To create the funneling based phantom sensation, the user was given simultaneous tactile stimulations to one's two index fingers, one in the right and the other in the left. The two fingertips were tracked using small markers (25mm x 25mm) by a head mounted camera and a "detached" augmented reality video imagery was presented to the user through a 47 inch monitor (nominal viewing distance: ~60cm). As for the augmented visual feedback, a small moving object ("a bouncing basketball") was rendered at the intended location of sensation between the two fingers (see Figure 3).

The OSGART [29] was used to recognize/track the small markers and generate the augmented video imagery. A webcam was worn on the head mounted fixture to produce a view close to one according to the actual line of sight. The user was asked to maintain a nominal distance (8cm) between the fingers (using an 8cm wide marking on the table), but was allowed small movements for natural and comfortable interaction. The inter-finger distance of 8cm was used and set equal to the experimental conditions used in [24, 27].

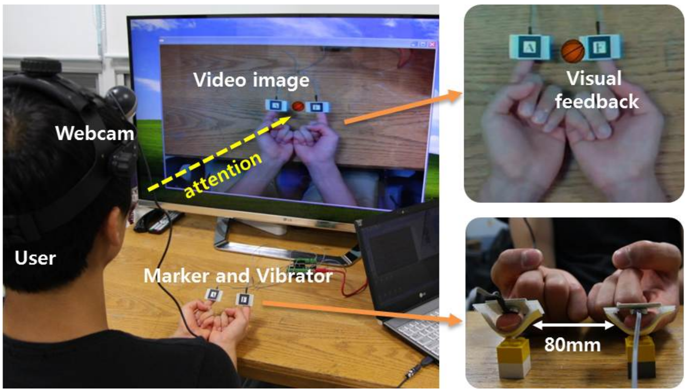


Fig. 3. The experimental set up for Experiments I and II. Vibratory stimulations were given to the two index fingers and the augmented visual feedback (e.g. bouncing basketball), detached from the fingers, seen through a large display monitor.

A common flat coin-type vibrator (11mm in diameter) was used (taped to the respective fingertips) and controlled by an Arduino board [2] (and interfaced to and synchronized with the OSGART based experiment software). For detailed specifications refer to [20]. It is controlled by a voltage input using a pulse width modulation signal with an amplitude between 0 to 5V, which in turn produces vibrations with frequency between 0 and 250 Hz and associated amplitudes between 0 to 2G (measured in acceleration, or 0 to 18μm in position) respectively. According to [20, 36], these values are well above the human's normal detection threshold (about 6 ~ 45db).

The experiment was designed as a 2x5 factor within-subject. The two factors were (1) inclusion of the visual feedback (with or without), and (2) intended locations of tactile illusion (five locations between the fingers labeled P1 ~ P5). Four survey questions were answered in a 7 Likert scale asking of the various aspects of the phantom tactile experience.

3.3 Detailed Procedure

Twenty paid subjects (15 men and 5 women) participated in the experiment with the mean age of 25.5. After collecting one's basic background information, the subject was briefed about the purpose of the experiment and instructions for the experimental task. A short training (3 minutes) was given for the subject to get familiarized to the experimental process. In addition to the head mounted fixture for the camera, the subjects wore ear muffs to prevent any bias from the sounds of the vibration. The ear muff was tested to make sure so that no sound could be heard during the experiment, and did not affect the outcome of the experiment.

The levels of stimulations were given with the intention to create phantom and real sensations at 5 equi-distanced locations between two fingers. The Linear variation of stimulus amplitudes methods of Alles [1] (Figure 4) was used with the stimulation duration set at 200ms. Preliminary studies and prior research has also confirmed that the aforementioned linear method and stimulation duration exhibited the best effect [24]. The visual feedback appeared at the intended location of sensation, at the time of the stimulations, stayed for 200 milliseconds (same as the tactile stimulation duration) and disappeared.

Each subject experienced, in a balanced order, a total of 60 positional feedbacks in all the 2 x 5 (10) conditions (6 repetitions each) with 10 second inter-stimulus rest interval, lasting about 40 minutes. For each condition, two exact same stimulation patterns were given, then subjects were asked to indicate the place of phantom sensations in terms of the five prescribed positions. We use the symbols L1~L5 to indicate the subject's response (as distinguished from the actual intended locations of sensations). The subjects were explicitly asked to report the place of tactile (e.g. rather than visual) sensation right after experiencing the stimulation. In addition, after all trials, they were asked to answer a short survey about their subject feelings (questions shown in Table 1).

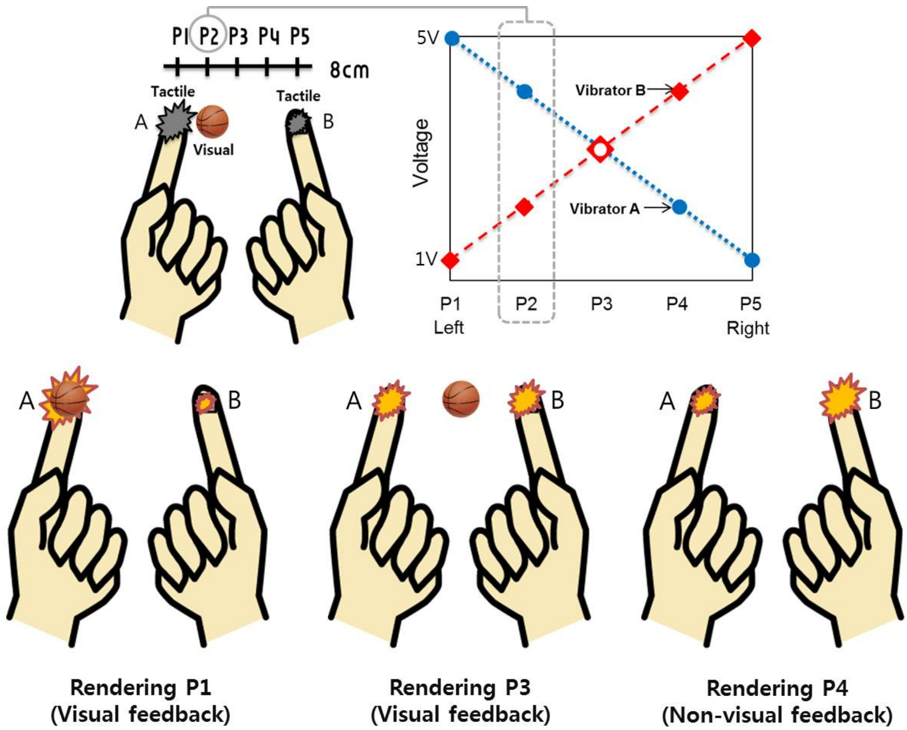


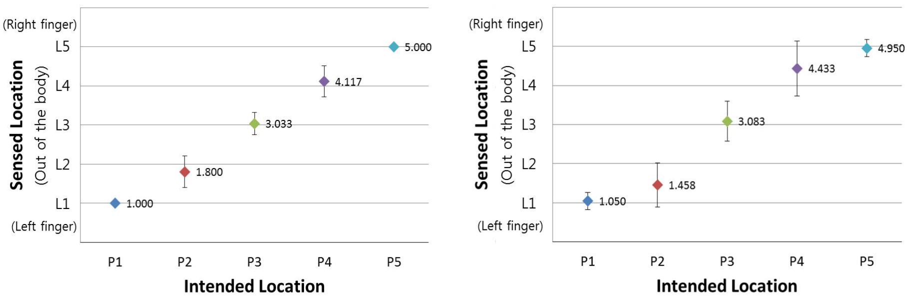
Fig. 4. The rendering method of stimulation originally proposed by Alles [1] for funneling to generate phantom sensations at five different positions (top). Examples of tactile stimulation: at P1 and P3 with visual feedback (bottom left and middle) and at P4 without visual feedback (bottom right). For example, to produce a phantom sensation at P2, a simultaneous stimulation of 4V at A and 2V at B are given as shown in the top left.

Table 1. The four survey questions regarding the subjective feel for the phantom sensation answered in 7 Likert scale

Q1	Were you able to perceive phantom sensation? (1: Not at all ~ 7: Very well)
Q2	When you perceive phantom sensation, did visual feedback affect you? (1: Not at all ~ 7: Very much)
Q3	How confidence are you about your answer to Q1? (1: Not confident at all ~ 7: Very confident)
Q4	How long did it take you to perceive the phantom sensation if any? (1: Instantly ~ 7: Few seconds)

3.4 Results

Figure 5(a) and (b) each shows the sensed/perceived locations of the tactile sensation (vertical axis), elicited by funneling, as reported by the users vs. the intended locations (horizontal axis) of sensation with visual effects and without. To our surprise, even without a virtually mediating object, phantom sensations were perceived at all five intended locations (Figure 5(b)). This is a first time discovery to our knowledge. There is still clear marked difference in the accuracy (or variance) for the intermediate locations, P2~P4. Note that P1 and P5 are where the vibrators are actually located, thus, a correct perception even without visual effect is naturally expected. Also note that the perceived locations were different among each other with statistical significances (see Table 2). Thus, a high localization controllability (~± 4mm) at approximately 2cm resolution was possible. Consistently to the statistical results, subjects reported that when no visual effects were given, it was difficult to differentiate between P1 and P2 (and similarly for P4 and P5), where the vibration motors were actually placed.



(a) With the associated visual effects (b) Without visual effects

Fig. 5. Accuracy of reported locations of the phantom sensation with respect to the intended: (a) with associated visual effect and (b) without

Table 2. Statistical differences (p-values) in the perceived locations (L1 ~ L5)

	L1-L2	L2-L3	L3-L4	L4-L5
With visual	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Without	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**

ANOVA revealed statistically significant differences existed in the senses locations between when the visual effect was given and when it was not, at all five locations except at the middle, L3 (Figure 6 and Table 3). Note that with funneling, when equal stimulation strengths are given at the two finger tips and it is plausible to think that it would be easier (that is, no help needed with the visual effect) to perceive the phantom sensation to come from the middle and make the proper response.

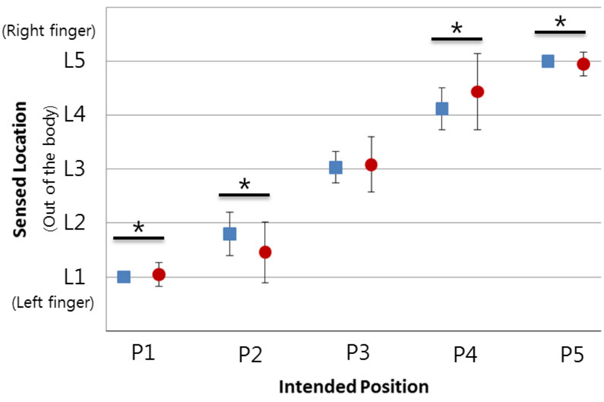


Fig. 6. A pair-wise comparison of the perceived locations between when visual effect is given (square) and not (circle). Star marks indicate those with statistically significant differences.

Table 3. Statistical differences (p-values) in the "differences" of perceived locations between when with associated visual effect (e.g. L1) and without (L1')

L1-L1'	L2-L2'	L3-L3'	L4-L4'	L5-L5'
0.013	**< 0.001**	0.352	**< 0.001**	**0.013**

Figure 7 shows the number of correct answers (i.e. correct when the perceived location matches the intended within a pre-specified threshold) in terms of score out of 100. Similarly to the above analysis, performances were generally worse without the associated visual effects for intermediate locations P2 and P4 being confused with

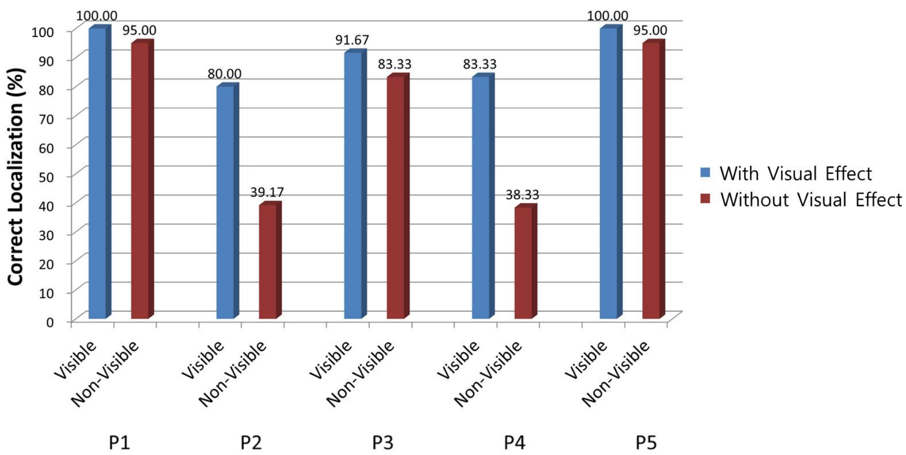


Fig. 7. Comparison of correct localization of visual and non-visual feedback based on stimulations across intended position

P1 and P5 respectively. We emphasize that while distinguishing of five distinct points within the 8cm distance was possible with visual feedback or without, the accuracy (how close the perceived is to the intended location of sensation) is expectedly lower when the visual feedback is absent.

Finally, Figure 8 shows the responses to the four survey questions, which are mostly consistent with the quantitative analysis. Subjects were conscious of the helpful effects of the visual feedback and confident of their phantom sensations.

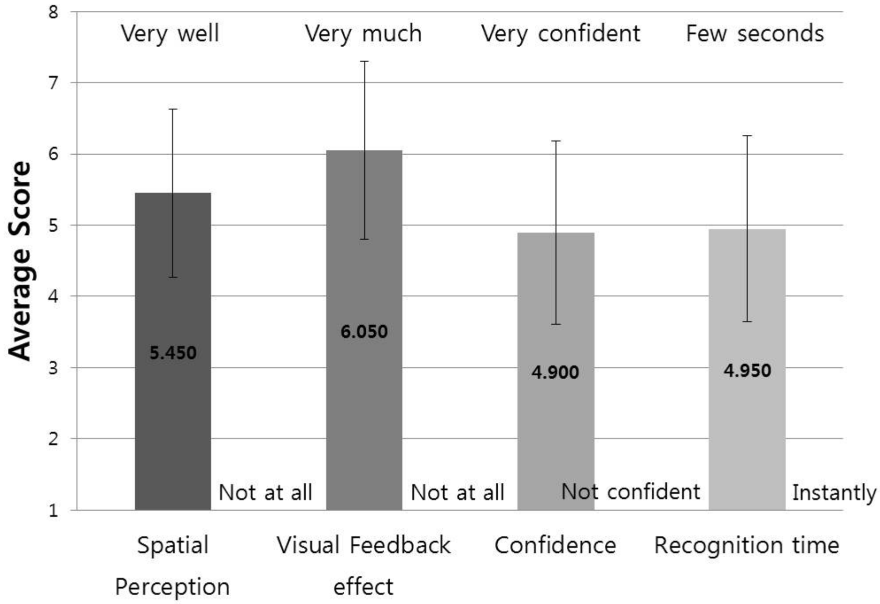


Fig. 8. The collective responses to the general usability/experience survey

4 Experiment II: Effects of Saltation with "Detached" Visual Object

4.1 Purpose and Hypothesis

The second experiment is mostly similar to the first except that the "detached" visual feedback effect to saltation was tested instead. We hypothesize for the existence and the improved quality of the phantom tactile sensation of the virtual objects when coupled with visual effects.

4.2 Experimental Design and Set Up

Again the basic experimental design and set up is mostly identical to the first one. Since saltation was used the vibro-tactile stimulations were timed rather than given simultaneously. The same "detached" augmented visual feedback, "bouncing basketball" was used. The experiment was designed as a 2x5x2 factor within-subject. The three factors were (1) inclusion of the visual feedback (with or without), (2) intended locations of tactile illusion (five locations between the fingers labeled P1 ~ P5) and (3) direction of the stimulation (from right to left or vice versa). Five survey questions were answered in a 7 Likert scale asking of the various aspects of the phantom tactile experience (see Table 4).

4.3 Detailed Procedure

Twenty paid subjects (15 men and 5 women) participated in the experiment with the mean age of 25.2 (a different pool from Experiment I). The experimental procedure was mostly identical to the first. Thus, we only describe the way saltation (i.e. timed stimulation) was administered in the treatments. Figure 9 pictorially describes how the timed vibro-tactile stimulations were given to create saltation effects. A total of three consecutive stimulations were given, each labeled S1, S2 and S3. The first two stimulation were given at P1 and the third at P5, intending to create a phantom sensation somewhere between P1 and P5. Inter-stimulus intervals of S1-S2 and S2-S3 were given with the intention to create phantom sensations at the prescribed positions, namely, P1~P5 (800ms-50ms) respectively (and stimulation duration of 80ms) based on recommended values for best effects from prior research and our own previous experiments [24].

Half of the saltation treatments were administered with right to left stimulations (at P5, then P1) and the other half, in the opposite directions (at P1, then P5). The visual effects were rendered at P1, P5 (where the actual stimulations were given) and at the intended location of sensation and stayed on for 80 milliseconds, same as the stimulation duration. As shown in Figure 10, the visual feedback was given, synchronized with the corresponding of the three tactile stimulations (S1~S3), at three locations (two at the finger tips where the actual stimulations are given and one in between at the intended location of sensation, one of P's). We stress again that the users were asked of the location and quality of the tactile feedback, not the visual.

Each subject experienced, in a balanced order, a total of 72 positional feedbacks in all the conditions: 60 times with visual feedback (2 directions, 5 intended locations of

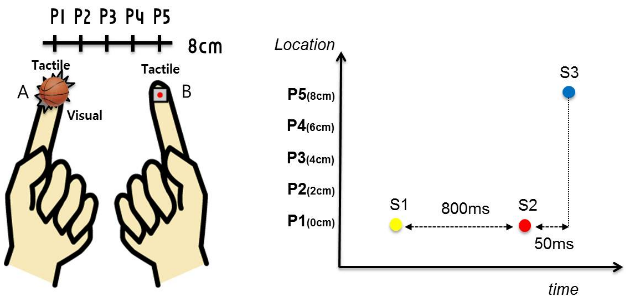


Fig. 9. The rendering method for saltation stimulation. Three timed stimulations of S1 (at P1), S2 (at P1) and S3 (at P5).

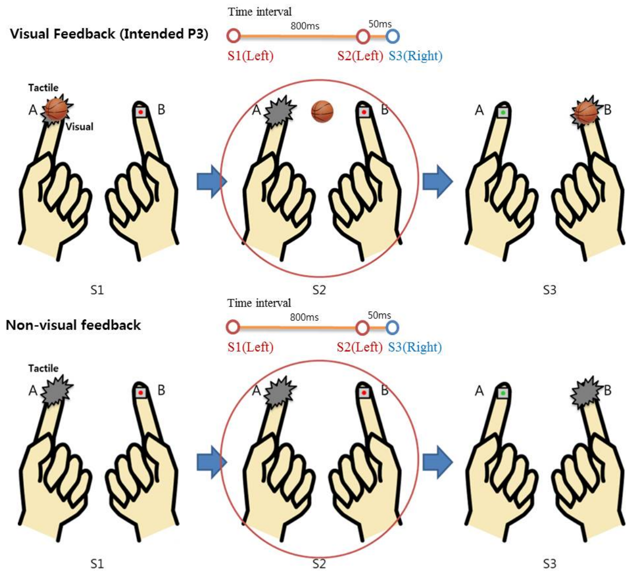


Fig. 10. Administering for the saltation effect with three timed vibro-tactile stimulations with visual effects (above) and without (below)

sensation and 6 repetitions) + 12 times without visual feedback (2 directions and 6 repetitions). Each stimulation was followed by a 10 second inter-stimulus rest interval. As it was so in the first experiment, for each treatment condition, two exact same stimulation patterns were given, then subjects were asked to indicate the place of phantom sensations in terms of the five prescribed positions (L1~L5). The subjects were explicitly asked to report the place of tactile sensation (e.g. rather than visual) right after experiencing the stimulation. In addition, after all trials, they were asked to answer a short survey about their subject feelings (questions shown in Table 4). The fifth survey question asked of the perception regarding the directionality.

Table 4. The five survey questions for Experiment II regarding the subjective feel for the phantom sensation answered in 7 Likert scale

Q1	Were you able to perceive phantom sensation? (1: Not at all ~ 7: Very well)
Q2	When you perceive phantom sensation, did visual feedback affect you? (1: Not at all ~ 7: Very much)
Q3	How confidence are you about your answer to Q1? (1: Not confident at all ~ 7: Very confident)
Q4	How long did it take you to perceive the phantom sensation if any? (1: Instantly ~ 7: Few seconds)
Q5	Were you able to recognize a particular direction? (1: Not at all ~ 7: Very well)

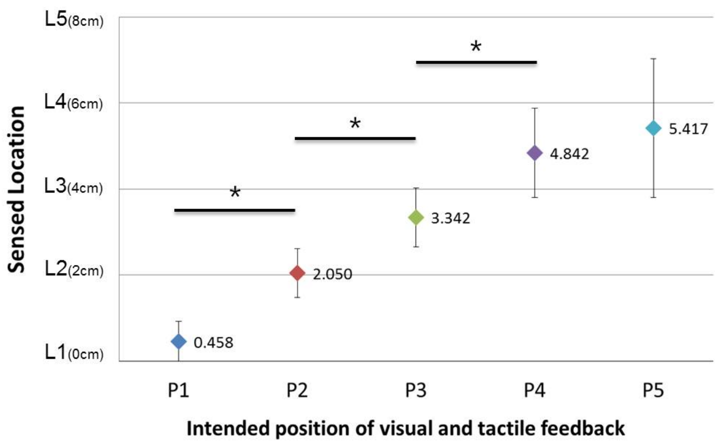
4.4 Results

Table 5 shows that the sensed tactile locations after saltation were statistically different from the actual location of stimulation for both when visual effect was given and, surprisingly again, when not given at all. Note that, according to Miyazaki, phantom sensation from saltation was not observed without visual feedback nor mediating object [27]. While our result is somewhat contrary, its extent was very small (see Figure 11(b)).

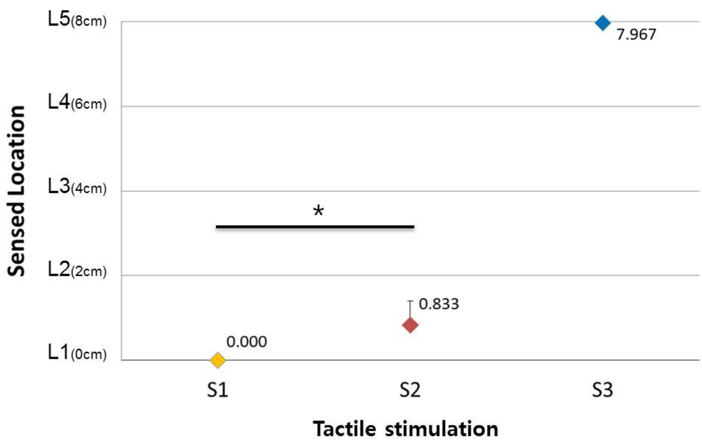
Table 5. Effects of saltation. Perceived location for S2 were in all cases different from S1 (both S1 and S2 stimluations were given at the same physical location) with statistical significance (p-values) indicating the existence of the saltation effect.

Intended position (with visual)	p-value for difference between S1 and S2 (Stim. Dir. L→R)	p-value for difference between S1 and S2 (Stim. Dir. R→L)
P1	**< 0.001**	**< 0.001**
P2	**< 0.001**	**< 0.001**
P3	**< 0.001**	**< 0.001**
P4	**< 0.001**	**< 0.001**
P5	**< 0.001**	**< 0.001**
Without visual	**< 0.001**	**< 0.001**

Figure 11(a) and (b) each shows the locations of the tactile sensation (vertical axis), elicited by saltation, as reported by the users vs. the intended locations (horizontal axis) of sensation with visual effects and without. When visual feedback was given, there were as usual five intended location of sensation according to the five respective locations of the visual feedback. However, only four statistically different sensed locations for S2 were found (i.e. L1, L2, L3, L4=L5; p-values L1-L2: <0.001; L2-L3: <0.001; L3-L4: <0.001; L4-L5: 0.088). There were significant differences in the accuracy (or variance) for the intermediate locations, between L3~L5 and P3~P5. Higher variance and lower accuracy/match (even at L1/P1) was observed as compared to the case of funneling, getting worse at the place of the third stimulation. When there was no visual feedback, we were only looking for (if it existed) one location of phantom sensation somewhere between the fingers. As mentioned above, this non-visual case did exhibit a phantom sensation at one location significantly different, though very small, from the place of actual stimulation, around 0.83cm away. Post-briefings with the subjects also reflected the observation this sensation was barely perceivable.



(a) With the associated visual effects



(b) Without visual effects

Fig. 11. Accuracy of reported locations of the phantom sensation with respect to the intended for saltation: (a) with associated visual effect and (b) without. Star marks indicate those with statistically significant differences.

Finally Figure 12 shows the responses to the five survey questions, which are consistent with the quantitative analysis. Subjects were conscious of the helpful effects of the visual feedback and confident of their phantom sensations. Subjects were also able to recognize the direction of the stimulation and the described the sensation to as soft bounce (15/20), hard contact (3/20) and moving vibration (2/20) in the post-briefing.

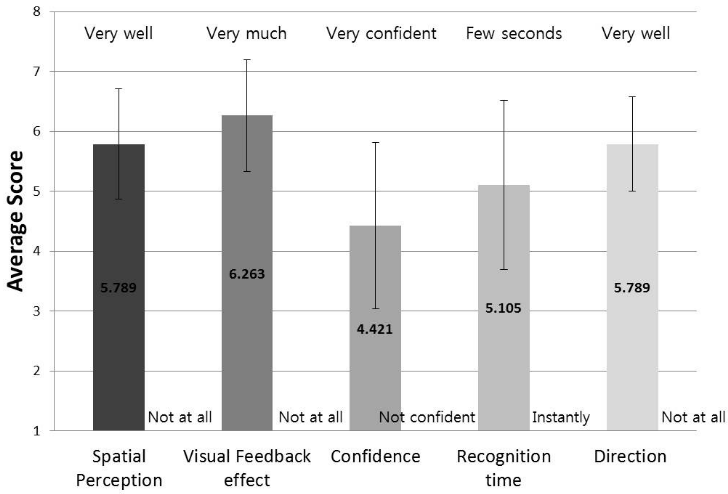


Fig. 12. The collective responses to the general usability/experience survey

5 Discussion and Conclusion

Previous research has found the phantom tactile sensations for virtual objects external to one's body. However, it required a virtual imagery to be attached to the user for directly extending one's body. This paper has investigated in whether similar phantom tactile sensations exist when the virtual object is visually detached from the user's body. Our results have shown that in addition to the perception of the phantom sensations with the "detached" visual feedback, the interaction experience was significantly enriched (vs. when without explicit visual feedback). We also discovered for the first time that for funneling, phantom sensations can be elicited without any visual feedback at all. We can further conjecture with both mediating visual feedback and the actual dynamic visual content, the tactile experience has to be improved even more with even higher localization controllability.

While the two tested phantom sensation techniques generated similar qualitatively enhanced tactile experience (both quantitatively and qualitatively), the funneling technique produced higher overall accuracy than saltation. On the other hand, saltation due to its nature seems fitting as a mean to provide directional tactile experience. The post-briefing also revealed the same. While both subjects answered both techniques did produce phantom sensations for certain, they also felt the funneling to have produced more efficient and stronger sensation.

The findings can be applied to the tactile interaction design using minimal number of actuators on a variety of media platforms including the mobile, holography and augmented reality. From a more practical perspective, we would be much interested in comparing the relative effect between (1) the usual single vibrator scheme and (2) using perhaps 2 or 4 vibrators for 1D or 2D [22] localized phantom tactile effect. The single vibrator scheme simply cannot be used for the case of "Out of the Body" virtual objects with virtual medium (e.g. two handed interaction with a dynamic holographic object). However, for the case of "Out of the Body" virtual objects with physical medium (e.g. two handed interaction with moving virtual objects on a mobile device), it is unclear whether there is sufficient benefit-to-cost in the additional effort to employ funneling or saltation. Single vibrator scheme has been quite successful in eliciting pseudo-haptic effects in smart phones and game controllers [8].

However the distinction must be made clear in terms of the role of the visual feedback. The single vibrator scheme can be explained to be a phantom or pseudo sensation directly caused by the visual feedback, while our paper has addressed the phantom sensation being strengthened reversely by the visual effect. Note that our performance measures were tactile experience, and not visual. This suggests, that although our experiment has only tested the case with virtual medium, it is plausible to expect that the tactile experience to be significantly richer than the case of single vibrator scheme, e.g. if applied to mobile devices. The plausibility is high also from the previous research indicating the weakened sensation when the "connecting" medium was virtual [24]. Either way, the combined effect can be explained by the Modality Competition Theory that the modal fusion will depend on the disparity among the stimulations in term of their consistency [33, 41]. For instance, the disparity between the location of the phantom sensation and the visual feedback seems less than that that in a single vibrator scheme.

In fact, our next focus will be to verify such a proposition. In addition, we are interested in the neurological or cognitive explanation to this phenomenon. The typical explanation for funneling and saltation has been based on the continuity of the body map in the somato-sensory area in the brain [9]. However, with the discovery of the "Out of the Body" phenomenon, it seems there may be a cognitive element to it too (reconfiguring of the body map is also possible but it is usually regarded a very slow process as demonstrated in the phantom limbs [33]).

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No 2012-0009232) and this research was funded by the Forensic Research Program of the National Forensic Service (NFS), Ministry of Security and Public Administration, Korea.

References

1. Alles, D.S.: Information transmission by phantom sensations. IEEE Transactions on Man-Machine Systems 11(1), 85–91 (1970)
2. Arduino, `http://www.arduino.cc/`
3. Barghout, A., Kammerl, J., Jongeun, C., Steinbach, E., Saddik, A.E.: Spatial resolution of vibrotactile perception on the human forearm when exploiting funneling illusion. In: IEEE International Workshop on HAVE (2009)
4. Bekesy, G.V.: Funneling in the nervous system and its role in loudness and sensation intensity on the skin. Journal of the Acoustical Society of America, 399–412 (1958)
5. Bolt, R.A.: Put-That-There, Voice and Gesture at the Graphics Interface. Computer Graphics, 262–270 (1980)

6. Borst, C.W., Baiyya, V.B.: Enhancing VR-based Visualization with a 2D Vibrotactile Array. In: Virtual Reality Software and Technology. ACM (2007)
7. Borst, C.W., Cavanaugh, C.D.: Haptic Controller Design and Palm-sized Vibrotactile Array. Technical Report, Center for Advanced Computer Studies, University of Louisiana, Lafayette (2004)
8. Bouncing Ball on Nokia N900, `http://bit.ly/x2-bouncingball/`
9. Chen, L.M., Friedman, R.M., Roe, A.W.: Optical imaging of a tactile illusion in area 3b of the primary somatosensory cortex. Science 302(5646), 881–885 (2003)
10. Cholewiak, R.W., Collins, A.A.: The generation of vibro-tactile patterns on a linear array: Influences of body site, space, and time. Perception & Psychophysics, 1220–1235 (2000)
11. Corradini, A., Wesson, R.M., Cohen, P.R.: A map-based system using speech and 3D gestures for pervasive computing. In: 4th IEEE International Conference on Multimodal Interfaces, pp. 191–196 (2002)
12. Eimer, M., Forster, B., Vibell, J.: Cutaneous saltation within and across arms: A new measure of the saltation illusion in somatosensation. Percept. Psychophys. 67(3), 458–468 (2005)
13. Flach, R., Haggard, P.: The cutaneous rabbit revisited. Journal of Experimental Psychology: Human Perception and Performance, 717–732 (2006)
14. Geldard, F.A., Sherrick, C.E.: The Cutaneous Rabbit: A perceptual Illusion. Science, 178–179 (1972)
15. Geldard, F.A.: Saltation in Somesthesis. Psychological Bulletin, 136–175 (1982)
16. Geldard, F.A.: Sensory Saltation: Metastability in the Perceptual World. Lawrence Erlbaum Associates, Hillsdale (1975)
17. Grasso, M.A., Ebert, D.S., Finin, T.W.: The integrality of speech in multimodal interfaces. ACM Transactions on Computer-Human Interaction (1998)
18. Hoggan, E., Anwar, S., Brewster, S.: Mobile multi-actuator tactile displays. In: Oakley, I., Brewster, S. (eds.) HAID 2007. LNCS, vol. 4813, pp. 22–33. Springer, Heidelberg (2007)
19. Israr, A., Poupyrev, I.: Tactile Brush: Drawing on Skin with a Tactile Grid Display. In: ACM CHI, pp. 2019–2028 (2011)
20. Jung, J., Choi, S.: Perceived magnitude and power consumption of vibration feedback in mobile devices. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 354–363. Springer, Heidelberg (2007)
21. Kilgard, M.P., Merzenich, M.: Anticipated stimuli across skin. Nature, 663 (1995)
22. Kim, Y., Lee, J., Kim, G.: Extending out of the body saltation to 2D mobile tactile interaction. In: 10th Asia Pacific Conference on Computer Human Interaction, pp. 67–74 (2012)
23. Lee, J., Spence, C.: Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. In: 22nd British HCI Group Annual Conference on HCI, pp. 185–192 (2008)
24. Lee, J., Kim, Y., Kim, G.J.: Funneling and Saltation Effects for Tactile Interaction with Virtual Objects. In: ACM CHI, pp. 3141–3148 (2012)
25. McGurk, H., MacDonald, J.W.: Hearing Lips and Seeing Voices. Nature (1976)
26. Mine, M.R., Brooks Jr., F.P., Sequin, C.H.: Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In: ACM SIGGRAPH, pp. 19–26 (1997)
27. Miyazaki, M., Hirashima, M., Nozaki, D.: The “Cutaneous Rabbit” hopping out of the body. The Journal of Neuroscience 30(5), 1856–1860 (2010)
28. Mizukami, Y., Sawada, H.: Tactile information transmission by apparent movement phenomenon using shape-memory alloy device. International Journal on Disability and Human Development, 277–284 (2006)

29. ARToolKit For OpenSceneGraph(OSGART), http://www.osgart.org/
30. Oviatt, S.L.: Multimodal interfaces. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 286–304 (2003)
31. Rahal, L., Cha, J., Kammerl, J., El Saddik, A., Steinbach, E.: Investigating the influence of temporal intensity changes on the apparent movement phenomenon. In: Virtual Environments, Human-Computer Interfaces and Measurement Systems, pp. 310–313 (2009)
32. Raisamo, J., Raisamo, R., Surakka, V.: Evaluating the effect of temporal parameters for vibrotactile saltatory patterns. In: International Conference on Multimodal interfaces, pp. 317–326 (2009)
33. Ramachandran, V.S., Blakeslee, S.: Phantoms in the Brain. Harper Collins Publishers (1998)
34. Richard, P., Burdea, G., Gomez, D., Coiffet, P.: A comparison of Haptic, Visual and Auditive Force Feedback for Deformable Virtual Objects. In: ICAT Conference (1994)
35. Richter, H., Hang, A., Blaha, B.: The PhantomStation: Towards funneling remote tactile feedback on interactive surfaces. In: Augmented Human International Conference (2011)
36. Sherrick, C.E.: A scale for rate of tactual vibration. Journal of the Acoustical Society of America (1985)
37. Sherrick, C.E., Cholewiak, R.W., Collins, A.A.: The localization of low-and high frequency vibrotactile stimuli. Journal of the Acoustical Society of America 169 (1990)
38. Tan, H.Z., Pentland, A.: Tactual displays for wearable computing. IEEE Personal and Ubiquitous Computing 1(4), 84–89 (1997)
39. Waller, D.: Factors Affecting the Perception of Inter Object Distances in Virtual Environments. Presence: Teleoperators and Virtual Environments, 657–670 (1999)
40. Warren, J.P., Santello, M., Tillery, S.I.H.: Electrotactile stimuli delivered across fingertips inducing the Cutaneous Rabbit Effect. Exp. Brain Res., 419–426 (2010)
41. Wickens, C.D., Sandry, D.L., Vidulich, M.: Compatibility and resource competition between modalities of input, central processing, and output. Human Factors 25(2), 227–248 (1983)

Precise Pointing Techniques for Handheld Augmented Reality

Thomas Vincent[1], Laurence Nigay[1], and Takeshi Kurata[2]

[1] Joseph Fourier University, UJF-Grenoble 1,
Grenoble Informatics Laboratory (LIG), UMR 5217, Grenoble, 38041, France
[2] Center for Service Research, AIST, Tsukuba Central 2, 1-1-1 Umezeno,
Tsukuba, Ibaraki, 305-8568 Japan
{thomas.vincent,laurence.nigay}@imag.fr, t.kurata@aist.go.jp

Abstract. We propose two techniques that improve accuracy of pointing at physical objects for handheld Augmented Reality (AR). In handheld AR, pointing accuracy is limited by both touch input and camera viewpoint instability due to hand jitter. The design of our techniques is based on the relationship between the touch input space and two visual reference frames for on-screen content, namely the screen and the physical object that one is pointing at. The first technique is based on Shift, a touch-based pointing technique, and video freeze, in order to combine the two reference frames for precise pointing. Contrastingly -without freezing the video-, the second technique offers a precise mode with a cursor that is stabilized on the physical object and controlled with relative touch inputs on the screen. Our experimental results show that our techniques are more accurate than the baseline techniques, namely direct touch on the video and screen-centered crosshair pointing.

Keywords: Handheld Augmented Reality, Interaction Techniques, Pointing.

1 Introduction

While still an open research area, Augmented Reality (AR) in terms of superimposition of graphics is now possible on camera-equipped handheld devices. However, issues related to interaction still need to be studied. In particular, pointing at physical objects through the live video playback of a handheld device with either direct touch or a screen-centered crosshair has limited accuracy [4, 11]. Nevertheless accurate pointing at physical objects would benefit several handheld AR applications including selection or in-situ positioning of digital annotations in dense physical environments such as paper maps.

Pointing accuracy in handheld AR is limited by various factors. First, interaction with handheld devices brings specific constraints [17]: the screen real estate is limited and direct touch on the screen, the de-facto standard input modality on such devices, is impaired by finger occlusion and an ambiguous selection point (i.e. the ”fat-finger” problem). Moreover, when considering handheld tablets which are larger but also heavier than phones, the trade-off between device handling (i.e., one or two handed hold) and

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 122–139, 2013.

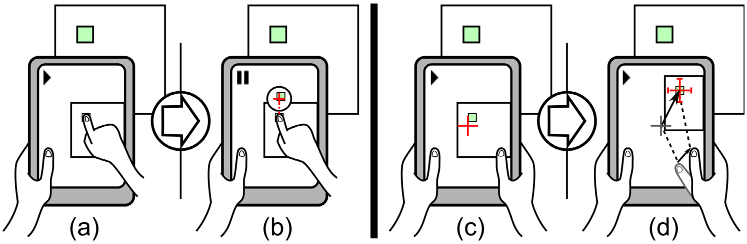


Fig. 1. Four techniques for pointing at physical objects through video on handheld devices: (a) *Direct Touch* on the live video; (b) *Shift&Freeze:* Shift [19] combined with freeze-frame; (c) Screen-centered *Crosshair;* (d) *Relative Pointing* with cursor stabilized on the physical object

touch interaction (i.e., available fingers for touch interaction and screen accessibility) need to be taken into account [20]. To address these issues, pointing accuracy on handheld devices has been studied and different techniques have been proposed [17, 19]. Yet, those techniques do not take into account the specificities of AR.

Indeed, in AR applications, when interacting through the video on handheld devices, the physical object one wants to point at is not stable on screen. As a consequence, it is also not stable within the touch input space. As the viewpoint is controlled by the device's pose, its stability is impaired by hand tremor and motion induced by the user's touch inputs. Furthermore, when using handheld tablets for AR applications, the trade-off is between viewpoint stability (i.e. a steady hold) and touch-screen accessibility. On the one hand, a steady hold with both hands (figure 1-c/d) only allows touch interaction with the thumbs in a limited region of the screen. On the other hand, when holding the tablet with one hand (figure 1-a/b), the other hand can interact with a larger area of the screen at the expense of more instability.

For handheld AR systems, pointing with a screen-centered crosshair has been studied [14, 15]. This technique is impaired by viewpoint instability. Freeze-frame techniques have been used to improve direct touch interaction by pausing the video [4, 10]. Nevertheless, one drawback of this approach is to prevent an up-to-date view of the physical scene.

To address the limitations of pointing for touch-based handheld AR systems, we propose two techniques:

- *Shift&Freeze* (figure 1-b) that addresses both direct touch accuracy limitations and viewpoint instability by combining Shift [19] with freeze-frame. Shift is a technique that extends *Direct Touch* pointing (figure 1-a) with a precise quasi-mode. We complemented this precise quasi-mode with freeze-frame to adapt Shift to handheld AR. So *Shift&Freeze* improves accuracy while still allowing direct touch for coarse but fast pointing.
- *Relative Pointing* (figure 1-d), which improves pointing accuracy without pausing the live video playback. To improve accuracy, it uses a cursor that is stabilized in the physical object's frame of reference. The cursor is controlled by indirect

relative touch inputs. As such, relative pointing in handheld AR does not share direct touch limitations and the cursor position is not impaired by viewpoint instability. Moreover this technique extends the screen-centered *Crosshair* pointing technique (figure 1-c) with a precise mode. This allows both coarse but fast pointing and accurate pointing when needed.

In this paper, we first review related work and then present the design rationale of our two techniques, *Shift&Freeze* and *Relative Pointing* in handheld AR. We then report two experiments comparing our techniques and the baseline techniques, namely *Direct Touch* on the video and screen-centered *Crosshair*. We conclude with a discussion of our results and directions for future work.

2 Related Work

We build on previous work on pointing techniques for touch-based handheld devices, as well as on pointing techniques for handheld AR and spatially aware interfaces.

2.1 Pointing Techniques for Touch-Based Handheld Devices

Much prior work has addressed how to improve pointing accuracy on touch-screen. Within the scope of our work, we examine pointing techniques on touch-based handheld devices that do not require prior knowledge of the targets, excluding for instance target expansion techniques as in *Starburst* [3]. Indeed for the case of in-situ positioning of annotations in handheld AR (e.g., annotations at any position on paper maps), there is no available knowledge of possible targets.

A first approach is zooming to enlarge the information space to a scale appropriate for accurate pointing [2]. When using zooming, the user is facing the classical trade-off between the level of zoom (for accurate pointing) and the visible context (for finding the area of interest). The interaction process can be quite tedious on handheld devices with limited screen real estate: zoom in to focus and zoom out for context. Based on zooming, *TapTap* [17] increases pointing accuracy. Two taps on screen are performed for pointing. The first tap selects a coarse area on the screen that is displayed zoomed in a pop-up view centered on the screen. The second tap performs the precise selection in the zoomed area and closes the pop-up view.

A second approach is to display a cursor to address both finger occlusion and selection point ambiguity. Potter et al. [12] proposed *Take-Off* that enables pointing adjustment and avoids finger occlusion by showing a cursor slightly above the finger position. One drawback of this technique is that the user does not know the position of the cursor until her/his finger touches the screen. Building on this technique, *Shift* [19] extends direct touch pointing with a precise quasi-mode. While in this mode, *Shift* displays a circular callout showing a copy of the screen area occluded by the finger and places it in a non-occluded location. The callout also shows a cursor representing the selection point of the finger, whose position can be adjusted by moving the finger. Validation is performed on finger lift. *MagStick* [17] also extends direct touch pointing by using a telescopic stick metaphor to enable further adjustment. When the finger touches the screen, it defines a reference point; then, by dragging the finger away from the reference point, the user can extend a telescopic stick centered on the reference point with the finger at one end and the cursor at the other end.

2.2 Pointing Techniques for Handheld AR and Spatially-Aware Interfaces

Seminal works on handheld AR like *NaviCam* [13] have paved the way for an active research area. For example, *TouchProjector* [4] allows users to move pictures on remote screens through direct touch on the live video of a handheld device. Handheld AR systems augmenting different kind of objects such as sights [1], printed conference proceedings [11], photo books [8] or paper maps [16] have been developed.

In handheld AR settings, the viewpoint in the augmented scene is usually controlled by the absolute device's pose in space (controlling the back-face camera viewpoint). As a handheld device is not self-stabilized (as opposed to the mouse for example), its pose is subject to hand jitter as for other freehand interaction techniques like laser pointers or handheld projectors [7]. As a consequence, the augmented scene the user interacts with is not stable on the screen. This is different from typical GUI situation on either desktop or handheld devices (see previous section) where the objects the user wants to interact with usually remain still on the screen during the interaction.

With such settings, pointing is usually performed with either a screen-center crosshair [8, 11, 14-16] or by direct input on the screen, using a pen or bare fingers [4, 8-11, 18]. Rohs et al. [14, 15] studied pointing with a screen-centered crosshair on a phone. They showed that the performance of this technique could be modeled with a two parts Fitt's law: physical pointing (i.e. moving the device in space) and virtual pointing (i.e. when the target is visible on screen). Hand jitter impairs accuracy of those interactions and different strategies have been proposed to improve interaction with handheld AR settings.

A first strategy is to increase target size on the screen by coming closer to the physical object or by zooming the live video. Zooming is compatible with both screen-centered cursor and direct input as well as with other strategies for improving interaction. *TouchProjector* user study shows that automatic zooming was overall the best performing technique: While zooming improves interaction, it does not render the image steady. The study also highlights that for precise manipulation a freeze-frame mode (which also performs automatic zooming) outperforms automatic zooming alone.

The freeze-frame technique belongs to the second strategy that overcomes viewpoint instability due to small hand motions. Indeed when pausing the live video, the viewpoint becomes steady. Freezing the frame also allows moving to a comfortable position for interaction. This approach is not compatible with a screen-centered cursor, but it has been proven useful to improve pen and touch interaction. In a user study, Lee et al. [10] showed that a video freeze mode improves accuracy for a drawing task with a pen through the handheld device video frame. They also noted that some users become lost when the live video is resumed as the viewpoint has changed. Another issue of freezing the frame is that the scene is no longer updated. *TouchProjector* overcomes this issue by updating the video snapshot with a digital copy of the remote screen one is interacting with. Unfortunately, a digital copy of the object of interest is not available for all AR scenarios. Freeze and zoom can be combined as previously explained in the case of *TouchProjector* and as in *TapTap* combined with video freeze for handheld AR [18]. In the latter, the combination of video freeze and zoom is a 'once' mode rather than a truly persistent mode. Nevertheless, any selection then requires two taps. Another way to stabilize the viewpoint is to use 'loose registration' as in *PACER* [11]. To interact with paper documents, they propose to display a digital copy of the document on the handheld device screen instead of the live video. This relaxes tracking requirements and allows for a coarse and filtered viewpoint to be used. Again, this requires a digital copy of the object of interest.

Finally, a third strategy consists of stabilizing inputs in the frame of reference of the physical object (or of its projection on the screen) rather than stabilizing the object of interest on the screen. *Snap-to-feature* [9] proposes to snap touch input on features of physical objects detected in the live video. This allows for better drawing of contours of physical objects on the screen without relying on freeze-frame or a digital copy of the scene. Our *Relative Pointing* technique is based on a similar strategy but does not rely on detecting features of the physical objects in the live video.

As opposed to on-screen content stabilization techniques that sever the live relation with the surrounding or use a digital copy of the scene, input stabilization offers the opportunity to improve accuracy without loosing the live relation with the surrounding. Both strategies can be complemented with zooming. Those approaches address limitations specific to the handheld AR context but not necessarily limitations of touch inputs. This is the challenge we addressed by designing *Shift&Freeze* and *Relative Pointing*, two pointing techniques that we introduce in the next sections.

3 Handheld AR Pointing

3.1 Design Rationale

To systematically analyze the issue of accurate pointing for handheld AR, we base our study on the relationships between the touch input space and two frames of reference for on-screen content: that of the screen and that of the physical object of interest.

With video freeze, the physical object's frame of reference is fixed within the image plane and thus it is stable on the screen (figure 2-b). This case is similar to GUI interfaces: Existing pointing techniques for handheld devices can be combined with video freeze. *Shift&Freeze* combines the existing pointing technique Shift [19] with video freeze.

When interacting through live video, the physical object is jittery on the screen (figure 2-a). In this case, we consider (see figure 3): (1) whether pointing is performed with or without an instrument (i.e. a cursor), and (2) in which frame of reference pointing is performed - either the frame of the screen or the frame of the physical object.

Direct Touch is the case of pointing in the screen frame without a cursor. Screen-centered *crosshair* makes use of a cursor and points in the screen frame. Those two techniques (with and without a cursor) are impaired by hand jitter as pointing occurs in the screen frame where the physical object of interest is not stable (figure 2-a).

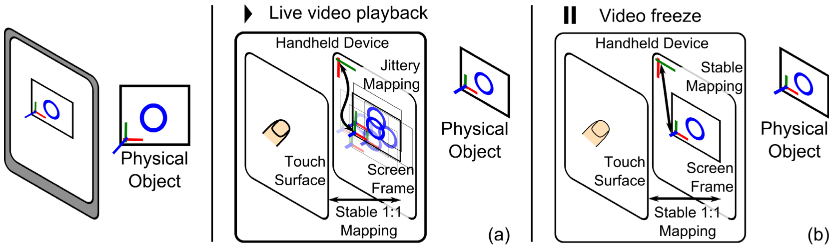


Fig. 2. Relationships between the frames of reference of touch input, of the screen and of the physical object on the screen: (a) with live video playback; (b) while the video is frozen

If pointing was performed in the frame of reference of the physical object rather than in the screen frame, pointing accuracy would not be impaired by hand jitter. Pointing in the physical object's frame of reference without a cursor instrument implies interaction directly on the physical object. We did not consider this case and focused on interaction with the touch-screen of the handheld device. Moreover such physical interaction seems cumbersome while holding a handheld tablet. Our *Relative Pointing* technique is the case where pointing is performed with a cursor in the physical object's frame of reference. In this case, we use an indirect relative mapping of touch inputs to cursor motions in the frame of reference of the physical object. As such, the cursor position is not impaired by hand jitter.

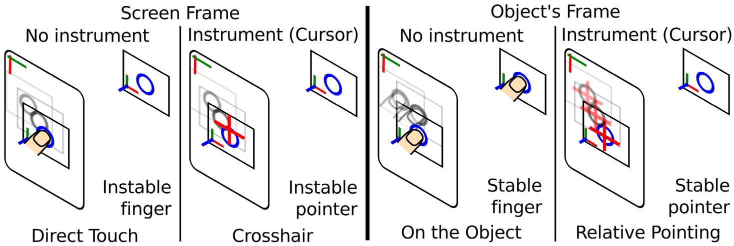


Fig. 3. Pointing through live video: four cases

This analytical framework based on the spatial relationships between the input space and two visual output frames of reference guided the design of our two techniques *Shift&Freeze* and *Relative Pointing*. Their respective designs result from a twofold strategy: *Shift&Freeze* is conceived as an improvement of *Direct Touch* and solves its accuracy problem by freezing the video and *Relative Pointing* is an improvement of *Crosshair* by adding a relative cursor stabilized on the remote physical object.

3.2 Shift and Freeze

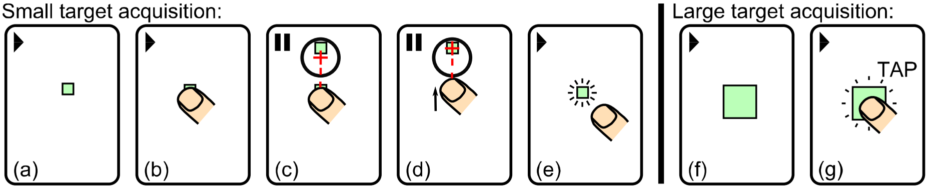


Fig. 4. Shift&Freeze walkthrough. (a-e) Small target acquisition with Shift and frozen video*;* (f-g) Large target acquisition with the *Direct Touch* technique (one tap on the screen).

Figure 4 shows a walkthrough of the two modes of our *Shift&Freeze* technique.

Scenario 1: (a) The user points the handheld device camera at the target so that the target appears on the screen. (b-c) After a short dwell time after finger contact, *Shift&Freeze* enters a precise quasi-mode and the video is frozen. A callout is placed above the finger and shows the area under the finger and a cursor at the current selection point position. (d) While in this mode, the video remains frozen and the user can adjust the position by moving its finger. (e) On finger lift, the target is selected and the live video playback is resumed.

Scenario 2: (f-g) For large enough targets where hand tremor and finger occlusion are not a problem, selection can be performed with a tap on the screen at the position of the target.

To cope with touch input limited accuracy, we chose to use *Shift* [19] for the following reasons. First, *Shift* does not require knowledge of existing targets to improve accuracy. Also, *Shift* extends *Direct Touch*, thus fast but imprecise pointing is still possible. Finally, similar techniques have been used to precisely place the cursor in text entry in commercial products and to improve accuracy when using 'loose registration' [11].

To cope with viewpoint instability in handheld AR settings, we combined *Shift* with freeze-frame. Touch-based pointing techniques in general and *Shift* in particular are designed for pointing at static targets on the screen. Instead of implementing freeze-frame as a mode, we complemented *Shift's* precise quasi-mode with video freeze. Compared to the original *Shift* technique, no extra user action is necessary to control video freeze/unfreeze. Nevertheless, as noted in [10], resuming the live video leads to a discontinuity of viewpoint that might disorient the user.

As such, the *Shift&Freeze* technique has the following properties: (1) By extending *Direct Touch,* this technique requires interaction overhead only when accuracy is required; (2) It allows precise pointing using *Shift's* callout on a frozen frame.

3.3 Relative Pointing

Figure 5 shows a walkthrough of the two modes of our *Relative Pointing* technique.

Scenario 1: (a) The user points the handheld device camera at the target so that the target appears on the screen. (b) In order to mitigate the instability due to hand tremor,

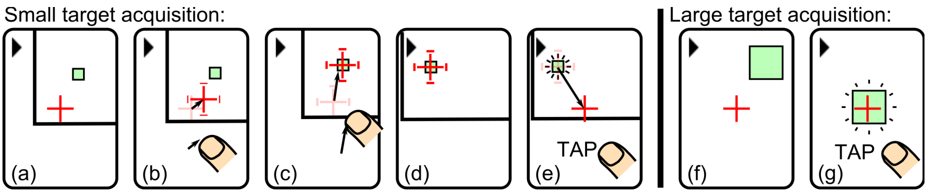


Fig. 5. Relative *Pointing* walkthrough. (a-e) Small target acquisition with a relative cursor stabilized on the physical object; (f-g) Large target acquisition with the *crosshair* technique.

when the user touches the screen and starts moving its finger on the screen, a relative pointing mode is triggered. (c) While in this mode, the cursor is no longer bound to the screen center. Instead, it is stabilized on the remote physical object at its current position. The user fine-tunes the cursor position by controlling the cursor indirectly with finger gestures on the screen. (d) On finger lift, no special action is performed. (e) The validation of a position is performed with a tap on the screen. Upon validation, a short animation moves the cursor back to the screen center, thus leaving the relative pointing mode.

Scenario 2: (f-g) When acquiring large enough targets, hand tremor is not a problem. In this case, the user does not need to use the relative pointing mode and can trigger a target acquisition at the position of the screen-centered cursor with a tap on the screen. This is similar to the screen-centered *Crosshair* technique.

To make relative pointing effective for handheld AR context, the following issues have been addressed.

Combining Absolute Physical Pointing and Touch-Based Relative Pointing. As the device's pose controls the camera viewpoint, the target in the physical world needs to be placed in the camera field of view before interaction with it can start. So, cursor-based relative pointing.needs to be combined with this absolute direct pointing in space. That is why we chose to extend the screen-centered *crosshair* pointing technique as it already uses a cursor and only relies on the device's pose for both viewpoint control and pointing. We extended this technique with a relative pointing 'once' mode where the cursor is no longer fixed at the center of the screen. Instead, the finger indirectly controls the cursor's position. This mode is triggered by finger motion on the screen. Lifting the finger from the screen does not deactivate this mode. This allows both finger clutching and checking the current cursor position before validation. A tap on the screen triggers the pointing validation. It is possible to cancel this relative pointing mode by pressing a button. Also, when tracking is lost, the relative pointing mode is cancelled. Finally, the cursor is bound to the screen. In case a change in the camera's viewpoint or a finger motion would otherwise make the cursor invisible on the screen (i.e., outside the screen), the cursor is automatically moved so that it remains visible on the screen.

Transfer Function. When dealing with indirect relative input, the transfer function (figure 6) that maps input motions to cursor displacements in the visual output space is of particular interest.

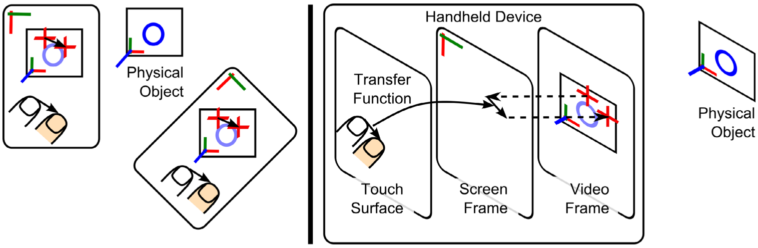


Fig. 6. With *Relative Pointing*: (left) Effect of screen rotation on the cursor's position; (right) The cursor is stabilized in the physical object's frame of reference and relative touch motions are applied on the screen

First, a transfer function that maps touch motions in the screen frame directly to cursor displacements in the physical object's frame is not appropriate. In this case, the relative rotation between the device and the physical object's frame, the distance between the device and the physical object, and the zoom factor would affect the displacements of the cursor on the screen. Yet the user is looking at the physical object through the live video on the screen. Therefore the control loop is between the finger motions and the cursor displacements on the screen (and not in the physical object's frame).

Instead, for *Relative Pointing*, the transfer function is applied in the screen frame. When a finger motion input is received, the cursor position is projected from the physical object's frame onto the screen; the cursor displacement is applied on the screen; and the new cursor position is projected back onto the physical object (figure 6-right). This guarantees that the behavior of the cursor on the screen is consistent when the device is rotated (figure 6-left) and when the viewpoint or the zoom factor changes. In short, we use the physical object's frame of reference to stabilize the cursor and the screen frame of reference to apply cursor displacements.

A second question is which transfer function should be used. Transfer function is the place for interaction improvements such as adjusting the control-to-display ratio dynamically according to the input device speed. Dynamic transfer function is a default feature for the mouse and touchpad inputs in common desktop environments. While dynamic transfer function has been studied for desktop environment, we are not informed of thorough evaluation of indirect mapping on handheld device's touch-screens [6]. A dynamic transfer function can be used with the *Relative Pointing* technique. For our developed technique and experiments, we used the transfer function of touchpad inputs on Mac OS X: `osx:touchpad?setting=0.875` [6]. With this configuration, the transfer function allows both reaching of most of the screen with fast movement and accurate positioning at lower speed.

To sum up, the *Relative Pointing* technique has the following properties: (1) By extending *Crosshair,* this technique requires interaction overhead only when accuracy is required. (2) By stabilizing the cursor on the remote physical object, it offers accuracy assistance without relying on video freeze.

4 Experiments

We ran two experiments to evaluate four handheld AR pointing techniques (the two baseline techniques and the two proposed techniques):

- *Direct Touch:* Pointing with selection at the finger press position;
- *Crosshair:* Screen-centered crosshair pointing where validation is triggered on finger press with a tap anywhere on the screen;
- *Shift&Freeze* with a 300 ms delay for escalation, a 44mm wide callout initially placed 22 mm above the initial touch position, and no zoom in the callout; and
- *Relative Pointing* as described above but without a cancel button.

All cursor-based techniques (i.e. *Crosshair, Shift&Freeze*'s callout and *Relative Pointing)* are using the same red square cross cursor, which is 7.7mm in size with a stroke width of 0.2mm.

The goal of the first experiment was to collect both user feedbacks and quantitative data on those techniques while performing a rather realistic task: placing marks on a wall map. Building on this experiment, we ran a second experiment to further evaluate those four techniques in a controlled experiment while acquiring small targets.

4.1 Experiment 1: User experience

For this experiment, we formulated the following hypothesis:

- H1: *Relative Pointing* and *Shift&Freeze* are preferred over *Crosshair* and *Direct Touch.* This is motivated by the extra precise mode offered by our two techniques. Moreover, this precise mode does not prevent the use of *Crosshair* or *Direct Touch* as a basic mode.
- H2: On tablet form factor, indirect cursor-based techniques are preferred over direct pointing techniques. So *Relative Pointing* is preferred over *Shift&Freeze* and *Crosshair* is preferred over *Direct Touch*. This is based on both the finger occlusion issue for direct touch input and the trade-off between tablet hold and screen accessibility.

Procedure, Apparatus and Participants. For each of the four techniques, we first explained the technique and participants had a chance to freely try it. Then, participants performed 5 different pointing tasks to place AR marks on a physical wall map with a handheld tablet. Each task was repeated three times. For each trial, participants started at 2.5m from the wall map. They were instructed to move freely in the room and to hold the tablet in portrait mode. Finally, a debriefing questionnaire and interview concluded each technique's experiment.

Before starting the experiment with the four techniques, participants started to perform each of the 5 tasks once with no interaction by only finding the required targets through the video on the tablet screen. This was introduced to help participants to become acquainted with the tasks and the experimental setting (in particular form factor and video quality). After this initial training, all participants started with the *Direct Touch* technique. The presentation ordering of the other three techniques was then counter-balanced across participants using a Latin-square. We used this design so that all participants share *Direct Touch*, the de facto standard interaction, as a common baseline. Experiments lasted about one hour including a debriefing discussion.

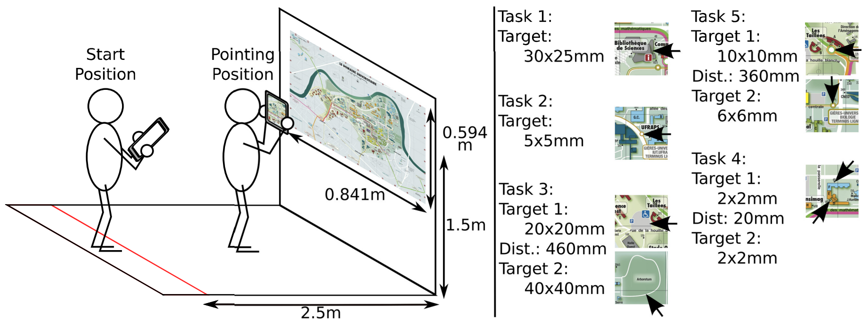


Fig. 7. Experimental set-up: (left) Participants started 2.5m away from the physical wall map and then could move freely to perform the pointing tasks; (right) the targets used for each task. Tasks 3, 4 and 5 consist of placing AR marks on 2 targets.

We use a map of our campus site in A1 format in landscape orientation (841mm x 594mm). It was placed vertically on a wall with the middle of the map 1.5m above the floor (figure 7-left). The targets of the 5 tasks are shown on the right of figure 7. Tasks 1 and 2 consist of placing a mark on a single target. Tasks 3, 4 and 5 consist of placing marks on two different targets. The target sizes range from 2mm to 4cm.

The experiment was conducted on an iPad2 (weight: 601g, screen resolution 1024x768 pixels (132 dpi)). The system provides touch input with the same resolution as the screen. We developed an ad hoc application for the experiment using OpenGL|ES 1.1[1], Vuforia SDK 1.5.9[2] and libpointing[3] [6]. This application runs at about 30 frames per second. The size of the camera images is 480x640 pixels and images are displayed full-screen. Statistical analysis was performed with R[4].

Twelve unpaid volunteers (4 female, 8 male; 1 left-handed and 1 ambidextrous), ranging in age from 22 to 45 years (mean 27 years), were recruited from our institution.

1 http://www.khronos.org/opengles/1_X/
2 https://www.vuforia.com/
3 http://www.libpointing.org/
4 http://www.R- project.org

All participants had previous experience with touch-based handheld devices (seven on a daily basis) amongst whom nine had used a handheld tablet before.

Results

User preference. The questionnaire was composed of 7 questions taken from the System Usability Scale questionnaire [5] (except questions 4, 5 and 6 that are applied to more complex systems). Responses were on four point Likert scale and gathered as a global usability score ranging from 0 (poor) to 21 (high). The overall median score is 17/21, *Crosshair* has the lowest median score (14/21), followed by *Relative Pointing* (16.5/21), then *Shift&Freeze* (17.5/21) and finally *Direct Touch* (18/21) (figure 8-left). Score differences are not statistically significant (Kruskal-Wallis rank sum test of score by technique: X^2=6.651, p>0.05).

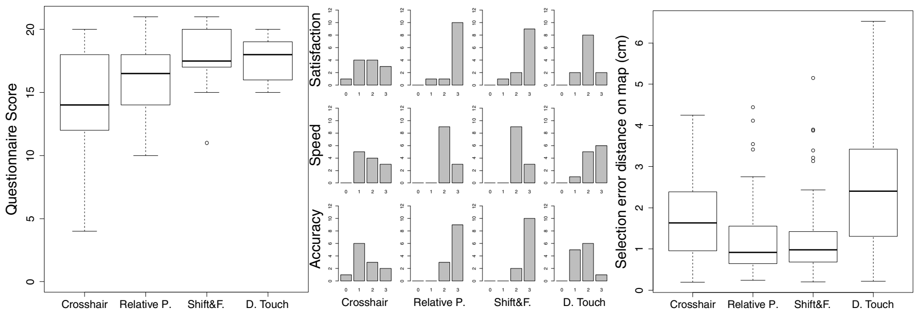


Fig. 8. For each technique: (left) Questionnaire scores; (center) Histograms of overall satisfaction, speed and accuracy rankings; (right) Boxplot of distances between the selection point and the target's center for Task 4 (2mm targets)

Participants also ranked each technique in terms of overall satisfaction, speed and accuracy on four point Likert scale (figure 8-center). Kruskal-Wallis rank sum tests found significant differences between techniques for overall satisfaction (X^2=14.3897, p<.01) and accuracy (X^2=24.2827, p<.0001) rankings by technique. A post-hoc pairwise comparison of overall satisfaction and accuracy rankings shows significant difference (with p<0.05 for overall satisfaction and p<0.01 for accuracy) for all pairwise comparisons except for the comparisons between *Shift&Freeze* and *Relative Pointing* and between *Crosshair* and *Direct Touch.* Table 1 gives satisfaction and accuracy ranking means.

Table 1. Means of satisfaction and accuracy ranking by techniques

Rank Means	Crosshair	Relative Pointing	Shift&Freeze	Direct Touch
Satisfaction	1.75	2.75	2.67	2
Accuracy	1.5	2.75	2.83	1.67

Finally, during the experiment debriefing, we asked participants for the techniques they found to be the fastest and the more precise and for the techniques they preferred overall (multiple answers were allowed). Six participants said *Relative Pointing* seems to be the fastest one, four said *Direct Touch,* three said *Crosshair* and one said *Shift&Freeze*. All but one participant said *Shift&Freeze* seems to be the most precise one and five said it was *Relative Pointing*. Eight participants preferred *Relative Pointing* and six preferred *Shift&Freeze*. Two more participants would have also preferred *Shift&Freeze* given that it provided zoom and a cancel option.

Selection Accuracy. We looked at the spread of selection points around the small targets of Task 4 (2mm wide). From 288 target selections, we removed 7 outliers noted during the experiment. The overall median of distances to the targets on the map is 1.7cm. The *Direct Touch* median (2.4cm) is more than twice that of *Relative Pointing* (0.9cm) and *Shift&Freeze* (1.0cm). The *Crosshair* median (1.6cm) lies in between (figure 8-right).

Discussion. These results support hypothesis H1. Indeed, *Shift&Freeze* and *Relative Pointing* are preferred over *Direct Touch* and *Crosshair*. Participants gave the best ranks in terms of accuracy and overall satisfaction to *Shift&Freeze* and *Relative Pointing*. Moreover participants have never mentioned either *Direct Touch* or *Crosshair* when asked for their preferred technique or for the most precise technique. However, these results do not support hypothesis H2. *Crosshair* received the lowest SUS scores, and *Shift&Freeze* was almost unanimously said to be the most precise technique. Moreover *Shift&Freeze* and *Relative Pointing* were almost equally preferred. So, indirect pointing techniques were not preferred over direct touch-based ones (i.e. *Direct Touch* and *Shift&Freeze)* even if the tablet form factor was presumably less convenient for direct touch-based techniques. Actually, some of our participants were used to direct touch input up to the point that they were tempted to tap on the cursor for the two indirect pointing techniques (*Crosshair* and *Relative Pointing).* The trend given by the measurable results is consistent with feedback gathered during the interviews.

Some participants complained about the handheld tablet form factor. A first reason is that due to its size and weight it is best held with both hands, but, as already explained, this impairs access to the screen with the *Direct Touch* and *Shift&Freeze* techniques. A second reason is that holding the tablet for AR application is different from other applications. The user needs to maintain the camera focus while interacting with the screen. Some participants felt that they held the tablet unsafely as they found it to be slippery and proposed to add some grips to the device. Also, the screen borders are not broad enough to allow all users to hold the tablet with their thumb on the side of the screen. This results in accidental touch inputs and an uncomfortable tablet hold when trying to hold the tablet with one hand in order to interact with the other one.

As for the distance from the wall map, most of the participants walked about the same distance for all tasks and all techniques. This is a surprising result since we expected the participants to adapt the distance to the map according to the difficulty of the task. Only one participant clearly adapted his distance from the map according to both target size and ease of manipulation of the technique. He did so up to the point that he did not walk at all for large targets with *Relative Pointing* (as he felt more comfortable with this technique).

The spread of selection points around small targets suggests that *Relative Pointing* and *Shift&Freeze* have higher accuracy than the two baseline techniques. It also suggests that *Direct Touch* is the least precise technique and that *Crosshair* has an intermediate accuracy. In the next controlled experiment, we further study small target acquisition.

4.2 Experiment 2: Performance

For the second experiment, we made the following hypothesis:

- H1: *Relative Pointing* and *Shift&Freeze* are more accurate than *Direct Touch* and *Crosshair* but they take longer to operate for small targets.
- H2: *Crosshair* is more accurate than *Direct Touch.* While both techniques are impaired by hand jitter, *Crosshair* does not suffer from finger occlusion.
- H3: *Relative Pointing* and *Shift&Freeze* offers similar accuracy. Both techniques overcome limitations inherent to touch input and hand jitter.

Procedure, Apparatus and Participants. This experiment was carried out utilizing the cyclical multi-direction pointing task paradigm. We used thirteen targets arranged in a circle on a remote screen. As the handheld tablet application uses computer vision to track the device's pose, the targets were overlaid on a background image. One target at a time was highlighted in black on the remote screen: this target must be selected by pointing at it on the tablet through the live video. In order to ensure a good visibility of the highlighted item regardless of its width, it was surrounded by a 3 cm wide white square with a green cross. Targets always appear in the same order: starting from the top item, the next item is always opposite and slightly clockwise from the selected one. One block thus consists of thirteen target selections plus the selection of the first target. The subjects were instructed to hold the tablet in portrait mode, to select the highlighted target as quickly and accurately as possible and to rest between blocks.

We used a single movement amplitude of 30 cm and 3 target widths (0.5 cm, 1 cm and 2 cm). We wished to have a consistent distance between the remote screen and the handheld tablet across participants and blocks. To do so, before each block, participants had to place the handheld tablet 1 meter (+/- 5cm) away from the remote screen by following indications displayed on the tablet screen. Those indications were hidden as soon as subjects started the block to avoid disturbing them during the experiment.

Presentation ordering of the four techniques and the three widths were counterbalanced using Latin squares. Each condition was presented three times including one time for training. The experimental design is:

4 Techniques x 3 Widths x 2 Blocks x 13 Selections = 312 acquisitions per subject, and *4 Techniques x 3 Widths x 13 Training Selections* =156 training acquisitions per subject.

For the handheld tablet, we used a similar apparatus as for the previous experiment. In addition, the targets were displayed on a 27" Apple Thunderbold display with 2560x1440 pixel resolution (109 dpi). The screen was placed vertically so that its center was 1.5m high from the ground. An ad-hoc application was developed to control target widths and to highlight the target on the remote screen.

Twelve unpaid volunteers (4 female, 8 male; 1 left-handed), ranging in age from 22 to 41 years (mean 30 years), were recruited from our institution. All participants had previous experience with touch-based handheld device (nine on a daily basis) amongst whom ten had used a touch-based tablet before.

Results. From 3744 selections, we removed 33 obvious outliers. Distance from the screen when selections were performed is on average 1.02m (1^{st} quartile: 0.99m, 3^{rd} quartile: 1.05m, range: 0.90m to 1.18m). This indicates that our experimental set-up that constrains participants to placing the handheld tablet at 1m (+/- 5cm) from the screen before starting each block was sufficient to confine the distance between the handheld tablet and the remote screen to a small range.

Errors. A Pearson's Chi-squared independence test between success of target acquisition and the 4 *Techniques* shows a significant dependence ($X^2 = 616.0356$, $p < .0001$). The overall error rate is 44.6%. This high error rate is explained by the choice of rather small target widths. The lowest error rate over the 3 target *Widths* is for *Relative Pointing* (20.1%), then *Shift&Freeze* (34.6%), *Crosshair* (49.4%) and the highest error rate is observed for *Direct Touch* (75.0%) (Figure 9-left).

We performed a 4 x 3 (*Technique* x *Width*) within subject analysis of variance on error rate by user. The *Technique* ($F_{3,143}$=50.835; p<.0001) and *Width* ($F_{2,143}$=57.286; p<.0001) main effects as well as the *Technique:Width* interaction ($F_{6,143}$=3.397; p<.01) were found significant. A post-hoc Tukey multiple means comparison found significant difference for all comparisons. Differences between *Relative Pointing* and *Shift&Freeze* and between *Shift&Freeze* and *Crosshair* were found significant with p<.012. All other differences were found significant with p<.0001.

We also performed a 4 x 3 (*Technique* x *Width)* within subject analysis of variance on median by user of distance between target center and selection point. Significant main effects were found for *Technique* ($F_{3,143}$= 42.605; p < .0001) and *Width* with p<.015 ($F_{2,143}$= 4.389). *Technique:Width* interaction was not significant. A post-hoc Tukey multiple means comparison found significant difference for all comparisons (with p<.01 for *Relative Pointing-Crosshair* comparison and p<.0001 for the others) except between *Relative Pointing* and *Shift&Freeze* and between *Shift&Freeze* and *Crosshair.*

Duration. The overall median of selection durations is 2.1 seconds, and medians of all selections for each technique are 2.7 seconds for *Relative Pointing,* 2.5 seconds for *Shift&Freeze,* 2.2 seconds for *Crosshair* and 1.0 second for *Direct Touch* (figure 8-right). We performed a 4 x 3 (*Technique* x *Width)* within subject analysis of variance on median of selection durations by user. Significant main effects were found for

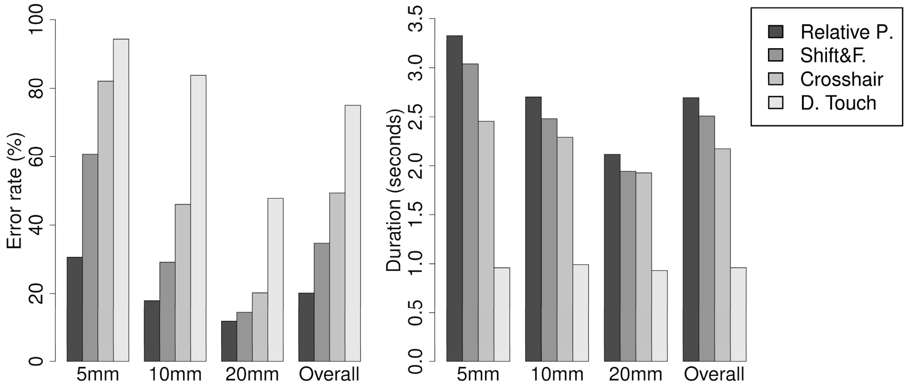


Fig. 9. (left) Error rates (%) and (right) selection durations (sec.) by *Technique* by *Width*

Technique ($F_{3,143}$= 67.781, p<.0001) and *Width* ($F_{2,143}$= 17.478, p<.0001). Effect of the *Technique:Width* interaction was also significant though with p=.0127 ($F_{6,143}$= 2.827). A post-hoc Tukey multiple means comparison found significant difference for all (with at least p<.01) except for two comparisons. Again, the differences between *Relative Pointing* and *Shift&Freeze* and between *Shift&Freeze* and *Crosshair* were not significant.

Discussion. The chosen tasks were quite hard to perform, which results in high error rates for all the techniques. Still we can observe that the different techniques offer different trade-offs between speed and accuracy or better said here, between duration and error rate (figure 9). The results partly support hypothesis H1. *Relative Pointing* is significantly more accurate and longer to operate than the two baseline techniques (i.e. *Direct Touch* and *Crosshair),* but this is not the case for *Shift&Freeze.* Indeed, *Shift&Freeze* does not show significant difference with *Crosshair. Crosshair* is significantly more accurate than *Direct Touch,* which supports hypothesis H2. Actually, *Direct Touch* is not adapted for those small target widths as indicated by both high error rates and no difference of duration across target widths. *Relative Pointing* and *Shift&Freeze* offers similar performance as indicated by non-significant difference of both duration and error distance. This supports hypothesis H3.

While participants held the tablet with both hands with *Relative Pointing* and *Crosshair,* they adopted different strategies with *Direct Touch* and *Shift&Freeze.* For *Direct Touch* and *Shift&Freeze,* participants used two different strategies: (1) holding the tablet with one hand and interacting with the other hand's finger (9/12 for *Direct Touch* and 6/12 for *Shift&Freeze)* and (2) holding the tablet with both hands and interacting with their thumb (3/12 for *Touch* and 6/12 for *Shift&Freeze).* This highlights the trade-off between holding the tablet and interacting on the screen for direct touch based techniques.

One drawback of this experiment is that the choice of the selection mode for *Shift&Freeze* and *Relative Pointing* was left to the participants. This results in different strategies as some users always used the precise mode while others adapted the mode according to the difficulty of the task. Nevertheless our goal was to evaluate our two techniques that include two modes.

5 Conclusion and Future Work

This paper provides a comprehensive study on precise pointing techniques for handheld Augmented Reality (AR). Our contributions are twofold. First we have presented an analytical framework for the design of interaction techniques for handheld AR that is based on the relationship between the touch input space and two visual output frames, namely the screen and the physical object. The usefulness of the framework is demonstrated by the classification of existing techniques and the design of two pointing techniques. Second we have presented two pointing techniques, *Shift&Freeze* and *Relative Pointing*. Their respective designs result from a twofold strategy: *Shift&Freeze* is conceived as an improvement of *Direct Touch* and solves its accuracy problem by freezing the video and using Shift's callout. *Relative Pointing* improves on the screen-centered *Crosshair* technique by stabilizing the cursor on the remote physical object. The two experiments revealed that those two techniques are preferred to the two commonly used techniques (*Direct Touch* and screen-centered *Crosshair*) and are more accurate than these baseline techniques. Further controlled studies must be performed to compare the two techniques, firstly, with less difficult tasks and then with a phone as the tablet form factor probably favors *Relative Pointing*. We also plan to run experiments with 3D physical objects (e.g., production machines). Several extensions to the two proposed techniques are envisioned including zooming for *Shift&Freeze* and testing different dynamic transfer functions (with or without known targets) for *Relating Pointing*. Finally, since our studies formulate the hypothesis of a perfect tracking of the device's pose, one further research avenue we must explore is the design of handheld AR pointing techniques that takes into account the imperfection of the underlying tracking system. We expect that precise pointing techniques in any context of use will become more and more crucial in the future, as a large range of richer and more complex handheld AR applications are designed.

Acknowledgements. This work has been supported by the French-Japanese ANR/JST AMIE project (http://amie.imag.fr). Special thanks to G. Serghiou for reviewing the article and to N. Mandran (LIG) for helping with the experiments.

References

1. Alessandro, M., Dünser, A., Schmalstieg, D.: Zooming interfaces for augmented reality browsers. In: Proc. MobileHCI 2010, pp. 161–170. ACM (2010)
2. Albinsson, P.-A., Zhai, S.: High precision touch screen interaction. In: Proc. CHI 2003, pp. 105–112. ACM (2003)
3. Baudisch, P., Zotov, A., Cutrell, E., Hinckley, K.: Starburst: a target expansion algorithm for non-uniform target distributions. In: Proc. AVI 2008, pp. 129–137. ACM (2008)

4. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: Mobile interaction through video. In: Proc. CHI 2010, pp. 2287–2296. ACM (2010)
5. Brooke, J.: SUS-a quick and dirty usability scale. In: Usability Evaluation in Industry, ch. 21. Taylor and Francis, London (1996)
6. Casiez, G., Roussel, N.: No more bricolage!: methods and tools to characterize, replicate and compare pointing transfer functions. In: Proc. UIST 2011, pp. 603–614. ACM (2011)
7. Forlines, C., Balakrishnan, R., Beardsley, P., van Baar, J., Raskar, R.: Zoom-and-pick: facilitating visual zooming and precision pointing with interactive handheld projectors. In: Proc. UIST 2005, pp. 73–82. ACM (2005)
8. Henze, N., Boll, S.: Who's that girl? Handheld augmented reality for printed photo books. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part III. LNCS, vol. 6948, pp. 134–151. Springer, Heidelberg (2011)
9. Lee, G.A., Yang, U., Kim, Y., Jo, D., Kim, K.-H.: Snap-to-feature interface for annotation in mobile augmented reality. In: Workshop on Augmented Reality Super Models at ISMAR 2010 (2010)
10. Lee, G.A., Yang, U., Kim, Y., Jo, D., Kim, K.-H., Kim, J.H., Choi, J.S.: Freeze-set-go interaction method for handheld mobile augmented reality environments. In: Proc. VRST 2009, pp. 143–146. ACM (2009)
11. Liao, C., Liu, Q., Liew, B., Wilcox, L.: PACER: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone. In: Proc. CHI 2010, pp. 2441–2450. ACM (2010)
12. Potter, R.L., Weldon, L.J., Schneiderman, B.: Improving the accuracy of touch screens: An experimental evaluation of three strategies. In: Proc. CHI 1988, pp. 27–32. ACM (1988)
13. Rekimoto, J., Nagao, K.: The world through the computer: Computer augmented interaction with real world environments. In: Proc. UIST 1995, pp. 29–36. ACM (1995)
14. Rohs, M., Oulasvitra, A.: Target acquisition with camera phones when used as magic lens. In: Proc. CHI 2008, pp. 1409–1418. ACM (2008)
15. Rohs, M., Oulasvitra, A., Suomalainen, T.: Interaction with magic lens: Real-world validation of a Fitt's law model. In: Proc. CHI 2011, pp. 2725–2728. ACM (2011)
16. Rohs, M., Schleicher, R., Schöning, J., Essl, G., Naumann, A., Krüger, A.: Impact of item density on the utility of visual context in magic lens interactions. Journal Personal and Ubiquitous Computing 13(8), 633–646 (2009)
17. Roudaut, A., Huot, S., Lecolinet, E.: TapTap and Magstick: improving one- handed target acquisition on small touch-screens. In: Proc. AVI 2008, pp. 146–153. ACM (2008)
18. Vincent, T., Nigay, L., Kurata, T.: Classifying handheld augmented reality: Three categories linked by spatial mappings. In: Workshop on Classifying the Augmented Reality Presentation Space at ISMAR 2012 (2012)
19. Vogel, D., Baudisch, P.: Shift: A technique for operating pen-based interfaces using touch. In: Proc. CHI 2007, pp. 657–666. ACM (2007)
20. Wagner, J., Huot, S., Mackay, W.E.: BiTouch and BiPad: Designing bimanual interaction for hand-held tablets. In: Proc. CHI 2012, pp. 2317–2326. ACM (2012)

The Unadorned Desk: Exploiting the Physical Space around a Display as an Input Canvas

Doris Hausen[1], Sebastian Boring[2,3], and Saul Greenberg[3]

[1] University of Munich (LMU), HCI Group, Amalienstr. 17, 80333 Munich, Germany
[2] University of Copenhagen, Njalsgade 128, Bldg. 24, 2300 Copenhagen S, Denmark
[3] University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada
doris.hausen@ifi.lmu.de, sebastian.boring@diku.dk,
saul.greenberg@ucalgary.ca

Abstract. In everyday office work, people smoothly use the space on their physical desks to work with documents of interest, and to keep tools and materials nearby for easy use. In contrast, the limited screen space of computer displays imposes interface constraints. Associated material is placed off-screen (i.e., temporarily hidden) and requires extra work to access (window switching, menu selection) or crowds and competes with the work area (e.g., palettes and icons). This problem is worsened by the increasing popularity of small displays such as tablets and laptops. To mitigate this problem, we investigate how we can exploit an unadorned physical desk space as an additional input canvas. With minimal augmentation, our *Unadorned Desk* detects coarse *hovering over* and *touching of* discrete areas ('items') within a given area on an otherwise regular desk, which is used as input to the desktop computer. We hypothesize that people's spatial memory will let them touch particular desk locations without looking. In contrast to other augmented desks, our system provides optional feedback of touches directly on the computer's screen. We conducted a user study to understand how people make use of this input space. Participants freely placed and retrieved items onto/from the desk. We found that participants organize items in a grid-like fashion for easier access later on. In a second experiment, participants had to retrieve items from a predefined grid. When only few (large) items are located in the area, participants were faster without feedback and there was (surprisingly) no difference in error rates with or without feedback. As the item number grew (i.e., items shrank to fit the area), participants increasingly relied on feedback to minimize errors – at the cost of speed.

Keywords: Augmented desks, digital desks, peripheral interaction.

1 Introduction

In everyday office work, people naturally arrange documents, tools and other objects on their physical desk so that they are ready-to-hand, i.e., within easy reach and where they can be retrieved without actively searching for them. People are able to do so because they are aware of these objects' spatial location [15] and can coarsely acquire those that are in their peripheral vision. In contrast, working with computers requires

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 140–158, 2013.

almost everything to visually happen on-screen. Yet because space is limited, the so-called desktop metaphor usually separates object placement into one of several workspaces (see Figure 1a): the *primary workspace*, which covers most of the screen, holds the currently active document, which people normally work on; the *secondary workspace* is the portion of on-screen space that contains a subset of artifacts related to the primary space's activities, e.g., icons and tool palettes; finally, the *off-screen workspace* holds the remaining artifacts, where users – through a series of operations – make them explicitly visible in a temporary fashion (e.g., menus, dialog boxes).

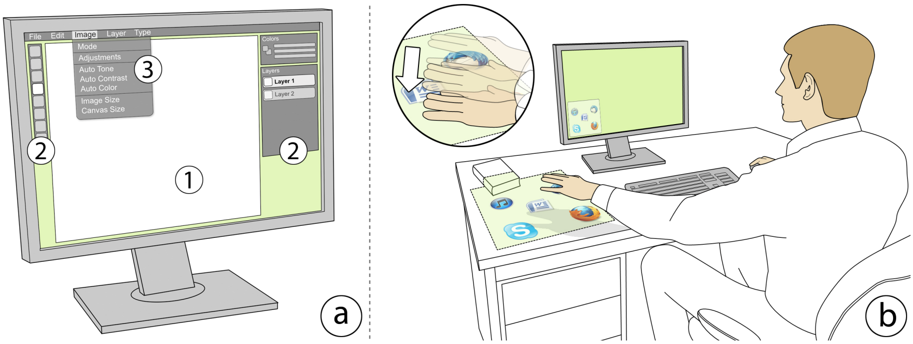


Fig. 1. (a) The three workspaces present in the desktop metaphor: the *primary workspace* (1) holds the active document people work on; the *secondary workspace* (2) holds items related to the activities in the primary workspace and is permanently visible; and the *off-screen workspace* (3) holds further items related to the document, yet people have to make them explicitly visible (e.g., menus). (b) The *Unadorned Desk* moves these workspaces onto a regular desk so that the primary workspace covers the entire display. When a person hovers over the interaction area on the desk, feedback may be given on-screen. Touching an item then selects it.

Yet, there is a tension between these workspaces. The primary and secondary workspaces spatially trade-off: the primary workspace dominates screen space, which leaves less space for its surrounding artifacts. This is especially true for tablets and other devices with rather small displays. The secondary and off-screen workspace also trade-off: it is much easier to select items in the secondary space, but only a few can be held there. In contrast, a huge number of items can be held in the off-screen workspace, but it is harder to select them (or to remember accelerator methods such as keyboard shortcuts) [21]. Instead of trying to fit everything on screen (directly or through menus), we investigate using the unadorned (i.e., unchanged except for a sensing device) desk as a further space to contain artifacts. Our hypothesis is that people can then easily select commonly used functions (e.g., tools or other windows) located on the desk's surface (see Figure 1b). This has several advantages. First, if we move artifacts from the secondary workspace to the desk, more display space can be allocated to the primary workspace. Second, if we move artifacts from the off-screen workspace to the desk, they will be easier to access. This also mimics the way we interact with everyday objects surrounding a document located on the desk (e.g., placing paints and brushes nearby for rapid retrieval while drawing).

Previous work on digital desks relies on a tight feedback loop, where visuals and interaction feedback were overlaid onto the regular desk surface i.e., by making the desk look and behave like a computer display. Examples include the use of projectors [20, 31], a tabletop computer as desk replacement [6], or by adding tablet computers next to the display as an interactive region [6]. These tend to be complex (or expensive) to set up. In contrast, the new generation of depth-sensing technologies mean that detecting touches and hovering is low-cost, such as via LeapMotion or Microsoft's Kinect camera. The problem is that these technologies do not provide visual feedback. This begs the question: is visual feedback on the desk necessary?

We are particularly interested in using the desk *as is* with the smallest possible alterations. In this paper, we take an extreme stance, where we provide either no feedback or feedback on-screen (rather than on the desk) solely on demand. Both approaches keep desk instrumentation to a minimum, thus allowing for the use of *any* desk – such as at cafes – to serve as a workspace. Using computer science terminology, this is a lower bounds investigation: we want to understand to what extent interaction is possible using minimal or no augmentation (i.e., no visual targets or confirmatory feedback on the desk).

To investigate how an *unadorned desk* can be used as input space, we built a prototype using a Microsoft Kinect depth camera mounted atop a regular desk. Our Unadorned Desk tracks a person's hand and allows for *hovering over* and *touching of* content. As we were interested in how people can interact with off-screen content while keeping their attention on their main task, feedback is either not provided, or is given on-screen and on demand. We conducted two experiments: the first *placement* experiment focused on placement strategies of participants. In the second *acquisition* experiment, varying numbers of virtual items were placed at predefined locations and participants had to retrieve them to find out which number is still usable for off-screen interaction. Our work offers two contributions: (1) a working prototype that makes use of an unadorned desk as input space by augmenting it with a depth camera. And (2), experimental results that inform the design of such interactions with respect to the amount of off-screen virtual items and the given on-screen feedback.

2 Related Work

Our work builds on several areas of research that relate to how people organize documents on their desk, peripheral and bimanual interaction, interfaces without direct visual feedback, and augmented desks in general.

Organizing the Desk. We routinely and fluidly arrange and manage documents on our physical desks without focusing much attention on it. We can do so because the document's physical arrangement on the desk offers context information about the status and importance of certain tasks [8]. Malone studied desk organizations and found that files and (even more so) piles are the most commonly used arrangements on a desk [24]. Files are usually ordered systematically (e.g., in an alphabetic order). Piles, however, are not organized deliberately, and people thus more likely use spatial organization for retrieval. Associated tools and materials are generally arranged so they are available for reuse, such as by placing them nearby and ready-to-hand during active use, or by organizing them into known locations (such as desk drawers) [12].

Many systems try to bring this traditional way of organizing a desk into the digital world. In *Data Mountain* [28], people can organize browser bookmarks on a virtual table, which proved to be faster than bookmarking in Internet Explorer 4. *BumpTop* simulates the desktop by allowing users to arrange documents in a virtual 3D space using physics [1]. Customization features in graphical user interfaces let people spatially arrange tools around the graphical desktop [12]. In contrast to these systems, we are interested in using the desk *as is* instead of mimicking it on-screen.

Augmented and Interactive Desks. There is a history of work where digital content is brought onto the surface of the physical desk. This not only provides a workspace larger than the constraints of a computer display, but – in some systems – also allows both physical and digital artifacts to be used in tandem. Early work focused on (partially) digitizing the desk. The *Digital Desk* [31] uses a projected interface on the desk. A video camera senses interactions with fingers and/or a pen, and can capture content of paper materials (i.e. interacting with paper). Rekimoto et al.'s *Augmented Surfaces* [26] are projected extensions to a laptop's display on a table or a wall. Users are able to drag content from their laptop onto the table where it is visible all times. Thus, the table serves as visual extension to the laptop's display. *Bonfire* [20] projects additional content next to a laptop's screen and allows touch input through cameras.

More recent prototypes augment the computer screen with a horizontal digital display ('surface') located underneath it. Surfaces typically allow for touch input, making sensing of user interaction easy (e.g., *Magic Desk* [6]). *Curve* [33] and *BendDesk* [30] merge the horizontal desk area and the vertical display area into one gigantic high-resolution touch-sensitive display, where they are seamlessly connected through a curve. Various studies investigated how particular touch regions on both the horizontal and vertical displays are used e.g., to show that the regions next to keyboard and mouse are best suitable for coarse interaction [6]. We build on this in that we use the areas left and right of the keyboard/mouse in our two studies.

Peripheral and Bimanual Interaction. Working with analogue documents on a desk often involves peripheral and bimanual interaction. *Peripheral interaction* offers coarse input styles in the periphery of the user's attention and thus quasi-parallel to the current primary task. The fundamentals for peripheral interaction are human capabilities such as divided attention (i.e., processing two tasks in parallel without switching channels [32]), automatic and habitual processes (i.e., carried out with little mental effort and hardly any conscious control [3]), and proprioception (i.e., being aware of one's own body, its posture and orientation [7]). Today's prototypes incorporating peripheral interaction mainly rely on TUIs (e.g., [4, 11, 17]) or freehand gestures [16, 18]. Our work adds to this by investigating how people interact coarsely in their periphery.

Bimanual (two-handed) interaction is the basis for peripheral interaction. While typically asymmetric, both hands influence each other leading to a kinematic chain [13]. Studies show that bimanual interaction can improve performance [9, 19]. At the same time, the body provides the kinesthetic reference frame, i.e., the user's sense of where one hand is relative to the body and the other hand [5]. Further, Balakrishnan et al. found that while separating visual feedback from the physical reference does affect performance, there is only a "remarkably small difference" when comparing interaction with and without visual feedback as long as "body-relative kinesthetic cues are available" [5]. We build on this as we separate feedback from interaction.

Interfaces without Direct Feedback. Spatial interaction does not necessarily rely on direct feedback or feedback at all. Gustafson et al.'s *Imaginary Interfaces* [14] make use of the visual short-term and visuospatial memory. By forming an "L" with the non-dominant hand a reference frame is created. *Spin & Swing* [2] depends on an imaginary circle around the user. By turning themselves, users navigate through the content displayed on a handheld device. The concept of *body-centric interactions* [10] employs the space around a person's body to hold mobile phone functions. For example, *Virtual Shelves* [22] positions items in a hemisphere in front of the user. *Point upon Body* [23] uses the forearm as interaction area, which can be divided at most into six distinct areas. *GesturePad* [27] and *BodySpace* [29] use different body locations for commands. As with our system, no direct feedback is provided. These systems rely primarily on spatial awareness and kinesthetic memory. Due to proprioception, users have a good understanding of where items are located and can easily – even with closed eyes – place and retrieve such objects [25]. These findings inspired us to mimic regular desk use as means for interacting with digital content.

3 Evaluating off-screen Interactions

In order to better understand how users can adapt to the novel input technology as well as how on-screen feedback for off-screen content would affect the interaction, we conducted two user studies. The first experiment aimed at understanding how people would spatially place various content items onto the desk that they would later retrieve. More precisely, we wanted to see whether people make use of special arrangements of their content. In the second experiment (which was tuned to use the results of the first study), we aimed to see how accurately participants could locate items placed in off-screen space as a function of the number of items in that space. The next section details the conditions and apparatus common to both experiments.

3.1 Conditions Common to Both Experiments

Although the tasks varied in both experiments, we had two conditions (additional to the experiment-dependent ones) that were the same in both studies: (1) the hand with which participants interacted in off-screen space, and (2) the type of feedback given during the task. In the following, we describe these two conditions in more detail.

Handedness: We chose to test our system with both hands. In the **dominant** hand condition, participants interacted with off-screen content using the hand they usually use to perform precise interactions (e.g., writing). In the **non-dominant** hand condition, they used the other hand. For each of the conditions, the interaction area was placed on the desk so that it was closest to the hand with which they had to interact in off-screen space (i.e., not reaching left of the keyboard using the right hand).

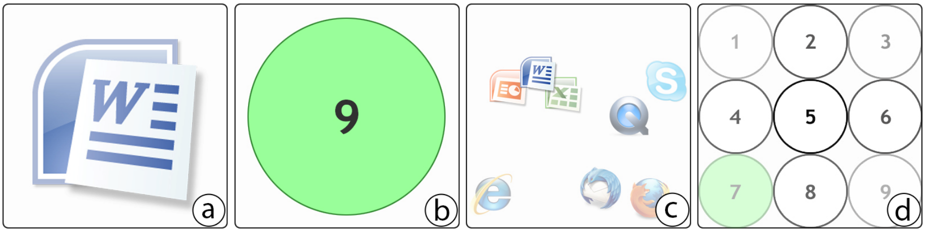


Fig. 2. In the *Single* feedback condition, the system showed the item closest to the participant's hand (a: 1st study, b: 2nd study). In the *Full* feedback condition, all items are shown with their correct spatial layout (c: 1st study, b: 2nd study). Here the participant hovers over the word icon (and item #5 respectively). In all conditions, transparency encoded the distance to that item.

Feedback: We had three conditions for on-screen feedback. In the **No Feedback** (**None**) condition, participants did not receive any feedback on the computer's display, forcing them to rely solely on their spatial memory and proprioception. In the **Single Item Feedback** (**Single**) condition, participants only saw the item that was closest to their hand, with the distance being encoded through transparency. That is, as participants moved closer to a respective item, the item's icon became increasingly opaque (see Figure 2a,b). In the **Full Area Feedback** (**Full**) condition, participants saw all items in the interaction area with correct spatial layout. As in the *Single* condition, the transparency of items again changed based on the distance between them and the participants' hands (see Figure 2c,d). That is, the item directly below the hand was more opaque than the surrounding items. The feedback area (400 × 400 pixels) was only shown on-screen while a participant's hand was inside the interaction area and invisible otherwise to not occupy valuable screen space. It was also located close to the interaction area (i.e., the bottom left or right corner of the display).

We used a within-subjects factorial design in both experiments: *2 Handedness* (*Dominant*, *Non-Dominant*) × 3 *Feedback* (*None*, *Single*, *Full*). *Feedback* was counterbalanced across participants. To minimize changing the camera's location for *Handedness*, we alternated participants so that the first participant had all three *Feedback* types with the *Dominant* hand and then again with the *Non-Dominant* one, while the second one started with *Non-Dominant* etc.

3.2 Apparatus, Setup and Participants

The *Unadorned Desk* uses a Microsoft Kinect depth camera mounted on a tripod facing upside down (see Figure 3) observing a sub-region of the desk within which a person could interact using the hand. The prototype runs on an Intel i7 3.4 GHz computer to allow for fast processing (i.e., 640 × 480 pixel frames at 30 frames per second).

Fig. 3. The Unadorned Desk: a Kinect tracks the user's hand

We use the Kinect depth camera to gather hand information within the tracked region. The camera provides depth images where each pixel in a depth frame encodes that pixel's distance to the camera in millimeters. At startup, the system takes a series of depth images, averages them (to reduce noise), and uses them as ground truth. Once running, it calculates the difference between the current depth frame and the calibrated depth image. The calculated difference image contains all points that are 'new' to the scene (e.g., a hand) with their distance to the desk. Using this point cloud, the system calculates the point of the hand closest to the corner of the interaction area that is the furthest away from the user (i.e., the tip of the middle finger). The vertical distance (depth) of that location to the desk further determines the hand's state: *touching* (depth < threshold), *hovering over* (depth ≥ threshold), or *absent* if no hand is detected. On-screen feedback is optionally provided once the user's hand enters the interaction area. When the hand touches an item, the system performs the action associated with that item.

In both experiments, participants were seated centrally in front of the computer's display. The depth camera captured a region of 40 cm × 36 cm (33.5 cm on the top edge due to slight camera distortion) next to the keyboard aligned with the desk's edges. For each *Handedness* condition, we moved the monitor, keyboard, mouse and chair to ensure that the participants are seated centrally in front of the display and close to the interaction area. The tracked region on the desk was empty. The computer display's background was set to a uniform color and had all desktop icons removed.

Each study used 12 participants. Sexes were mixed (first: study 6 female; second: 4 female), and ages ranged from 19 to 30 (average was 24). Each person only participated in one of the studies to minimize learning effects. Handedness varied, 9 were right-handed in the first study, and all in the second. Each session lasted up to 1.5 hours, and all participants received $20 as compensation for their time.

3.3 Hypotheses

We had similar hypotheses in both studies:

H1. Item retrieval time would increase as the number of off-screen items increased.
H2. Error rate would increase as the number of off-screen items increased.
H3. Item retrieval time would increase when no feedback was present.
H4. Offset and error rates would decrease with feedback present.

4 Study 1 – Placing and Retrieving Content

The purpose of our first study is to understand how participants would use of the *Unadorned Desk* to organizationally place and later retrieve an item, and the effect of having an increasing number of items placed within that space. In particular, we were interested in (1) how they arrange a given number of items on their desk, and (2) the offset (and the item's size respectively) when retrieving items to ensure successful pointing in the periphery. Our *Handedness* × *Feedback* factorial design was extended to include a third *Sets* condition, which is the number of items participants had to place and retrieve in the off-screen space. We used well-known, easily identifiable applications, which had meaning to our participants: Word, Excel, Power Point, Firefox, Thunderbird, Skype, QuickTime, and Internet Explorer. For each condition, the amount of items was ascending (to increase difficulty): 2, 4, 6, and 8 different items.

4.1 Tasks and Procedure

The experiment consisted of two phases for each combination of *Handedness* and *Feedback*: placing items and later retrieving them. We instructed users to place items off-screen in a position of their own liking. However, items had to have a minimum distance of 47.6 millimeters (and 50 pixels respectively) to avoid overlaps of them, which would make retrieval more error-prone. Each set of items they had to place was shown on the monitor during the placement task (see Figure 4a), so that participants were aware of all items and could group them if that would aid their memory. To place an item, participants first had to hit the spacebar to indicate they were starting the task, at which point timing began. Once the trial was active, they could move their hand into the interaction area and place the item by touching the desk's surface. When feedback was given, already placed items were shown to give participants a feeling of the location of other items (see Figure 2a,c). Participants repeated this step until they had placed all icons in the current set in the physical off-screen space.

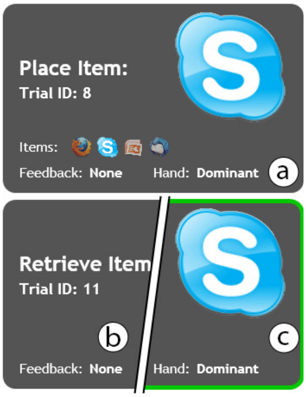


Fig. 4. Commands: Placing an item (a), and retrieving it (b: before trial activation, c: after)

For retrieval, the system notified participants on-screen of which item to retrieve before the trial began (Figure 4b). They then had to hit the spacebar to activate the trial (Figure 4c). Users would then retrieve that previously placed off-screen item. Retrieval worked exactly like the placement: hit spacebar for time measurement and touch a location to retrieve the item. Afterwards, the system prompted them with the next item until all items were retrieved. If the wrong item was retrieved, the participant was not informed, the trial was not repeated and the experiment continued but the error was recorded. For each *Feedback* and *Handedness* combination, participants placed 4 *Sets* of items (*2*, *4*, *6*, and *8* items) once and then retrieved each of them *4* times. We collected *24* placement sets (*480* item retrievals).

For placement, we recorded all *x,y* locations (as the center) of placed items. For the retrieval task, we measured the time from the beginning of a trial (i.e., hitting the spacebar) until they touched the desk's surface. We further recorded the location they touched, the distance to the actual item (*x,y* location), and the amount of items that were closer than the correct one (i.e., errors). We manually counted the participants' gazes, whether they looked at the interaction area, the feedback area, or both (the experimenter pressed a key for each gaze, which was recorded). Finally, we asked participants to fill out a device assessment questionnaire: once after completing one *Feedback* and *Handedness* condition, and again at the end of our study.

4.2 Results

We used *heat maps* to uncover how people would freely place items on the desk. We then compared *retrieval time*, *retrieval offset*, and *gazes* using repeated measures within-subjects analyses of variance (ANOVA). For pair-wise post hoc tests, we used Bonferroni-corrected confidence intervals to retain comparisons against $\alpha = 0.05$. When the assumption of sphericity was violated, we used Greenhouse-Geisser to correct the degrees of freedom. All unstated p-values are $p > 0.05$.

We performed a 2 × 3 × 4 (*Handedness* × *Feedback* × *Items*) within-subjects ANOVA. As we did not find any significant main effects or interactions for *Handedness*, we aggregated over *Handedness* for all subjects in subsequent analyses. For heat map analysis, we mirrored interactions performed in the area right to the keyboard to bring those into the coordinate system of the one left to the keyboard.

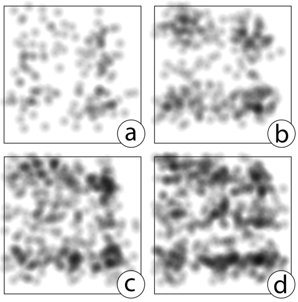


Fig. 5. Heat maps (a-d: 2 to 8 items) show that users tend to arrange items in grids

Strategic Placement of Items. Through a heat map analysis (see Figure 5) we found that many participants tended to arrange items based on an imaginary grid (thus item placement was not random). Further, participants followed other semantic patterns: first, some placed items in a single row as in the dock in Mac OS X. During retrieval with feedback, participants then hovered over that line to find the correct item. Second, some hierarchically grouped similar items together (e.g., all browser icons). They would later retrieve the item by first going to the general group area containing that item, and then selecting the particular item. Finally, the more frequently they use an application based on their personal usage outside the study, the closer they would place it to the keyboard. Items used less often are thus further away from the primary interaction space. Participants did consider that areas further away would require more physical effort to access an item. However, all participants made use of the *entire* area, as they felt more comfortable to access items placed further apart from each other.

We calculated three *Distances* (*Closest*, *Average*, and *Highest*) between items that they had placed off-screen. Participants placed items with an average distances for all conditions between 207.4 and 231.6 millimeters (M=219.2; SD=9.7). To understand whether *Feedback* or the *Set* of items had an influence on the distances between items, we performed separate 3 × 4 (*Feedback* × *Set*) ANOVAs for each *Distance*. For the closest distance, we found a significant main effect for *Set* ($F_{1.953,21.487} = 184.76$, $p < 0.001$) and post hoc multiple means comparisons revealed that the distance increases with a decreasing *Set* of items (all pairs except *6* and *8* items differ with $p = 0.011$) regardless of *Feedback*. *Feedback* had an effect on the highest distance between items, where we found significant main effects for both *Feedback* ($F_{2,22} = 15.49$, $p < 0.001$) and *Set* ($F_{3,33} = 128.74$, $p < 0.001$). Smaller *Sets* lead to lower distances between items except for *6* and *8* items (all $p < 0.001$). More importantly, in the *None* feedback condition, participants placed items further away. The differences further increase with the

Set size. Particularly for *8* items, *None* significantly differed from the other two (all $p < 0.05$), and from *Single* for *Set* sizes *2* and *6* (all $p < 0.05$). Thus, when relying on feedback, participants felt more comfortable placing items closer to each other. Interestingly, *Single* and *Full* did not differ for any *Set* size, and there was no significant difference between all three conditions for the *Set* with 4 items, which we attribute to participants using the four corners of the area.

Retrieval Time. We compared retrieval times from the moment participants hit the spacebar until they retrieved an item. We only took into account the correct retrieval times (even so, we did not find significant differences between retrieval times with and without errors). We performed a 3 × 4 (*Feedback* × *Set*) within subjects ANOVA and found significant main effects for *Feedback* ($F_{2,20}$ = 31.098, $p < 0.001$) and *Set* ($F_{1.609,17.698}$ = 15.583, $p = 0.011$). Figure 6a suggests that retrieval times slightly increase with larger *Sets*. However, *Feedback* influences retrieval times. Separate ANOVAs for each *Set* showed that *No Feedback* was always faster (all $p < 0.001$). Furthermore, the two conditions with visual feedback were more strongly affected by the *Set* of items. Overall, *None* was the fastest (*M*=1.40s, *SD*=0.36s), followed by *Full* (*M*=2.47s, *SD*=0.88s), and *Single* (*M*=2.68s, *SD*=1.06s).

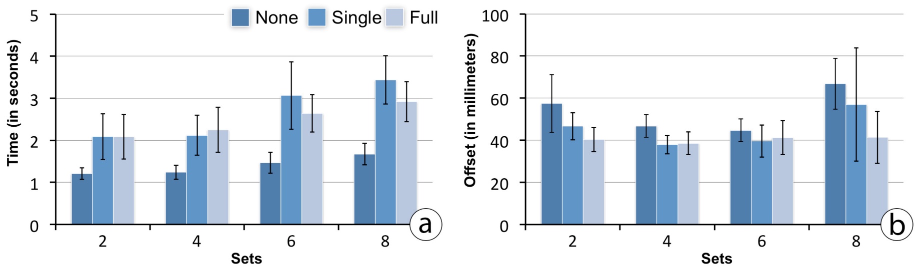


Fig. 6. Results of the placement study: (a) retrieval time for one item for all feedback conditions and sets; (b) offset for correct retrievals measured as Euclidean distance between the item's center and the touch's location. Error bars represent 95% confidence intervals.

Offset. We compared the offset (the distance between the touch point and the item's center). We chose to only include successful retrievals to eliminate cases where participants did not remember an item's location and thus randomly touched the desk. We performed a 3 × 4 (*Feedback* × *Set*) within subjects ANOVA and only found a significant main effect for *Feedback* ($F_{2,22}$ = 4.201, $p = 0.027$) but no effect for *Set* and no interactions. Figure 6b summarizes the results for different *Sets* and *Feedbacks*: *Full* had the smallest offset between the touch and the item's center (*M*=36.6 millimeters), followed by *Single* (*M*=41.1 millimeters) and *None* (*M*=48.9 millimeters). Figure 6b also reveals that, in order to have 95% successful selections regardless of *Feedback* and *Set*, an item with a radius of at least 85 millimeters is sufficient.

We were also interested in the impact of *Feedback* on wrong retrievals (i.e., touch was closer to an incorrect item than to the correct one). We normalized the data by dividing the number of incorrect closer items by the maximum number of possible wrong items (i.e., 1 for a set of 2, 3 for a set of 4, etc.). We performed a 3 × 4 (*Feedback* × *Set*) within subjects ANOVA and found a significant main effect for *Feedback* ($F_{1.22,13.419}$ = 4.914, p = 0.039). Post hoc tests revealed that only for a *Set* with *6* items *None* was more error-prone than the other two conditions (p = 0.04). We believe that – particularly with no visual cues on the desk – participants made use of space to more easily retrieve an item. That is, a larger offset still leads to correct retrieval. In summary, the chance for an erroneous selection with *None* is 20% (*SD*=12%), and 15% (*SD*=8%) for *Single* and *Full*. This can be lowered, however, by increasing the required minimum distance between items.

Gaze Analysis. We told participants to minimize looking at the interaction area, and instead imagine that they were concentrating and looking at their primary on-screen task. We did not instruct them with respect to using the feedback window, that is, they could freely make use of it. We report gazes averaged across both placement and retrieval phase. There were no gazes to the feedback area in the *None* condition.

For *Gazes* to the *Interaction Area*, we performed a 3 × 4 (*Feedback* × *Set*) within subjects ANOVA and found significant main effects for *Feedback* ($F_{1.126,12.383}$ = 7.948, p = 0.012), *Set* ($F_{3,33}$ = 14.494, p < 0.001) and a *Feedback* × *Set* interaction ($F_{2.15,23.645}$ = 8.618, p < 0.001). Post hoc tests revealed that for *6* items participants gazed at the interaction area more often in the *None* condition compared to *Single* (p = 0.027). For *8* items, they gazed more often using *None* compared to the other two conditions (all p = 0.016). For *Gazes* to the *Feedback Area*, we did not test the *None* condition (as there was no feedback area) and performed a 2 × 4 (*Feedback* × *Set*) within subjects ANOVA and found a significant main effect for *Set* ($F_{2.121,23.334}$ = 7.274, p = 0.002). Pairwise comparisons showed that *Gazes* to the *Feedback Area* increase with larger *Sets* (*2* and *4* differ from *6*, all p = 0.044, and *2* differs from *8*, p = 0.021). Overall, when *No Feedback* was presented, participants gazed at the interaction area on the desk more often (0.24 times per trial), compared to *Single* (0.11) and *Full* (0.12). In conditions that had *Feedback*, participants gazed at the feedback area 0.72 (*Full*) and 0.66 (*Single*) times per trial. Thus, participants 'left' their fictive primary task more often (i.e., looked away from it) when feedback was presented.

4.3 Discussion

During placement, we observed that participants used the whole interaction area, even though they stated that retrieval was easier if the item was placed closer to them. Placement was reasonably systematic, each following some kind of spatial organization. We noticed an increased time for placement and found significant differences for item distances with *No Feedback*. We believe that participants put more effort into a good arrangement (with reasonably spaced items) to allow for easier retrieval afterwards, which was especially important when there was no visual feedback.

During the retrieval stage, the *None* condition caused two problems for participants: (1) they had to remember where they put items, and (2) they were not informed whether they actually had correctly acquired an item. Interestingly, participants stated afterwards that – when feedback was provided – they felt pressured to point more precisely although this would not have been necessary (i.e., the selected item was always the one closest to the touched location), resulting in longer retrieval times for conditions with feedback. One participant stated that he started to search instead of think, which slowed him down. Our analysis of gazes supports this view: participants more often looked away from their fictive primary task when feedback was given. In fact, they looked more at the feedback area (when available) than at the interaction area when no feedback was given. *Feedback* did help participants to remember locations and decreased their offset for larger *Sets*, but also slowed them down.

Recall that these interaction techniques are to allow coarse interaction in the periphery (preferable with minimal attention). Our results suggest a suitable tradeoff between the item's sizes and the overall number of items. We observed that participants had problems recalling their spatial layout with *6* or more items. Nevertheless, the results also indicate that participants were able to successfully retrieve *2* or *4* items – even without feedback. While the number of manageable items in real life scenarios could be quite large (e.g., participants may want to place many items meaningful to their task on the interaction area), others have argued that a small number of such items could comprise a large number of the actions people actually do [12]. Examples are frequently or recently used commands. Nevertheless, this first experiment suggests that having more items decreases the probability of a correct retrieval. Quite possibly, our results could be affected by less than optimal placements of items on the desk, e.g., due to a lack of visual cues on the experiment's desk. For this reason, we conducted a 2nd study that spatially separated items into a grid (a layout applied by many in this first study), and that did not require to memorize locations, which is hard to achieve anyhow in a lab study setting, especially for long-term memory.

5 Study 2 – Targeting Content

To prevent memorizing (our lab study is only able to test short-term memory) where items were placed and eliminate the potential influence of unfavorable placement, we presented our participants with a predefined layout. Based on the 1st study, where participants had arranged items in a grid, we created grid-like layouts with pre-placed items, which was visible to them on-screen during each of the trials. We added a variable *ItemSize* with three levels: *Small* (10 cm wide), *Medium* (13.3 cm), and *Large* (20 cm). To fill the entire interaction area, we decided to fill the grid accordingly. That is, we had 16 (4 × 4) *Small*, 9 (3 × 3) *Medium*, and 4 (2 × 2) *Large* items. In this experiment, we were interested in getting more insights on item locations and size with respect to retrieval time, offset and errors.

5.1 Task and Procedure

Both, task and procedure were similar to the retrieval task of the first study (though items are already pre-placed on the desk). At the beginning of each trial, the system showed participants which item they had to retrieve (see Figure 7a-c). As before, they activated the trial by hitting the spacebar and retrieved the respective item from off-screen space by touching the respective location. If they retrieved the correct item, the system prompted them with the next item to retrieve. If they touched the wrong one, the system notified them that the trial was incorrect, increased the item's error count, and asked them to retrieve it again. However, to avoid frustration, the system moved on to the next item after three failed attempts. Participants had to retrieve each of the different *ItemSizes* three times for all *Handedness* and *Feedback* combination, thus requiring every participant to perform 522 retrievals. However, we excluded the first block as training block. We logged: *task time* from the moment the spacebar was hit until they either successfully retrieved the item or missed it; the *Euclidean distance* (i.e., *offset*) of the touch to the item's center; and the number of *errors* (we allowed a maximum of 3 errors per item). As in the first study, we manually tracked whether the participant looked at the interaction area on the desk, on the feedback area on the screen or both of them. After each *Feedback* and *Handedness* combination, participants filled out the same device assessment questionnaire used in the first study as well as a closing questionnaire.

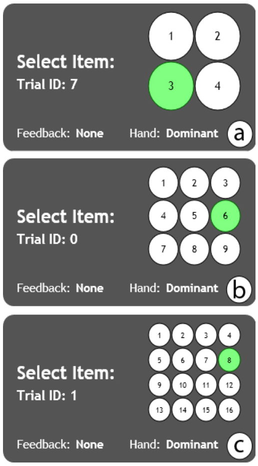


Fig. 7. The target item (green), among all other items. a-c: *Large*, *Medium*, and *Small* items

5.2 Results

We performed a 2 × 3 × 3 (*Handedness* × *Feedback* × *ItemSize*) within subjects ANOVA. As in the first study, we did not find any significant main effects or interactions for *Handedness*. Thus, in subsequent analyses, we aggregated over *Handedness* across all participants. We also excluded all erroneous, unsuccessful retrievals from analyses of retrieval time and offset, as we ended a trial after three incorrect retrievals). Because of this, we excluded 6.5% of all trials.

Retrieval Time. Regarding retrieval time for an item, we performed a 3 × 3 (*Feedback* × *ItemSize*) within subjects ANOVA and found significant main effects for *ItemSize* ($F_{1.272,13,997} = 15.269$, $p < 0.001$) and *Feedback* ($F_{2,22} = 19.037$, $p < 0.001$). We further found an *ItemSize* × *Feedback* ($F_{4,44} = 5.414$, $p < 0.001$) interaction. Post hoc multiple means comparisons showed that for all *ItemSizes* retrieval time differed significantly for the *None* condition (users needed less time) compared to the other two (all $p = 0.017$). Further, for *Single* and *Full*, the retrieval time for the *Small* items differed significantly from the shorter retrieval time for the *Medium* and *Large* items ($p < 0.001$). Overall, *None* was the fastest (M=1.68s), followed by *Single* (M=2.25s), and *Full* (M=2.33s). Figure 8a summarizes these results.

Retrieval Offset. For the analysis of the offset of successful retrievals (measured as Euclidean distance between the touch and the item's center), we normalize the distance as we had different *ItemSizes*. To do so, we divided the measured offset by the

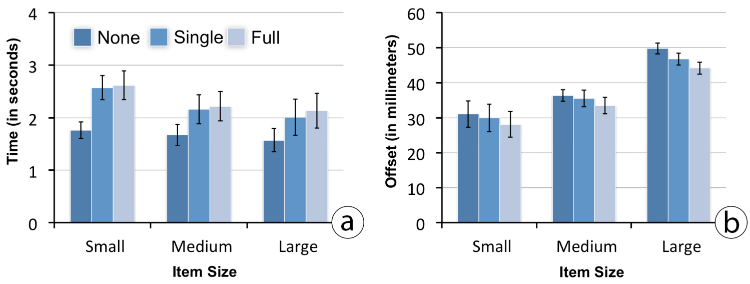


Fig. 8. Results of the targeting study: (a) retrieval time for one item for all feedback conditions and item sizes; (b) offset for correct retrievals measured as Euclidean distance between the item's center and the touch's location. Error bars represent 95% confidence intervals.

maximum possible offset (i.e., the item's actual size). With the normalized data, we performed a 3 × 3 (*Feedback* × *ItemSize*) within subjects ANOVA and found significant main effects for *ItemSize* ($F_{2,22}$ = 39.318, p < 0.001), and *Feedback* ($F_{2,22}$ = 4.918, p = 0.016), but no *Feedback* × *ItemSize* interaction. Pairwise comparison of different *ItemSizes* across all *Feedback* conditions further revealed that participants were always relatively closer to the item's center (yet physically further away) for *Large* items (p = 0.007). Overall, participants had the smallest offset for *Large* items (46.9% of the item's width), followed by *Medium* (52.1%), and *Small* (59.5%). However, when looking at the non-normalized offset (see Figure 8b), the results are the exact opposite: participants had the least offset for *Small* items (29.7 mm), followed by *Medium* (34.6 mm) and *Large* (46.9 mm) ones.

We normalized errors since we had a different amount of items depending on the *ItemSize*. We divided the errors by the number of items in the grid for each trial. With these values, we performed a 3 × 3 (*Feedback* × *ItemSize*) within subjects ANOVA and found significant main effects for *ItemSize* ($F_{2,22}$ = 88.909, p < 0.001), *Feedback* ($F_{1.309,14.4}$ = 10.587, p = 0.002), and a *Feedback* × *ItemSize* ($F_{2.126,23.385}$ = 4.036, p = 0.028) interaction. Post hoc tests showed that the *None* condition differed significantly from the other two for the *Large* (all p = 0.018) and from the *Full* condition for the *Small* items (p = 0.008). However, *Feedback* conditions do not differ significantly for the *Medium* ones. For all *ItemSizes*, *None* was the most error prone (*M*=0.41, *SD*=0.23), followed by *Single* (*M*=0.22, *SD*=0.16) and *Full* (*M*=0.18, *SD*=0.10).

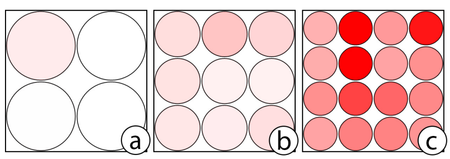


Fig. 9. Heat maps showing errors (aggregated upon all feedback conditions, mirrored for the right interaction area) for a) large, b) medium and c) small items. Saturation indicates errors.

To understand the error-prone performance, Figure 9 visualizes the locations where users had the most errors as heat map. As trend, one can see that for *larger* items the corner furthest away from the user caused the most errors. However, the *smaller* items get, the more errors occur in the center, which can be explained by the desk's edge (and the borders of the interaction area respectively) being a reference frame. This made it easier to target items close to the borders and harder in the center. In a second analysis, we excluded the items further away: for *Large* items, we excluded the top left item, for *Medium* items the three items furthest away, and for *Small* items the six items furthest away. We performed the same *3* × *3* (*Feedback* × *ItemSize*) within subjects ANOVA using the reduced set and found significant main effects for *ItemSize* ($F_{2,22} = 23.941$, $p < 0.001$), *Feedback* ($F_{1.332,14.648} = 9.973$, $p = 0.003$), but no *Feedback* × *ItemSize* interaction. Post hoc tests revealed that both, *Single* and *Full*, differed significantly from *None* only for *Small* items (all $p = 0.045$). This substantiates that the corner furthest away was the most error-prone. Nevertheless, *None* is still the most error-prone across all *ItemSizes,* with the least errors for *Large* items with 0.037 errors per trial (*Single*: 0.012, *Full*: 0.019).

Gaze Analysis. We instructed participants in the same way as we did in the first experiment. For *Gazes* to the *Interaction Area*, we performed a within subjects ANOVA on *Feedback* and found a significant main effect ($F_{1.136,12.495} = 10.485$, $p = 0.004$). Multiple means comparisons revealed that users gazed more often at the *Interaction Area* in the *None* condition compared to the other (all $p = 0.022$). We again excluded the *None* condition for *Gazes* to the *Feedback Area*, and performed a within subjects ANOVA on the remaining two *Feedback* factors and did not find a significant effect.

Overall, *None* had the most gazes to the interaction area (0.23 times per trial), compared to *Single* (0.05) and *Full* (0.06). In *Feedback* conditions, participants gazed at the feedback area 0.69 (*Full*) and 0.65 (*Single*) times per trial.

5.3 Discussion

The second study re-enforces the findings from the first study. As before, *No Feedback* led to shortest retrieval times. Retrieval time also increased for *Small* items when feedback was present, yet it did not change when no feedback was given. Naturally, *Small* items required participants to select more precisely. The absolute offset from the center of an item for *Large* items (with 4.69 cm) would almost not suffice for *Small* items (as they only had a radius of 5 cm and a width of 10 cm respectively). Users seemed to make use of space for larger items (it did not matter how close to the center they touched the item) and adjusted their offsets for smaller ones.

No Feedback caused significantly more errors with the corner further away from the user included in the analysis. Similar to Magic Desk [6], where Bi et al. found that completion time was longer for areas further away from the keyboard, our users had problems acquiring targets further away. When we excluded items further away from analysis (i.e., only considering that half of the interaction area closer to the participant), the *No Feedback* condition only differed significantly from the others for the *Small* items. However, the error rate for *Small* items was high regardless of feedback. Thus, items with a size of 10 cm or less are generally too small to be manageable in the periphery on an unadorned desk independent of the provision of feedback.

6 General Discussion

In both studies, more items in the interaction area require lower offsets between a touch and the item's center – in the first study to ensure that the correct one is still the closest item, and in the second study because items got *smaller* as their number increased. As we hypothesized in H1, both studies showed that retrieval time increases as the number of items in the interaction area increases. While H2 suggested that error rate increases with more off-screen items, our experiments only partially support this. We did not find evidence for more errors when increasing the item number (up to *8*) in the first study. Similarly, we did not find a significant effect in the second study for *Medium*-sized items, but did find a significant effect for *Small* ones. Thus, H2 (i.e., more errors with more items) is only supported for *10* or more items. H3 suggested that participants' time to retrieve items would increase when no feedback was present. Indeed, in both studies retrieval times were shorter when participants did not have *Feedback*, which fully supports this hypothesis. And finally, in H4 we hypothesized that the participants' offsets would increase and their error rate decrease when feedback was given. Yet, our results at best show a tendency towards more errors and larger offsets without feedback. In the first study, there was no significant effect for offset, and a significant effect on errors only for *6* items, but not for *8*. In the second study, we found an effect for *Small* and *Large* items (but not for *Medium* ones) between *No Feedback* and *Full Feedback* (yet not for *Single*). When only analyzing that half of the interaction area closer to the participant, *No Feedback* only differs significantly from the other two for *Small* items. Thus, our results therefore do not support H4 and only show a tendency towards *No Feedback* increasing offset and errors.

The first study showed that participants made use of the whole interaction area, even with a small number of items. In the second study, we found that items located closer to keyboard and mouse, are less error prone than those located further away. This suggests that a rectangular shape might not be the most suited interaction space. In in-situ experiments, however, users would have a better reference frame (i.e., items on the table that convey meaning) instead of just the blank desk – which ultimately would influence on the results.

Our study showed that simple interaction on an unadorned desk is possible, albeit with a modest number of items and a reasonable item size (the first study revealed 85mm to achieve 95% successful retrievals, which would have sufficed for the second study). As the number of items increased, both retrieval times and error rate increased as well. However, previous studies on peripheral interaction showed that this interaction style needs to be trained and learned to be effective [3, 17], which naturally is not possible in a short-term laboratory experiment. Abandoning feedback leads to faster retrieval times and functions (in terms of offset and errors) for a small numbers of items. Our findings suggest that the amount of items on the desk should be limited to less than ten. Similar to the shape of the interaction area, we expect this number to be higher if the desk contains more physical objects that serve as a visual cue or anchor and participants are used to the system and place meaningful items on the desk.

Overall, participants enjoyed interacting with the unadorned desk, and considered it to be fairly easy. All were able to carry out the interaction equally well with their dominant and non-dominant hand, which strengthens our understanding that it is a peripheral interaction style. Interestingly, some of them were also irritated by this kind of interaction as they thought that the entire hand (and its palm respectively) acts as input, where in fact only a single point of the hand was tracked. Nevertheless, those participants adapted to the interaction fairly quickly.

7 Conclusion and Future Work

We presented the *Unadorned Desk*, which supports peripheral coarse interaction and extends the input- and workspace beyond a computer's display. The *Unadorned Desk* relies on hand tracking by a depth camera (Kinect). Our studies showed that users are capable of interacting with virtual items on the desk, for small numbers of items even without on-screen feedback. It is a lower-bounds performance study, as we deliberately did not place anything on the desk's surface to indicate an item's virtual location.

Our current experimental implementation suffers from three limitations that restrict its deployment for everyday use. First, as with most optical tracking systems, the system is susceptible to false detections when sunlight hits the tracked area. That problem also occurs with our depth camera, as the sun's infrared light does interfere with the structured, infrared light of Microsoft's Kinect. Second, the system requires mounting a depth camera atop a desk, which is unsuitable for situations where rapid setup and teardown is required (e.g., temporary desks). This limits our ability to study the Unadorned Desk during anticipated everyday use. Third, the prototype does not yet address the fact that not every interaction on the desk is actually meant as input to the computer (e.g., retrieving a book). While emerging technologies will likely address the first two limitations, more research is needed to find an appropriate, distinct, yet not distracting gesture. Despite these limitations, our prototype allows us to evaluate implications for interaction on unadorned desks and to envision example applications such as those shown in the video figure[1].

There are still many unanswered questions for future work. Our first experiments were carried out in an artificial lab setting, which brings with it usual concerns about external validity. The primary task was placement and retrieval, rather than one's actual work. The items had no special significance. Interferences with other tasks carried out at the desk are not explored yet. Repeating the study in field cases could reveal nuances not seen in the lab. Our interaction area was rectangular, of a given size, and uncluttered; all these could both be varied to see how it affects performance. It was also in 2D (albeit with a hover plane). Yet a 3D interaction space is possible, e.g., virtual piles where a user can navigate through it with the hovering hand. Finally, ours was a lower bounds study of an unadorned desk. There could be many possible

[1] Video Figure of the Unadorned Desk: `http://youtu.be/ePQxR3EzJ_I`

ways of introducing modest adornments that indicate position. Although this would now introduce desk artifacts, it could improve performance significantly.

Acknowledgments. This work is partially funded by the iCORE/NSERC/SMART Chair in Interactive Technologies, Alberta Innovates Technology Futures, NSERS, and SMART Technologies Inc., and the German state of Bavaria. The authors thank Andreas Butz, Matthew Dunlap, Richard Fung, Ulrike Kister and the Interactions Lab at the University of Calgary for their valuable input and their support throughout this project. We further thank everyone participating in our two experiments.

References

1. Agarawala, A., Balakrishnan, R.: Keepin'it real: pushing the desktop metaphor with physics, piles and the pen. In: CHI (2006)
2. Altakrouri, B., Kawsar, F., Kortuem, G.: Spin&Swing: Spatial interaction with orientation aware devices. In: Pervasive (2010)
3. Bakker, S., van den Hoven, E., Eggen, B.: Design for the periphery. Eurohaptics (2010)
4. Bakker, S., van den Hoven, E., Eggen, B., Overbeeke, K.: Exploring peripheral interaction design for primary school teachers. In: TEI, pp. 245–252 (2012)
5. Balakrishnan, R., Hinckley, K.: The role of kinesthetic reference frames in two-handed input performance. In: UIST (1999)
6. Bi, X., Grossman, T., Matejka, J., Fitzmaurice, G.: Magic desk: bringing multi-touch surfaces into desktop work. In: CHI (2011)
7. Boff, K.R., Kaufman, L., Thomas, J.P.: Handbook of perception and human performance (1986)
8. Bondarenko, O., Ruud, J.: Documents at hand: Learning from paper to improve digital technologies. In: CHI, pp. 121–130 (2005)
9. Buxton, W., Myers, B.A.: A study in two-handed input. In: CHI, pp. 321–326 (1986)
10. Chen, X., Marquardt, N., Tang, A., Boring, S., Greenberg, S.: Extending a mobile device's interaction space through body-centric interaction. In: MobileHCI (2012)
11. Edge, D., Blackwell, A.F.: Peripheral tangible interaction by analytic design. In: TEI (2009)
12. Greenberg, S.: The computer user as toolsmith: The use, reuse, and organization of computer-based tools (1993)
13. Guiard, Y.: Asymmetric division on labor in human skilled bimanual action: The kinematic chain as a model. Journal of Motor Behavior 19(4), 486–517 (1987)
14. Gustafson, S., Bierwirth, D., Baudisch, P.: Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In: UIST, pp. 3–12 (2010)
15. Harrison, S., Dourish, P.: Re-place-ing space: The role of place and space in collaborative systems. In: CSCW, pp. 67–76 (1996)
16. Hausen, D., Boring, S., Polleti, J., Butz, A.: Exploring design and combination of ambient information and peripheral interaction. DIS Work in Progress (2012)
17. Hausen, D., Boring, S., Lueling, C., Rodestock, S., Butz, A.: StaTube: Facilitating state management in Instant Messaging Systems. In: TEI, pp. 283–290 (2012)
18. Hausen, D., Wagner, C., Boring, S., Butz, A.: Comparing Modaltities and Feedback for Peripheral Interaction. CHI Extended Abstracts (2013)

19. Kabbash, P., Buxton, W., Sellen, A.: Two-handed input in a compound task. In: CHI (1994)
20. Kane, S.K., Avrahami, D., Wobbrock, J.O., Harrison, B., Rea, A.D., Philipose, M., LaMarca, A.: Bonfire: a nomadic system for hybrid laptop-tabletop interaction. In: UIST (2009)
21. Lane, D.M., Napier, H.A., Peres, S.C., Sandor, A.: Hidden costs of graphical user interfaces: Failure to make the transition from menus and icon toolbars to keyboard shortcuts. International Journal of HCI 18(2), 133–144 (2005)
22. Li, F.C.Y., Dearman, D., Truong, K.N.: Virtual shelves: interactions with orientation aware devices. In: UIST, pp. 125–128 (2009)
23. Lin, S., Su, C., Cheng, K., Liang, R., Kuo, T., Chen, B.-Y.: PUB-Point upon body: Exploring eyes-free interaction and methods on an arm. In: UIST, pp. 481–487 (2011)
24. Malone, T.W.: How do people organize their desks?: Implications for the design of office information systems. ACM Transactions on Office Information Systems 1(1) (1983)
25. Mine, M.R., Brooks Jr., F.P., Sequin, C.H.: Moving objects in space: Exploiting proprioception in virtual-environment interaction. In: SIGGRAPH, pp. 19–26 (1997)
26. Rekimoto, J., Saitoh, M.: Augmented surfaces: A spatially continuous work space for hybrid computing environments. In: CHI, pp. 378–385 (1999)
27. Rekimoto, J.: GestureWrist and GesturePad: unobtrusive wearable interaction devices. Wearable Computers, 21–27 (2001)
28. Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., Dantzich, M.V.: Data Mountain: Using spatial memory for document management. In: UIST, pp. 153–162 (1998)
29. Strachan, S., Murray-Smith, R., O'Modhrain, S.: BodySpace: inferring body pose for natural control of a music player. In: CHI EA, pp. 2001–2006 (2007)
30. Weiss, M., Voelker, S., Borchers, J.: Benddesk: Seamless integration of horizontal and vertical multi-touch surfaces in desk environments. In: Adjunct Proceedings ITS (2009)
31. Wellner, P.: Interacting with paper on the DigitalDesk. Comm. of the ACM 36(7) (1993)
32. Wickens, C.D., McCarley, J.S.: Applied Attention Theory
33. Wimmer, R., Hennecke, F., Schulz, F., Boring, S., Butz, A., Hussmann, H.: Curve: revisiting the digital desk. In: NordiCHI, pp. 561–570 (2010)

GSR and Blink Features for Cognitive Load Classification

Nargess Nourbakhsh[1,2], Yang Wang[1], and Fang Chen[1]

[1] National ICT Australia (NICTA), Sydney, Australia
[2] The University of Sydney, Sydney, Australia
{nargess.nourbakhsh,yang.wang,fang.chen}@nicta.com.au

Abstract. A system capable of monitoring its user's mental workload can evaluate the suitability of its interface and interactions for user's current cognitive status and properly change them when necessary. Galvanic skin response (GSR) and eye blinks are cognitive load measures which can be captured conveniently and at low cost. The present study has assessed multiple features of these two signals in classification of cognitive workload level. The experiment included arithmetic tasks with four difficulty levels and two types of machine learning algorithms have been applied for classification. Obtained results show that the studied features of blink and GSR can reasonably discriminate workload levels and combining features of the two modalities improves the accuracy of cognitive load classification. We have achieved around 75% for binary classification and more than 50% for four-class classification.

Keywords: Cognitive load, galvanic skin response, eye blink, machine learning.

1 Introduction

Being continuously aware of user's mental status is an important step in making intelligent systems interacting with people. Such systems can properly change their interface and interactions to match the imposing workload with the current working memory of their user. In this way, the optimum performance will be obtained and many human errors will be avoided. Therefore, it is necessary to measure mental load accurately and in real-time.

Cognitive load is commonly used to refer to the load that performing a particular task imposes on the person's cognitive system [19]. Different methods have been applied for quantifying cognitive workload; however, not all of them are useful for developing adaptable systems. Subjective (self-reporting) [18] and performance-based measurement [3] techniques have been widely used and, regarding implementation, are usually the most convenient methods. However, asking subjects to rate the experienced mental workload means several interruptions and distractions from performing the principal tasks. Moreover, both methods are post-task processing and can be done when the task is finished, thus are not useful for real-time cognitive load assessment. In contrast, human behaviors and physiological responses can continuously and non-intrusively demonstrate user's cognitive states while performing the intended task.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 159–166, 2013.

Several physiological signals have been used for cognitive load measurement: signals from heart [20], eye [27], brain [2, 12] and skin [17]. Galvanic skin response (GSR), which is electrical conductance of skin, is a low-cost, easily-captured, robust physiological signal. Previous studies have used skin conductance in detecting emotions [15] or differentiating between stress and cognitive load conditions [16], and a few ones have found relations between GSR features and mental workload [17, 25]. Some others have tried but did not obtain satisfactory results for detecting cognitive load from GSR [7, 11].

Speech [10], pen input [21] and eye movements [5] are instances of behavioral signals used in cognitive load measurement. Eye activity can reveal valuable information about mental workload. In contrast with some eye based features (such as pupil dilation) which can only be gathered through an expensive eye tracker, eye blink can be obtained with an acceptable accuracy through a conventional camera. Therefore it is a low-cost and easily-obtained signal which can be used for cognitive load measurement. Some previous works have studied blink variations in regard to modality (visual versus acoustic) [13] or location (central versus peripheral) [6] of presenting stimuli. Another research has measured the blink rate in resting, reading and talking conditions [1]. A few studies have examined blink features in two cognitive load levels and found them related with mental load level [5, 8].

There are various application domains for cognitive load measurement, from brain-computer interactions to air traffic control. However, in some domains such as driving and education it is essential to be able to measure this load at a low cost, with short preparation time and minor restriction of users' movements. Considering that GSR and eye blink are suitable measures in such situations, in this paper we have explored features of these two signals captured during arithmetic experiments consisting of four cognitive load levels. We have assessed how useful every single feature is for classifying mental workload level and how combining features from the two signals affects the accuracy of cognitive load classification.

Support vector machines (SVM) and Naïve Bayesian classifiers are two popular machine learning algorithms in human-computer interaction studies [14]. Some previous works have used SVM for recognizing drowsiness [28] or different emotions [9, 22, 24] from physiological features. Naïve Bayesian classifiers have been used for detecting human emotions from facial expressions [23] or physiological signals [4, 9]. In this study, we have used these two types of classifiers for cognitive load classification of GSR and blink features.

2 Experiment

The data was collected from thirteen healthy 24 to 35-year-old volunteers who signed consent forms before the experiment and were awarded with movie vouchers for their participation. The experiment included 8 arithmetic tasks with 4 difficulty levels. Each subject performed two trials of each task level and the whole eight trials were performed in a randomized order. In each task four numbers were shown one by one, each for three seconds. Subjects were supposed to add-up these four numbers and select (by clicking the mouse using their right hand) the correct answer from three numbers which were next presented on the screen. First to fourth difficulty levels respectively included binary numbers (0 and 1), one-digit numbers, two-digit numbers and three-digit numbers. Before appearing the first number of each task, a slide containing one, two or three 'x' symbols (according to the number of digits in the task) was presented for three seconds. There was no time limit for answering and the background was always black. There was a 6-second rest time between consecutive tasks. After finishing the experiment, subjects rated task difficulty levels in a questionnaire (ranging from 1 to 9).

To collect galvanic skin response, the GSR device from ProComp Infiniti of Thought Technology Ltd was used and the sensors were attached to the subject's left hand finger (all subjects were right-handed). The sampling frequency was 10Hz. Eye activity data was recorded with a remote eye tracker (faceLAB 4.5 of Seeing Machines Ltd) which operated at a sampling rate of 50Hz and continuously recorded eye data. A 21" LCD monitor and a usual computer mouse were used for presenting the tasks and obtaining user inputs.

3 Cognitive Load Measurement

Figure 1 shows the average subjective ratings of the task difficulty levels. One-way analysis of variance (ANOVA) on the self-reporting scores showed a highly significant difference between task levels ($F_{3,48}$=108.63, $p<0.05$).

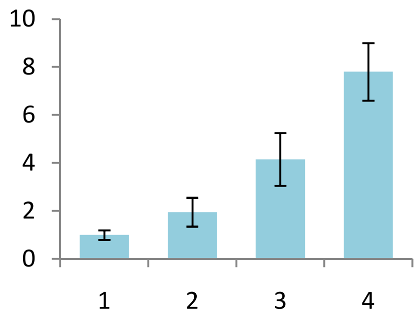


Fig. 1. Subjective rating of task difficulty levels

We also examined the response time for each task. By response time we mean the time between *disappearing the last (fourth) number* of the task and *selecting the answer*. Average response time of all subjects for each task difficulty level are shown in Figure 2. It can be seen that response time has a direct relation with the task difficulty level: harder tasks take longer response time. Results of ANOVA test on response time of different task levels are significant ($F_{3,48}$=62.59, $p<0.05$). These observations about subjective rating and response time show that the designed tasks have effectively manipulated the cognitive load.

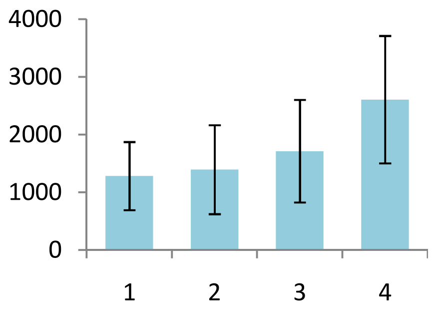


Fig. 2. Response time for each task difficulty level (milliseconds)

Two GSR and two blink features were calculated for each task:

- accumulative GSR (summation of GSR values over task time)
- GSR power spectrum (frequency power)
- blink number (number of blinks in the task)
- blink rate (number of blinks in the task divided by task time)

The time between *appearing the first number* and *inputting the answer* was considered as the task time in which every feature was computed. We observed that GSR and blink values are highly subjective, that is they differ from person to person. In order to omit the subjective differences, we calibrated each feature of task *j* of participant *i* by dividing it by the average of all similar features of all tasks of that subject:

$$calibrated\_feature(i,j) = \frac{feature(i,j)}{\frac{1}{m}\sum_{j=1}^{m} feature(i,j)} \qquad (1)$$

where m is the number of tasks (m=8). Furthermore, we averaged each feature between tasks with same difficulty levels for each subject. Figures 3 and 4 show the average values of the studied (calibrated) features in the four task levels.

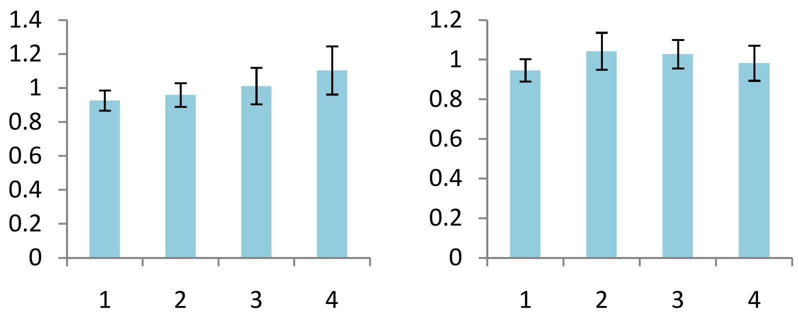


Fig. 3. Average GSR features of all subjects for the four task levels: accumulative GSR (left), GSR frequency power (right)

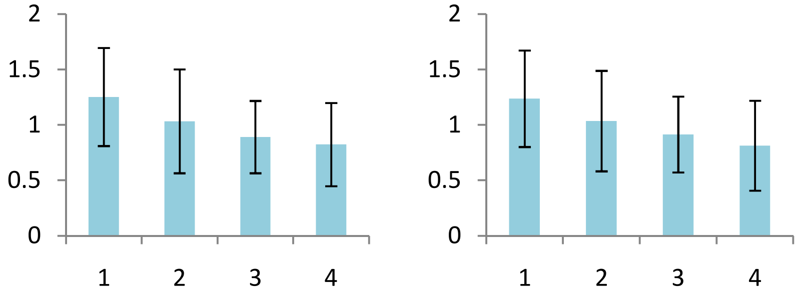


Fig. 4. Average blink features of all subjects for the four task levels: blink number (left), blink rate (right)

One-way ANOVA test was applied to statistically evaluate cognitive load level discrimination of each feature. Table 1 represents the results of statistical analysis (ANOVA test) of the studied features for four task difficulty levels. As can be seen, the results are significant for all four features.

Table 1. ANOVA results of features for four task difficulty levels

Feature	Results
Accummulative GSR	$F_{3,48} = 7.22$, $p < 0.05$
GSR Frequency Power	$F_{3,48} = 4.07$, $p < 0.05$
Blink Number	$F_{3,48} = 3.37$, $p < 0.05$
Blink Rate	$F_{3,48} = 3.22$, $p < 0.05$

4 Cognitive Load Classification

In this study, support vector machines (SVM) and Naïve Bayes classifiers were applied for cognitive load classification. For every feature, we have examined two- and four-class classification while the former means considering levels one and two as low load and levels three and four as high load. The cross validation method was leave-one-subject-out. In other words, in each round the classifier was trained by the data of the all subjects except one and data of the remaining subject was used for testing. The classification accuracies of all rounds were averaged.

Tables 2 to 5 show the cognitive load classification accuracies of the single features for two and four classes. Results of all features are reasonable, GSR features outperform blink features in two-class classification and results of blink number are better than those of blink rate. It is also worth mentioning that in most cases the two types of classifiers have very near or even similar (Table 4) performances on the single features. The largest difference is in classifying by use of accumulative GSR (Table 2) where classification accuracies of Naïve Bayes learners are about 5% higher than those of SVM in both 2-class and 4-class classifications.

Table 2. Classification accuracies of accummulative GSR

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	66.4%	34.6%
Naïve Bayes	71.2%	40.4%

Table 3. Classification accuracies of GSR frequency power

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	66.4%	37.5%
Naïve Bayes	65.4%	35.6%

Table 4. Classification accuracies of blink number

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	62.5%	40.0%
Naïve Bayes	62.5%	40.0%

Table 5. Classification accuracies of blink rate

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	57.5%	31.3%
Naïve Bayes	55.0%	32.5%

In the next step, we examined cognitive load classification using combinations of GSR and blink features. The combination of blink number and GSR frequency power resulted in the highest classification accuracies which can be seen in Table 6. Comparison with tables 3 and 4 reveals that combining blink number and GSR frequency power improves the classification accuracy in both two- and four-class classifications up to about 10% in the former and 16% in the latter. It can be observed that for combination of the two modalities (Table 6), similar to single feature cognitive load classifications, the classification accuracies of SVM and Naïve Bayes classifiers are close.

Table 6. Classification accuracies of blink number + GSR frequency power

Classification Algorithm	2-Class Classification	4-Class Classification
SVM	71.5%	53.6%
Naïve Bayes	75.0%	50.0%

5 Conclusion

We have applied classification algorithms on blink and GSR features and combinations of them. Accumulative GSR, power spectrum of GSR, blink number and blink rate were significantly distinctive and had reasonable accuracies in both two- and four-class classification of cognitive load using support vector machines and naïve Bayes classifiers. Combining GSR and blink features improved the classification accuracy. As our next step towards fully automated and more precise cognitive load detection, we will apply automatic feature selection (e.g. [26]) of physiological cognitive measures.

References

1. Bentivoglio, A.R., Bressman, S.B., Cassetta, E., Carretta, D., Tonali, P., Albanese, A.: Analysis of Blink Rate Patterns in Normal Subjects. Movement Disorders 12, 1028–1034 (1997)
2. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
3. Chandler, P., Sweller, J.: Cognitive Load While Learning to Use a Computer Program. Applied Cognitive Psychology 10, 151–170 (1996)
4. Chanel, G., Kronegg, J., Grandjean, D., Pun, T.: Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals. In: Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) MRCS 2006. LNCS, vol. 4105, pp. 530–537. Springer, Heidelberg (2006)
5. Chen, S., Epps, J., Ruiz, N., Chen, F.: Eye Activity as a Measure of Human Effort in HCI. In: International Conference on Intelligent User Interfaces, pp. 315–318. ACM Press (2011)
6. Fogarty, C., Stern, J.A.: Eye movements and blinks: Their Relationship to Higher Cognitive Processes. International Journal of Psychophysiology 8, 35–42 (1989)
7. Haapalainen, E., Seungjun, K., Forlizzi, J.F., Dey, A.K.: Psycho-Physiological Measures for Assessing Cognitive Load. In: International Conference on Ubiquitous Computing. ACM Press (2010)
8. Holland, M.K., Tarlow, G.: Blinking and Mental Load. Psychological Reports 31, 119–127 (1972)
9. Horlings, R., Datcu, D., Rothkrantz, L.J.: Emotion Recognition Using Brain Activity. In: International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing, pp. II.1.1–II.1.6. ACM (2008)
10. Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., Leino, T.: Effect of Cognitive Load on Speech Prosody in Aviation: Evidence from Military Simulator Flights. Applied Ergonomics 42, 348–357 (2011)
11. Ikehara, C.S., Crosby, M.E.: Assessing Cognitive Load with Physiological Sensors. In: Hawaii International Conference on System Sciences, p. 295a (2005)
12. Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., Zarjam, P.: Measuring cognitive workload with low-cost electroencephalograph. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 568–571. Springer, Heidelberg (2011)

13. Lipp, O.V., Neumann, D.L.: Attentional Blink Reflex Modulation in a Continuous Performance Task is Modality Specific. Psychophysiology 41, 417–425 (2004)
14. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces. Journal of Neural Engineering 4, R1–R13 (2007)
15. Nakasone, A., Prendinger, H., Ishizuka, M.: Emotion Recognition from Electromyography and Skin Conductance. In: International Workshop on Biosignal Interpretation, pp. 219–222 (2005)
16. Niculescu, A., Cao, Y., Nijholt, A.: Stress and Cognitive Load in Multimodal Conversational Interactions. In: International Conference on Human-Computer Interaction, pp. 891–895. Springer (2009)
17. Nourbakhsh, N., Wang, Y., Chen, F., Calvo, R.A.: Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. In: Australian Computer-Human Interaction Conference, pp. 420–423. ACM Press (2012)
18. Paas, F.G.W.C.: Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. Journal of Educational Psychology 84, 429–434 (1992)
19. Paas, F.G.W.C., Van Merriënboer, J.G.: Instructional Control of Cognitive Load in the Training of Complex Cognitive Tasks. Educational Psychology Review 6, 351–371 (1994)
20. Paas, F.G.W.C., Van Merriënboer, J.G.: Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. Journal of Educational Psychology 86, 122–133 (1994)
21. Ruiz, N., Taib, R., Shi, Y., Choi, E.H., Chen, F.: Using Pen Input Features as Indices of Cognitive Load. In: International Conference on Multimodal Interfaces, pp. 315–318. ACM Press (2007)
22. Sakr, G.E., Elhajj, I.H., Wejinya, U.C.: Multi Level SVM for Subject Independent Agitation Detection. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 538–543 (2009)
23. Sebe, N., Lew, M.S., Cohen, I., Garg, A., Huang, T.S.: Emotion Recognition Using a Cauchy Naive Bayes Classifier. In: 2002 Proceedings of the 16th International Conference on Pattern Recognition, vol. 1, pp. 17–20. IEEE (2002)
24. Setz, C., Arnrich, B., Schumm, J., Marca, R.L., Tr, G., Ehlert, U.: Discriminating Stress From Cognitive Load Using a Wearable EDA Device. Technology 14, 410–417 (2010)
25. Shi, Y., Ruiz, N., Taib, R., Choi, E.H., Chen, F.: Galvanic Skin Response (GSR) as an Index of Cognitive Load. In: Computer/Human Interaction Conference Work-in-Progress, pp. 2651–2656. ACM Press (2007)
26. Wang, W., Li, Z., Wang, Y., Chen, F.: Indexing Cognitive Workload Based on Pupillary Response under Luminance and Emotional Changes. In: International Conference on Intelligent User Interfaces, pp. 247–256. ACM Press (2013)
27. Xu, J., Wang, Y., Chen, F., Choi, E.: Pupillary Response Based Cognitive Workload Measurement under Luminance Changes. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part II. LNCS, vol. 6947, pp. 178–185. Springer, Heidelberg (2011)
28. Yeo, M., Li, X., Shen, K., Wildersmith, E.: Can SVM Be Used for Automatic EEG Detection of Drowsiness During Car Driving? Safety Science 47, 115–124 (2009)

Information Holodeck: Thinking in Technology Ecologies

Sharon Lynn Chu and Francis Quek

2202 KnowledgeWorks II, Kraft Drive
{sharilyn,quek}@vt.edu

Abstract. Information can be persistently represented on a multitude of devices beyond a single screen and session. This paper explores how technological *display and device ecosystems* (*DDEs*) may support human thinking, learning and sensemaking. We propose a theoretical foundation that extends Vygotsky's sign mediation triangle to include digital information. Through a process we call *objectification*, perceivable objects, e.g. interface objects, tangible technologies, can be associated with signs to support thinking. We present a qualitative study of learning in a testbed *DDE* with 12 graduate students. We developed a method that traces digital objects within 'thinking episodes' to help us evaluate how technology configurations support *objectification.* Our findings relate two storylines of how *DDE* technologies may afford objectification. Our work advances a method informed by psychological theory to examine device ecologies for their potential for learning, and articulates affordances for the design of technology that can help to support higher thought.

Keywords: Ecology, technology, devices, displays, thinking, sensemaking, objectification, embodied interaction, design.

1 Introduction

In the TV Sci-Fi series, Star Trek Voyager, there is an Emergency Medical Hologram Mk 1 (EMH) doctor played by Robert Picardo. Although the EMH was designed to simulate a personality, at the beginning of the series he/it was just a program restricted to the Sick Bay (which had 'Holo-emitters' to support his materialization), and could take physical form for periods no longer than necessary to see an emergency patient. But because Voyager was stranded far from home without medical support, the technical staff 'stabilized his matrix', so he could remain persistently corporeal. Over time, they extended the technology coverage to enable him to materialize in larger portions of the spaceship, and acquired a 'mobile holo-emitter' that enabled him to be material continuously. He became 'real'.

This technology tale is relevant to our deployment and use of technology because just like the doctor, information with which we interact is at the same time digital bits and materialized representations on screens and devices. Just as an ecosystem of technological enablements support the materialization of the doctor, a multitude of devices are increasingly able to support persistent representations of information materializations

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 167–184, 2013.

beyond the restrictions of a single screen and a single session. The question is how may our information become 'real' and why this may be important for learning.

We have a good understanding of how technology augments our human physical skills (e.g., digging), and our cognitive skills (e.g., memory) to a certain extent. It has proven more difficult for us to understand how technology can augment the higher human thought. Investigations on information, communication, artifact and media ecologies, have addressed mostly philosophical, behavioral, design and technical perspectives of these new environments. This paper explores how these technological *display and device ecosystems* (*DDEs*) in the physical world may function to support human thinking, learning and sensemaking.

First, we review and synthesize the literature on technology ecologies, and then present a model to make sense of the process of thinking in *DDEs* derived from the theories of the Russian psychologist, Lev Vygotsky. We describe a study that uses our theoretical framework to investigate how students think in the *DDEs* that they form, and conclude the paper by discussing technology affordances and their implications for the design of *DDEs* for learning.

2 A Review of Technology Ecosystems

A range of research has hitherto applied the metaphor of a biological ecosystem to human activities with technology for illustrative purposes and to stimulate intellectual discussions. However, this body of work does not always form a coherent whole, and it is a challenging undertaking to present a comprehensive synthesized account. We reviewed many technology ecology notions in the literature, but in the interest of space, we shall simply list them and expand only on the most relevant ones here.

The overall message underlying the different positions in the literature is that artifacts, devices, systems and products cannot be studied in isolation but can only be truly understood when seen in the broader perspective of the universe they inhabit. Depending on the position taken, the universe can consist of one's physical context, other artifacts used, or one's practices and culture using technology. We classify the perspectives into three categories: philosophical positions, empirical study results and technical frameworks.

Among the theoretical or philosophical positions taken on technology ecology concepts, one can find the 'media ecology' by McLuhan [1], Altheide's [2] 'ecology of communication' or 'communicative ecology', the 'information ecology' by Nardi & O'Day [3], Tungare et al.'s [4] 'personal information ecosystem', Krippendorf's [5] 'ecology of artifacts', and Rick's [6] proposition of a 'classroom ecology of devices'. Gibson [7], Suchman [8] and Norman [9] have also used the ecological metaphor.

Research of technology ecologies that present an empirical study of some sort include Huang, Mynatt & Trimble's [10] 'display ecology', Enquist, Tollmar & Corry's [11] 'interaction ecology', Dearman & Pierce's [12] 'computing with multiple devices', Forlizzi's [13] 'product ecology', Jung et al.'s [14] 'personal ecology of interactive artifacts', and Coughlan et al.'s [15] 'device ecology'. From the management sciences, Bailey & Barley [16] present an extensive 'shadowing' study of 'teaching-learning ecologies', tracking knowledge as it moves through six engineering firms. Their focus however was on people and not technology.

Some of the technical frameworks that have been proposed to implement technological ecologies are Loke & Ling's [17] use of petri nets to represent the state of devices, the 'task migration framework' by Pyla et al. [18], and Pierce & Nichols' [19] framework based on instant messaging to enable multi-device user experiences.

Among the different conceptions of ecologies that we reviewed, only Coughlan et al. [15] and Jung et al. [14] presented formal empirical studies about technology use in a learning context. Rick [6] points out the importance of a classroom ecology, but does not provide any supporting study. Coughlan et al.'s investigation informs the design of ecologies by studying transitions in foci across devices (a tabletop computer with a mirrored projection, laptops, a telephone) in three short controlled activities, carried out in a "technologically-enhanced indoor space". Communication across devices was provided by a Central Management System, and instant messaging. The focus of their study was on how device ecologies can support collaboration. Study results presented a set of "seams" that represent disconnects in a device ecology that can affect users' behaviors. Their study however gave little indication of how one can understand whether or how learning has occurred within the context they constructed.

Jung et al. [14] studied one's network of personal artifacts through the lens of 'factors' and 'layers' within a 'personal ecology of interactive artifacts', described as a "set of all physical artifacts with some level of interactivity enabled by digital technology that a person owns, has access to, and uses". They make use of two methods called the *Personal Inventory*, based on a simplified version of the Repertory Grid Technique, and the *Ecology Map*, which consists of sketching using sticky notes to probe about a person's device ecology. Their exploratory study with ten graduate students found that perceived attributes of an artifact can be classified into two categories, designed properties (physical, functional, informational, interactive aspects) and subjective values (experiential, emotional, social). They further specify the different types of relations that artifacts in a personal ecology can have, based on: purpose of use, context of use, or subjective meaning. Their study results, although very helpful to understand the nature and types of technological ecologies, again do not consider the process of learning.

We conceive of a technology ecology where devices function, not as individual gadgets, but in ecosystems to deliver an experience. As such, we define a *DDE* as *a mesh of interacting displays and devices that enable the manifestation and manipulation of digital information to deliver a cohesive learning experience* [20].

3 A Model of Thinking in *DDEs*

In the context of our definition of a *DDE* above, we present a high-level model of how thinking can be understood with regards to interactions with digital information through physical technological devices. Figure 1 illustrates our model. People's thoughts (what we label as *Thought objects,* TOs), can be encoded into *information*, that we describe as *digital objects* (DOs). Through technology in the *DDE,* DOs can be *externalized* as *Manifest Objects* (MOs) that are perceptible to the human senses. Examples of MOs can be displayed documents, images, or file icons. These MOs, however, also have the capacity to mediate further *thinking* – a process we call *objectification*. Externalization of TOs as MOs allows the user's perceptual and spatial abilities to participate in the thinking process. This idea is related to that of 'distributed cognition' [21], which states that human knowledge and cognitive processes are offloaded into the environment as external representations. Our contribution is that we suggest the mechanism by which cognition is distributed into the environment through the *DDE*. In the next section, we make use of Vygotsky's sign mediation theory to describe a mechanism by which the different processes in our high-level model take place. Other perhaps more commonly known theories that have been derived from or are closely associated with Vygotsky's theory include, for instance, activity theory and distributed cognition.

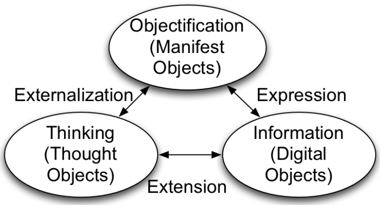


Fig. 1. High-Level Model of Thinking in *DDEs*

The Sign Mediation Triangle and Materialized Thought

We turn to Vygotsky's theories to understand thinking in *DDEs* because he proposes a way by which things in the environment may be brought into the very process of thinking. According to sign mediation theory, language is conceived of as a psychological tool by which both cultural (interpersonal) and psychological (intrapersonal) thought are '*mediated*' [22, 23]. *Signs* are self-generated linguistic stimuli [22] that extend the operation of human cognition beyond the confines of the strictly biological system. Figure 2 illustrates Vygotsky's *sign* mediation theory. For example, a student in algebra may be introduced to the summation concept, the TO: $a + b + c$ .... She understands and is able to perform the operation. However, if she had to think of details of the concept each time she applies it, the limits of her memory, attention, and mental processing would make further advancement untenable. Thus, she encodes this concept as a mental '*sign*' – the concept of '*summation*'. She is able to think of the operation simply as $\sum$, and to employ this in further learning (e.g. $\sum_{n=1}^{6}(n^2-1)$). As the *sign* becomes 'internalized' it becomes in essence the object in her thinking. She can 'unpack' the *sign* as needed to attend to the details.

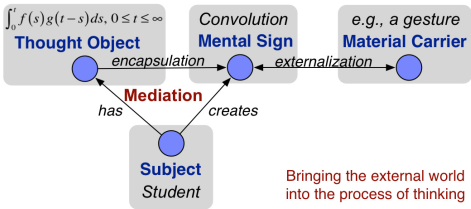


Fig. 2. Vygotsky's sign mediation theory

In Vygotsky's model, *signs* may take the form of both internal or external symbols (a stick between a child's legs becomes his horse, and a block represents an idea [22]), or as an abstract entity grounded in language. Externally instantiated signs are referred to as '*material carriers*' (MC) of thought [24] (see Figure 2). Any perceivable *object* (spatial location, gestures, objects or even sounds) in the environment can opportunistically and temporally be appropriated for use as MCs to assist thinking by bringing spatial ability and perception into play. In theory, the MC can be anything that may or may not resemble the mental object. In our example, the MC for convolution can be a specific hand gesture or a written * symbol.

In the early 1900s, Vygotsky obviously never encountered the magic of computation and modern display technology. The number of MCs one can entertain at any one time is limited by the meanings one can assign and recall for amorphous objects and space. We advance a theoretical framework, in Figure 3, that extends the model of the *sign* and MC, to include the 'magic' of digitality. The top of Figure 3 from *thought object* (TO) to mental *sign* and *material carrier* (MC) replicates Vygotsky's mediation theory. In the information world, a TO can be *encoded* in a *digital object (*DO). The DO can at times even *extend* the TO, such as with information about how the convolution formula was derived. The mental *sign*, as well, thus becomes *associated* with the DO. Through *DDE* technologies, the DO is *expressed* as a MO, e.g. as a website displayed in a browser on a mobile device. If one mentally appropriates this MO in the process of thinking, in essence shedding 'intentional regard' [25] to the object, the MO becomes synonymous with the MC, i.e. the binding process of *objectification.*

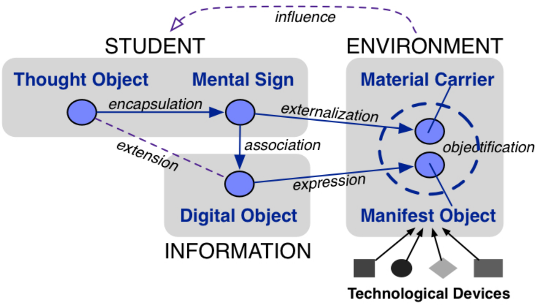


Fig. 3. Extension to the sign triangle

Our model expands the power of MOs to support thinking through MCs in two ways. First, MOs are iconic and provide mnemonic reference in ways that arbitrary objects and space cannot. This potentially expands the number of MCs one can employ over longer periods of time. Second, digital media that are encodings of one's *thought objects* (and hence associated with one's mediating *signs*) can serve as an external long-term detailed representation of a piece of knowledge, thus extending the depth of thinking one can handle. Our model describes a thinking process that is different from simply opening a document on a screen to refresh one's memory about an idea because then the document simply becomes something one queries for information rather than wields in the process of thinking. With multiple users, the MO, being the perceivable component of the model, becomes (in co-located situations) or may be made (in remote scenarios) the shared object among all users (Figure 4). We envision users taking advantage of the

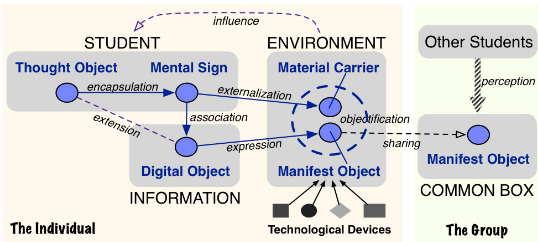


Fig. 4. Thinking in groups

broader interactions made possible by a *DDE* to place, organize, and possibly step back and view networks of MCs.

4 The Study

We conducted a qualitative study of the manifestation of the thinking process in *DDEs.* Adopting the hermeneutic perspective, the inquiry was to seek "understanding rather than explanation" and to uncover systems of meaning from the participant's vantage point [26]. We follow the methodology specified by Patterson & Williams [26] who advance that a hermeneutic approach begins with establishing a point-of-view (our 'objectification framework'), from which an 'organizing system' is derived to meaningfully organize, interpret and present data [27].

Study Methodology

The study enabled a set of students to experience a *DDE* for the purpose of doing a knowledge discovery assignment over a two-month period, and capturing their experience through self-reports and interviews. The participants were 12 computer science students in a graduate class. The assignment, to be completed in teams of three, was to research and write a report about the emergent field of 'Physical Computing'.

Fig. 5. Suite of devices available to study participants

The students were each given a 'testbed suite' of devices comprising an *iPod Touch*, an *iPad*, and a 27" *iMac* to use as their own throughout a semester. The *iMac,* embedded in a custom casing that allowed it to be laid on the table horizontally or at an angle, was endowed with a touch overlay to enable touch interaction (see Figure 5). The rationale for the dissemination of devices of various form factors was that we wanted to provide the students with an experience of a heterogeneous *DDE.* Additionally, to provide a basic information architecture that crosses devices in the ecology, we installed the free file sharing service *Dropbox*, the notetaking service, *Evernote,* with a paid subscription, and the paid PDF reader *GoodReader*[1] on the *iPad.* The first two services are based on the cloud, and *GoodReader* can be paired with *Dropbox* to allow data synchronization and transfer. We provided the students with a tutorial session on how to use the three services/applications prior to the study.

The class was held in our research center, which contains several large display screens spread out in different meeting rooms, and a large interactive vision-based

[1] www.dropbox.com, www.evernote.com http://www.goodreader.net/goodreader.html

touch screen prototype [28]. All students had constant access to the building. They were asked to use technology as much as possible while doing the assignment, including the device suite given to them, the large display screens around the center, and any other devices that they own such as laptops and other desktops. They could also use the devices freely for any other purposes.

Data Collection

We conducted semi-structured interviews and surveys to capture the students' experiences with technology before they were given our suite of devices (pre-experience), and after they have used them for the semester (post-experience). Each interview lasted for about an hour and was audio recorded. The purpose of doing a pre- and post-interview was not to perform an experimental comparison, but rather, to be cognizant of the initial conditions under which our participants joined the study.

The pre-experience interviews and surveys collected data about the students' behaviors and 'strategies of use' of the devices, as well as their attitudes toward and perceptions of the devices and processes. More specifically, the survey, which was completed in the presence of the interviewer, asked about the list of devices used and owned, the duration and frequency of use, familiarity with devices, purpose of use, perceived usability of devices (measured on a 7-point likert scale by adapting the IBM usability questionnaire [29] with dimensions like ease-of-use, comfort, efficiency, satisfaction), and data sharing methods. Qualitative comments were also encouraged during the completion of the survey. The interview addressed similar themes as the survey, and added questions about the role of devices in their idea generation and paper writing processes, and device interactions in their existing ecologies.

The post-experience interviews asked about the process of assignment completion; impact of the devices on practice; problems with devices, information sharing, and the writing process; context/situations of devices use; influencing factors of information use; personal sensemaking; work distribution and team coordination; use of file/data sharing services; longer term device use; general assessment of the ecology; desired changes; meeting contexts; idea generation process; and, information presentation.

Data Analysis

All pre- and post-interviews were fully transcribed with timestamps. Survey data were imported into a spreadsheet for analysis. To separate relevant from irrelevant verbiage in the transcripts, two coders did a first round of analysis of all the transcripts to identify 'thinking episodes'. Following Barker's ([30] in [16]) concept of "behavioral episodes", we define a thinking episode as a "coherent run of [thinking] in which the constituent [processes] have a constant direction, a purpose". Some examples include 'filtering out important keywords', 'searching for information', and 'annotating papers'. We acknowledge that thinking is a process with no explicit end point or necessary resulting outcome. However, the method of tracking thinking episodes allowed us to identify possible situations when the students engaged in thinking, and

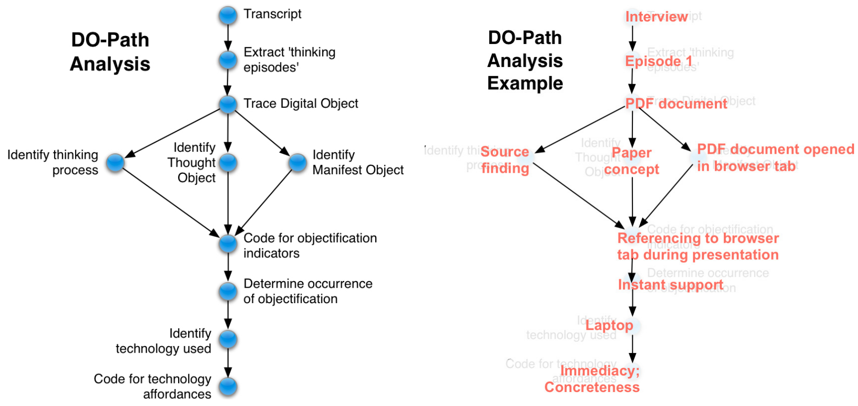


Fig. 6. Left. Analysis procedures; Right. Example analysis

was thus presented with opportunities for learning. It is not our interest in this paper to evaluate the uptake of knowledge or retention per se. Moreover, in this paper, we understand the commonly used term of sensemaking as being a necessary part of thinking.

In order to make sense of our data on the students' experiences of their ecologies, we engaged in what we call a 'DO (Digital Object)-path analysis'. The steps involved in our analysis, illustrated in Figure 6 (Left) together with an example analysis (Right), were as follows: each thinking episode was analyzed to distinguish the different components of TOs, DOs, MOs, and technologies used. We noticed that certain objects had a high frequency of occurrence in our collated list. **To focus our analysis thus, we decided to trace the use of three of the common DOs: *GoogleDocs* documents, *Evernote* notes, and PDF documents.** We shall call these our three focus objects for our study. All thinking episodes where any of the three DOs were manifested as MOs were copied to a separate spreadsheet for analysis. We then analyzed the different technology configurations in which the DO in question was used in each of the relevant thinking episodes.

By tracing the model components in context of use, we determined whether there were indications that *objectification* had occurred in any form for the particular task at hand. Two separate coders identified *objectification* indications by making a judgment as to whether the MO (e.g. *Evernote* note displayed on a tablet screen) brought spatial ability and perception into the process of thinking. Subsequently, a feature extraction was done, whereby we identified the characteristics of the technologies, or technology configurations, that seemed to have supported or hindered the *objectification* process. The characteristics were further categorized along uncovered themes.

5 Findings

To frame our findings, we will first describe the starting context of our participants from results of the pre-interviews/surveys, with an emphasis on processes of thinking with devices. This also conveys an understanding of the work strategies of the participants in the testbed *DDE* of the study, since practices, as we anticipated, did not change significantly over the two months. We then describe case scenarios from our DO-path analysis, before expounding on derived technology affordances for *DDEs*.

Initial Experience of *DDEs*

It was evident that some of the students already had a rather extensive ecology of devices prior to the study. The most common devices that they possessed were laptops, desktops and smartphones. Fewer owned tablets, music players and large displays (Figure 7). Among those who owned them, laptops and tablets were used daily. Desktops were mostly used only two to three times per week. All, apart from two, indicated that they used their smartphones everyday. The use of music players and large displays was more sporadic.

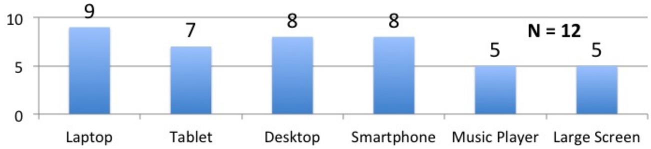


Fig. 7. Distribution of devices owned at start of study

Laptops and smartphones were mostly ranked first in familiarity, followed by tablets and music players. Students were generally least familiar with desktops and large displays. Their use of the devices can be classified into three types: *instantaneous* use, e.g. checking email, calendar, news, social networks (tablet, smartphone); *extended* use, e.g. reading papers, video streaming (laptop, tablet, large display, music player); and *long-term* use, e.g. backup of work, managing Internet (desktop, music player).

All except one had experience with *Dropbox* prior to the study, with an average duration of use of a year. Only three of the students had used *Evernote*, with an average of two months of use. All of the others however made use of other notetaking applications such as *GoogleDocs* and *TextMate*. Among those who already had a tablet, only one of them made use of *GoodReader*. The other students used other PDF readers/annotators such as *Adobe Reader* and *QuickOffice*. In terms of usability, the laptop (M = 2.57) was ranked the highest, followed by the tablet (M = 2.41), the desktop (M = 2.39), the smartphone (M = 1.92), the music player (M = 1.88), and finally the large display (M = 1.85).

In the existing *DDEs* of the students, the thinking process for assignments mostly followed the standard loops of foraging and sensemaking [31], with the prominent use of the laptop and desktop throughout. Actions in which a thinking process was evident on these devices included for instance, "just open[ing] up Microsoft Word and start typing in ideas", taking notes in a text file on *Dropbox* "when I'm surfing and I find something interesting", "categorizing my papers through…folders", and "start[ing] to kind of do a treelike structure from the cited references".

Those who had a tablet reported making use of it only occasionally for purposes of work. We could not identify many thinking episodes in their accounts of use of tablets. Among the few that we found such as for annotating readings, reading information in the browser or notetaking in the native 'Notes' application, the annotations were only "like an intermediary step to a final annotation", the website was "just [to take] a look at", and the notes (e.g. "on page X second paragraph is interesting") would remain on the tablet.

The use of non-digital materials, such as sticky notes, notebooks and pieces of paper, was evident mostly in the 'foraging' part of the process. Paper materials appeared to have been used only as quick, temporary MCs and rarely had their content transferred to digital in the same exact form. One participant, for instance, recounted that she would write down 'idea fragments' on paper, and then "when I finally get my idea, I would put it on TextEdit", after which she would throw away the paper pieces. Or for reading papers, another participant "tend to print them out", "take some notes on it. And then turn it into some thrash". A number of the participants also reported relying only on memorization, "I remember some keywords of it, so I can get it when I want to", without the use of any devices.

Objectification as Manifested in Testbed *DDEs*

Recall that our analysis traces three kinds of 'focus objects' (PDFs, Evernote notes, & GoogleDocs) used by our participants as DOs. Our DO-path analysis of the post-interviews brought to light different ways in which the *objectification* process occurred in the *DDEs* that the students experienced. We follow one of our three 'focus' objects as it was involved in two example storylines extracted from thinking episodes.

Storyline 1: *Identifying relevant sources for the assignment.* One team employed a strategy by which they brainstormed a set of keywords and organized these on a whiteboard as a concept map, after attempting to use the vision-based interactive touch screen and failing. From this, they selected a subset of keywords to explore, and used these for literature searches. One of our participants related the following process, which we map to our model:

1. He searched for sources using the keywords and left these in tabs in his laptop browser. These were not saved to disk nor named. The browser tabs served as a 'bag of finds' whose relevance was yet to be determined ("I didn't save them on my computer. They were opened in my browser actually, because I wanted to filter what I have and see if I need to *take these* or not.").

2. He brought his bag of finds on his laptop to the next meeting and presented them to the team, and selected some to *take* ("I just renamed the paper to keep track of *which is which*. If I saw robotics … so this is *the robotics paper*. If there is a lot of robotics papers, I will say *robotics 1, 2, 3,* and or I would tie it to the name of the author."). It is at this point that the participant associates particular PDFs with *signs*. Relating this to Figure 3, the 'concept of robotics paper as relevant to physical computing' was the *thought object*, the *sign* was the name of the file, and the *manifest object* was either the displayed PDF document or the named file icon in the folder.

3. He would copy the PDFs to his *iPad* for further reading, and refined the naming of the PDF according to his understanding of the paper ("So after you read the paper … OK … this paper is about *this.* A way I used to do is to just rename the paper to keep track of *which is which.*").

In this storyline, we observe the formation of *signs* and the use of *manifest objects* as MCs to help the participant make sense of how each paper (representing a specific concept) fits in ‘Physical Computing’. This strategy was seen in two other participants. One had put her ‘bag of papers’ on her *iPad* and had them open in *GoodReader* tabs to share and discuss with her teammates (“… papers that were open on it … since we were talking about what we researched. I was able to say *I found* ***this*** *paper, it talks about this and this*”). She was using the *iPad* on which the PDF was materialized as an MO to support her discussion and thinking (the team did not actually read the paper on the *iPad* screen, the participant merely used it as an objectified *sign* to reference the information the PDF contained). A third participant employed an almost identical strategy with a slight difference in that he did an initial filtering of his ‘bags of finds’ on his laptop browser tabs, and winnowed these to 15 - 20 papers that he saved and named in his *Dropbox*. He copied these to his *iPad* for further reading and annotation. There was also evidence of his employment of filenames as meaningful *signs* to give him mnemonic access to the PDFs, and his use of the displayed PDFs as MCs in the process of reasoning about ‘Physical Computing’.

Storyline 2: *Creating the assignment report*. One of the participants organized his paper building from an outline written in a *GoogleDocs* document. This outline served as the conceptual frame on which he worked: “I have the outline and each point I know which PDF is linked to what, so I would read the PDF and while I am reading, I get an idea, and I start writing on the paper.” He read the paper on his *iPad* and wrote in *GoogleDocs* on his laptop simultaneously. He mentally associated each outline item in his *GoogleDocs* outline with several PDFs.

Here, we see that he has a two-level conceptualization of the PDF paper he is reading. The idea of an outline item is a *thought object.* Both the mnemonic name of the PDF and the concomitant name of the outline point serve as *signs* for the TO. The TO has two different simultaneous MOs. The outline item in the opened *GoogleDocs* document is a MO with which he associates the TO, but it is also clear that he “knows which PDF [it] is linked” to. A second MO is thus the display of the PDF on his *iPad*. Further, since the PDF summoned to the *iPad* is open to reading and inquiry, the contents of the PDF themselves become another level of TOs, *signs*, and MOs. At each level, *objectification* may occur to allow the use of the MO as an MC for reasoning. It is precisely this nature of *signs* that are able to hold entire concepts for thinking, and be unpacked into its contents for inquiry at a finer degree of abstraction that makes the theory a potent vehicle to understand thinking in *DDEs*.

This same two-level strategy was employed by another participant who used PDFs in *Dropbox* folders to support conceptualization. The mnemonic names of the PDFs served as the *signs* and the preview display of the PDFs (quick look feature in the *Macbook*) in the *Dropbox* folder served as the MO or MC when he referred to it while writing. The participant wished he could add more information to the *sign* at the level of the whole PDF, stating “Originally we discussed annotating them (the PDFs), but we found that it was kind of difficult to actually do it because we just wanted to have a summary of each paper but Dropbox on the iPad would not let us create a text file [of the same name] next to the papers.” This shows that the participant was thinking at the level of the PDF as a whole, and wanted to associate meta-information at the level of PDF within the folder structure.

Another three participants exhibited the same two-level conceptual structure using yet other methods to provide overarching structure of how the *sign* associated with each PDF relates to the paper as a whole. Two of them employed a citation manager, e.g. *Endnote*, *Mendeley*, to maintain citations and to add an annotation at the level of the whole citation for each PDF. An entire PDF thus could be used as a contained unit of thinking for paper writings. Our last participant recorded to use this strategy employed *Evernote* for her higher-level organization. She had an *Evernote* note for each PDF document, describing its contents at the level of its place in the whole 'Physical Computing' paper. *Evernote* allowed her to attach the PDF directly to the note, enabling the same two-level sign-MC strategy.

Technology Affordances of *DDEs*

To recap, our theoretical framework used for analysis is that *digital objects* may be materialized through DDEs into *manifest objects*, and that these MOs may be appropriated through a process of *objectification* to serve as *material carriers* that support thinking. We extracted 50 thinking episodes that related to *GoogleDocs* documents, *Evernote* notes, and PDF documents. Our DO-path analysis traced the three 'focus' MOs in terms of thinking processes, TOs, DOs, MOs, technologies used, relevant quotes relating to affordances and experience, and the occurrence of an MO-MC binding (coded as hindrance or support).

The second part of the DO-path analysis, of which a sample is shown in Figure 8, coded the quotes extracted in the first part of the analysis for *objectification* indicators and 'support' (in the moment or/and in the long-term) or 'lacking' characteristics of the technology configurations used. Indicators included for example referencing materials as thoughts, reports of pointing to objects as ideas, and use of space as an organizing structure. Figure 9 shows the distribution of thinking episodes in which the MO was able to be coded as either having a support or hindrance role in relation to the students' thinking processes. The synthesis of the technology affordances that we found supported, were lacking, or

Thinking Process	Technology Configuration	Objectification Indications	Support Characteristics	Lacking Characteristics
Concept formation/ Outlining	• A laptop with EVN hooked to a large display	• Reference to points listed in Evernote	• Common ground • Simultaneous focus	• None
Brainstorming/ Concept formation	• Three co-located personal laptops, each with EVN	• None	• None	• Simultaneous focus • Common ground
Concept formation	• iMac with EVN • One personal laptop	• Gesturing and pointing to picture in Evernote	• Spatiality • Accessibility • Immediacy	• None
Presenting ideas	• Tablet with EVN	• Reference to points in Evernote while presenting	• Transparency • Immediacy • Persistence	• None
Sharing information	• Tablet with EVN	• None	• None	• Transparency • Consistency • Expectation of interaction

Fig. 9. Objectification indicators and Technology affordances from DO-path analysis

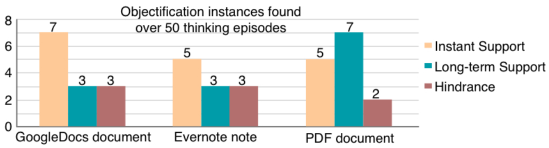


Fig. 8. Role of *Manifest Objects* in thinking

hindered *objectification* to occur is presented in Figure 10. The features were derived from quote-supported storylines like Storyline 1 and 2 above.

Storyline 1, for example, shows evidence of the creation of mental *signs* as part of the reasoning workflow, and how *objectification* may have occurred by which MOs, materialized from the PDFs, served as MCs to support thinking. The *persistence* of the ‘bags of finds’ in the laptop browser allowed the participants to think of these finds as ‘stuff I found’ without necessarily forcing them to a premature commitment to meaning or interpretation. The mobility of the *iPads* supported *opportunistic action* so that the bags of finds and the named PDFs could be used to support thinking. The *transparency* of action between searching, renaming, and reading allowed for “a complete experience”.

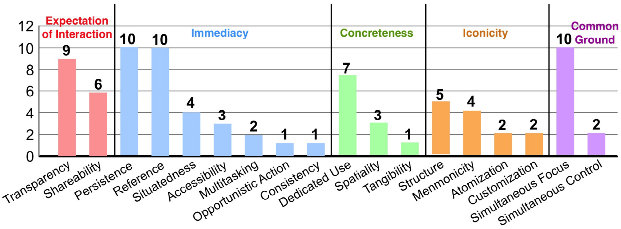


Fig. 10. Technology affordances for *objectification*

In Storyline 2, all the participants regarded the entire PDF as the DO related to a particular TO at the level of the organization of the entire paper being written. The participants all employed a strategy of encapsulating this level of conceptualization of the paper with a very short description that served as the MO. This allowed the participant to bring perception into the process of thinking by using the MO to serve as an MC. The extent at which this *objectification* process occurred however was highly dependent on the *transparency* of interaction to annotate/associate meta-data to the high-level description of the TO, and the *persistence* of the display of the MO. The *immediacy* of being able to quickly drill down to the content of the PDF document as well supported the unwrapping of successive layers of abstraction and encapsulation.

6 Discussion

We organize the discussion of our work under three themes: the *objectification* framework as a method for the study of technology ecologies; the experiences of students in *DDE*s; and technology affordances to aid in the design of *DDEs.*

An Objectification Framework for Thinking in DDEs

Digital information is ‘magical’. We can duplicate, copy, summon to visual presentation, search, etc. almost at will. Yet, despite this malleability of pixels on screen, we often revert to physical, material things to help us make sense of information because they more readily help us to objectify thought. Starting from Vygotsky’s sign mediation theory, we proposed that *thought objects* can be related to *digital objects* that can be *materialized* in DDEs as *manifest objects*. Through *objectification,* the student appropriates the MO to serve as a *material carrier*, thereby bringing the MO into the loop of thinking. Our DO-path analysis, as presented in the two example storylines, showed the promise of our *objectification* framework, and affordances in the *DDE* that appears to support *objectification*. We successfully traced the components of our model in different situations in our data transcripts to elucidate the thinking process.

Other similar 'path analysis' methods have been proposed before for the study of technology use, notably in works that argue for the importance of context in HCI. For example, in 'activity theory' (AT), the unit of analysis is human activity, mediated by the use of artifacts or tools. One traces the subject, object, actions and operations in an analysis approach using AT [32, 33]. In a similar vein, situated action models identify the "activity of persons-acting in setting" as the focus of study. Using the approach, one traces the relations between individual and environment at a "very fine-grained level" as the human engages in dynamic uses of artifacts [32]. Distributed cognition (Dcog) looks at the "cognition of a system in terms of its function". It requires one to trace "representation [states] inside and outside the head" as the central unit of analysis [34]. Our method of tracing *DOs* does not exclude artifacts (devices and displays), individuals (students), or their activities (thinking processes). Adopting a different focus, it provides a way instead to make sense of the flow of information, literally 'food for thought', from the environment to the mind of the individual (or vice-versa).

Experiencing the *DDEs*

With currently available technologies, students already build ecologies with their personal devices in daily life. As highlighted by Jung et al. [14], "every newly designed interactive artifact will inevitably become a part of someone's ecology". The question is however whether it will be an *effective* part of his/her ecology. We found many situations when the student was faced with 'gaps' in his/her *DDE*, which hampered the smooth flow of his/her thinking. He/she then had to find other technologies that could bridge the 'gaps'. In this sense, the *DDEs* that one constructs today are 'ad-hoc ecologies': the different components do not function as a synergistic whole by design, but places the burden of synchronization and file control on the user [4].

Furthermore, we found that the use of the individual devices was mostly characterized by what we call the 'portal thinking' effect. We posit that many current interfaces and devices are designed based on the Cartesian view that external things and thinking processes are mutually exclusive. The result may be that the screen then acts only as a conduit through which information is summoned. With computational advancements (e.g. larger screens, mobile systems) we may have multiple 'portals', but they remain data straws to draw information to view. It is unfortunate thus that the potential of interface, visual and physical components of devices to be MCs for thinking often fails to be harnessed, and instead, technology is used as mere input, output or 'projection' devices.

Designing DDEs for a Learning Experience

We believe that there is a lack of designed support to optimize technology's role of augmenting humans' higher thought. From our findings, we proposed technology affordances that appears to help the process of *objectification*. These can be classified into seven groups, although they are all interrelated at certain levels. Below, we discuss the central theme of each group, while providing considerations for system design features that can potentially embody each theme. It must be noted however that our proposition of technology affordances in *DDEs* for learning does not imply that one can design such a complete ecosystem for learning. Variations in individual learning styles and the diversity of subject matter to be learned would preclude such rigidly constructed ecosystems. Drawing insight from the inveterate paper ecology that has supported learning since Gutenberg's printing press, we suggest that learning ecologies, be they physical or digital, have to remain ad-hoc but with designed support. The paper ecology, that includes writing implements, tables, bookshelves and libraries, books, and the myriad paper clips, staplers, rulers, and folders, has evolved over more than half a millennium and provides a vast set of components that are flexible enough for the individual to appropriate them to construct his/her environment for specific learning experiences. The digital ecology as well, needs to evolve technologies that possess affordances but yet are malleable enough to support learning.

Affordance 1: ***Expectation of interaction (includes transparency, shareability):*** It is key that components of *DDEs* are able to not only interoperate in some way, but also provide an expectation of interaction to users. Our writing implements for example can be expected to write with few impediments on paper. Of course, the paper ecology is constrained by materials and physical laws (e.g., we do not have to worry about book 2.0 falling through the surface of table 3.1 because of incompatibilities) while all interactions in *DDEs* have to be designed and implemented. Also, the cultural longevity of paper has built expectations and constraints (e.g., pencils do not work on leather portfolios) into the user community that digital technologies cannot always rely on. In our study for instance, the students reported that they decided to use the whiteboard for brainstorming particularly because they knew that they would be able to take a picture of it with their *iPad* cameras later on. Conversely, a clear example of failure of this aspect in our study is one instance where the students spent one entire meeting only to set up shared *Dropbox* folders and *Evernote* notebooks. Work in middleware and system interoperability, as well as direct interaction methods, are important to enable this affordance of 'expectation of interaction' to take place.

Affordance 2: ***Immediacy (includes persistence, reference, situatedness, accessibility, multitasking, opportunistic action):*** *Immediacy* concerns features that allow the user to display, manipulate, and use DOs across devices without going through one or more indirect actions. For example, if a user has a physical paper that she wants a friend to read, she drops it on the table in front of him. However, if she had the document on her laptop, she may have to put it into *Dropbox* and tell him where to get it before they can discuss its contents. The lack of *immediacy* in this scenario hinders the *opportunistic* use of the document as a focus of discussion or thought. *Immediacy* is closely associated with *transparency*. Transparent interoperability across platforms supports immediate action as do *consistency* of interaction techniques (consistent ways to move and manipulate MOs across platforms is critical to support their use as MCs). In our study, participants used *persistence* to allow information to stay immediate and more easily participate in their thinking: on the laptop, they aligned their *Word* document and the PDF papers side by side. Others used their *iPad* as a 'persistent' secondary display for the PDFs.

Affordance 3: ***Concreteness (includes dedicated use, spatiality, tangibility):*** Components that possess the characteristic of 'concreteness' in a form appears to better support *objectification*. In our study we saw that the whiteboard that affords the use of space and the *iPad* that opens applications using the full screen real estate, for example, aid information to become what Heidegger [35] calls a 'thing', something tangible for the student to grasp in her thinking process. A technological environment that requires the user to hold thoughts in mind while looking for the appropriate device to record them offers little support for objectification. Work on embodied interaction, physicality, and tangible user interfaces that address how to enable the user to easily assign thoughts to concrete materials contribute to this affordance.

Affordance 4: ***Iconicity (includes structure, mnemonicity, atomization, customization):*** Objects to which we have assigned meaning become more easily internalized than neutral objects. We have seen the use of file renaming and organization of folder structure as instances when *iconicity* has enabled MOs to be used as MCs of thought. The few works in HCI that have looked at familiarity of interfaces are potent informants of this affordance. *Atomization* is a related feature that supports the association of DOs to TOs. A TO is typically an atomic idea at some level of abstraction, in the same way that a 'unit of analysis' specifies an entity that is a coherent whole at a certain level. Take the idea of convolution in our earlier example. A Wikipedia page on convolution would be an apropos DO for the concept, but a whole book on signal processing would not (even if it contains a section on convolution). Features to support *atomization* (e.g., bookmark individual components of larger text documents) have been investigated for example in hypermedia research.

Affordance 5: ***Common ground (includes simultaneous focus and control):*** To be able to focus on thinking in groupwork, the technology should provide support for students to easily create shared MCs. A common MO may not necessarily lead to the same mental *signs* for two different people. This is essentially the question of intersubjectivity [36]. Physical things inherently allows for several users to have simultaneous focus and control. A page on the table can be seen by everyone around the table; several users can write on the whiteboard at the same time. In digital technologies however, the students always needed a separate 'situating channel' (e.g. speech, instant messaging, comment lines) to set common ground together with the 'information channel', where work is carried out. This factor also appeared to have prompted the more intensive use of *GoogleDocs,* which has a simultaneous editing capability, over *Evernote* for notetaking. Work in computer-supported collaborative would be relevant to further inform the design of this affordance into technologies.

7 Conclusion

This paper contributes significantly to the area of technology ecologies, focusing on how they may help us to think and to learn. We identified a gap in the literature of the plethora of technology ecology/ecosystem concepts previously proposed in terms of investigating learning. Second, we extend Vygotsky's sign mediation theory to the digital world by proposing how *thought objects* may relate to *digital objects*, and their manifestation through technology ecologies. Third, we propose a way through the concept of *objectification* and method of DO-path analysis to examine device ecologies for their potential for learning. And fourth, we articulate affordances that can inform either design guidelines or evaluation metrics for the development of *device and display ecologies* that allow us to think effectively.

One limitation of our study may be that it was conducted with computer science students who may have had a different perspective on technology. Researchers from other domains like history may require additional support to piece together fragmentary information. Such extension can be a rich domain for future research. Although we do not claim to propose a holistic explanatory theory of cognition, we do believe that the *objectification* framework manages to uncover basic thinking processes irrespective of the student's domain expertise. Knowledge discovery is foundational to any form of learning. In a future study, we plan to employ our model to make sense of students' in-situ speech on top of their self-reported experiences of learning using technology. In the face of changing perspectives, HCI researchers should begin investigating and designing technologies that help us think, or in our Star Trek tale, that make the *doctor* 'real'.

Acknowledgements. This project has been supported by NSF grant No. IIS 1059389.

References

1. Strate, L.: Studying Media As Media: McLuhan and the Media Ecology Approach. MediaTropes I, 127–142 (2008)
2. Altheide, D.: An Ecology of Communication: Toward a Mapping of the Effective Environment. The Sociological Quarterly 35(4), 665–683 (1994)
3. Nardi, B., O'Day, V.: Information Ecologies: Using Technology with Heart. First Monday 4(5) (1999)
4. Tungare, M., et al.: Personal Information Ecosystems and Implications for Design. Computing Research Repository (CoRR) cs/06120812006
5. Krippendorff, K.: The Semantic Turn: A New Foundation for Design, p. 368. CRC Press (2005)
6. Rick, J.: Towards a classroom ecology of devices: Interfaces for collaborative scripts. In: Wksp at CSCL 2009: "Scripted vs. Free CS Collaboration: Alternatives and Paths for Adaptable and Flexible CS Scripted Collaboration", Rhodes, Greece (2009)
7. Gibson, J.J.: The ecological approach to visual perception. Lawrence Erlbaum Associates (1979)
8. Suchman, L.A.: Human-Machine Reconfigurations: Plans and Situated Actions. Cambridge U. Press (2007)
9. Norman, D.: The Psychology of Everyday Things. Basic Books, New York (1988)
10. Huang, E.M., Mynatt, E.D., Trimble, J.P.: Displays in the Wild: Understanding the Dynamics and Evolution of a Display Ecology. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A. (eds.) PERVASIVE 2006. LNCS, vol. 3968, pp. 321–336. Springer, Heidelberg (2006)
11. Enquist, H., Tollmar, K., Corry, A.: Interaction Ecologies. In: Wksp at Pervasive Computting (2007)
12. Dearman, D., Pierce, J.: It's on my other Computer!: Computing with Multiple Devices. In: CHI 2008 (2008)

13. Forlizzi, J.: The Product Ecology: Understandng Social Product Use and Supporting Design Culture. International Journal of Design 2(1), 11–20 (2008)
14. Jung, H., et al.: Toward a Framework for Ecologies of Artifacts: How are Digital Artifacts Interconnected within a Personal LIfe? In: NordiCHI 2008, pp. 201–210.
15. Coughlan, T., et al.: The Conceptual Framing, Design and Evaluation of Device Ecologies for Collaborative Activities. IJHCS 70(10), 765–779 (2012)
16. Bailey, D., Barley, S.: Teaching-Learning Ecologies: Mapping the Environment to Structure Through Action. Organization Science 22(1), 262–285 (2011)
17. Loke, S., Ling, S.: Analyzing Observable Behaviours of Device Ecology Workflows. In: 6th International Conf. Enterprise Information Systems (2004)
18. Pyla, P.S., Tungare, M., Holman, J., Pérez-Quiñones, M.A.: Continuous User Interfaces for Seamless Task Migration. In: Jacko, J.A. (ed.) HCI International 2009, Part III. LNCS, vol. 5612, pp. 77–85. Springer, Heidelberg (2009)
19. Pierce, J., Nichols, J.: An Infrastructure for Extending Applications' User Experiences Across Multiple Personal Devices. In: UIST 2008. ACM (2008)
20. Chu Yew Yee, S., et al. The Physicality of Technological Devices in Education: Building a digital experience for learning. In: ICALT 2012 (2012)
21. Hollan, J., Hutchins, E., Kirsh, D.: Distributed Cognition: Toward a New Foundation for HCI Research. TOCHI 7(2), 174–196 (2000)
22. Vygotsky, L.S.: Mind in society: The development of higher psychological processes. Harvard U. Press (1978); Cole, M., et al. (eds.)
23. Vygotsky, L.S. (ed.): Thought and Language (Edited and translated by E. Hanfmann and G. Vakar). MIT Press, Cambridge (1987); Kozulin, A. (ed.)
24. McNeill, D.: Gesture and Thought. University of Chicago Press, Chicago (2005)
25. Dourish, P.: Intentionality. In: Where the Action Is: The Foundations of Embodied Interaction, pp. 134–138. MIT Press, Boston (2004)
26. Patterson, M., Wiliams, D.: Collecting and Analyzing Qualitative Data: Hermeneutic Principles, Methods, and Case Examples. Advances in Tourism Applications. Sagamore Publishing (2002)
27. Tesch, R.: Qualitative Research: Analysis Types and Software Tools. Falmer Press, New York (1990)
28. Verdie, Y., Fang, B., Quek, F.: Mirrortrack: tracking with reflection - comparison with top-down approach. In: ICMI-MLMI 2009, pp. 247–250 (2009)
29. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. IJHCI 7(1), 57–78 (1995)
30. Barker, R.G.: The Stream of Behavior. Appleton-Century-Crofts, New York (1963)
31. Card, S., Piroli, P.: Sensemaking Processes of Intelligence Analysts and Possible Leverage Points as Identified Through Cognitive Task Analysis. In: Int'l Conf. on Intelligence Analysis (2005)
32. Nardi, B.: Studying context: A comparison of activity theory, situated action models and distributed cognition. In: East-West Conf. on HCI, St. Petersburg, Russia (1992)
33. Kaptelinin, V., Nardi, B., Macaulay, C.: The Activity Checklist: A Tool for Representing the "Space" of Context. In: Interactions 1999, pp. 27–39 (1999)
34. Halverson, C.A.: Activity Theory and Distributed Cogntion: Or What Does CSCW Need to DO with Theories? CSCW 11, 243–267 (2002)
35. Heidegger, M.: Being and Time. Harper & Row Publishers, New York (1962); Macquarrie, J. (ed.)
36. Gillespie, A., Cornish, F.: Intersubjectivity: Toward a dialogical analysis. J. Theory of Social Behaviour 40, 19–46 (2010)

Managing Personal Information across Multiple Devices: Challenges and Opportunities

Simone Beets and Janet Wesson

Department of Computing Sciences, Nelson Mandela Metropolitan University,
P. O. Box 77000, Port Elizabeth, 6031, South Africa
{Simone.Beets2,Janet.Wesson}@nmmu.ac.za

Abstract. Due to the constantly increasing volume of personal information (PI) and the current trend towards mobile devices, there is a growing need to provide access to PI across multiple devices. It has become difficult for a user to manage his/her PI across these devices. The current hierarchical systems used to organize PI do not support accessing PI across multiple devices. The aim of this paper is to discuss the outcomes of an interview study that was conducted to determine how users currently manage PI across multiple devices and to identify what problems are experienced using these devices. Results showed that participants found it difficult to access PI across their devices and do not know beforehand what information they need to access. These problems could be solved by providing an information visualization tool installed on their devices which provides a single user interface to facilitate an overall view of their PI.

Keywords: Personal Information Management, Multiple Devices, Interview Study.

1 Introduction

Personal information management (PIM) involves the daily activities or tasks that users need to perform using a set of information items, such as documents, media and calendar events [1]. A user's personal information (PI) increases constantly as information is currently being stored on a number of different devices, platforms and applications [2]. This has led to a high level of dispersion of PI, referred to as the information fragmentation problem, and an increased difficulty in managing, accessing and using this information. It has become difficult for a user to access to his/her PI at any time regardless on which device the information is stored. The goal of this research is to provide support for accessing PI across multiple personal devices.

The current method used to organize PI is in hierarchies of files and/or folders [3]. Although the hierarchy is a familiar organization method, it suffers from several limitations due to its restrictive nature [3]. Current PIM solutions mainly focus on enhancing PIM on a single device or on a subset of PI [4]. Alternative applications, such as Dropbox (www.dropbox.com) and TeamViewer (www.teamviewer.com), provide support for accessing PI, but also have several shortcomings.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 185–192, 2013.

The aim of this paper is to describe the results of an interview study that was conducted to determine how participants currently manage PI across different devices and identify the problems that participants experience in managing PI across these devices. The results of this study will inform the design of a tool to support accessing PI across multiple devices using information visualization (IV) techniques.

Section 2 discusses related work regarding PIM. Section 3 introduces the interview study and describes the themes identified from the interview data. Design implications are identified in Section 4 based on the results of the interview study. The paper concludes by discussing the main contribution of this research and future work to be completed.

2 Related Work

PIM is a large, active area of research [1], [5]. The goal of PIM is to enable a user to access his/her PI relevant to his/her location, in the most appropriate form, while supporting the tasks of PIM [1]. PIM tasks were originally identified in [6] to include acquiring, organizing, storing, maintaining, retrieving and producing PI. PIM tasks were then simplified to include keeping (storing), managing (organizing and maintaining) and finding (searching and browsing for retrieval) [1]. Lower-level PI tasks include creating, sorting, moving, naming, assigning properties, copying, distributing, deleting and transforming PI [1]. The keeping (storing), organizing and finding (viewing and retrieving) PIM tasks across multiple devices form the focus of this paper.

The type of PI organization has a large impact on how the PI is viewed [3]. Jetter *et al.* [7] suggested that a user interface (UI) needs to be designed to support accessing PI to assist a user to develop his/her own processes, structures and views. Gomes *et al.* [8] stated that a meaningful IV technique may be the solution for the difficulty in finding relevant PI.

3 Interview Study

3.1 Interview Method

Face-to-face, one-on-one, in-person, informal, semi-structured interviews were conducted with selected participants [9]. Thematic analysis, in combination with coding techniques, was used to analyze the interview data. The interview study was used to establish how PI is currently being managed across multiple devices. PI types considered in this study included email, calendar events, document files and media, such as images and video.

The interviews were conducted with ten academic staff and postgraduate students from the Department of Computing Sciences at the Nelson Mandela Metropolitan University (NMMU), using purposive sampling based on the participants' computing knowledge and experience, and their use of multiple devices for PIM. The participants ranged between 21 and 50+ years of age. All participants had at least six years of computer experience and all but one participant managed their PI daily using a digital device.

Four main questions were asked in the interview regarding PIM across multiple devices, which included the following:

1. **How many digital devices do you currently use to store PI?**
   Sub-questions: What type of device is each of these? What platform(s) does each device use? Which is the main device for managing PI? If you travel, which device do you take with you? Are your devices used for personal or work information or for a combination?
2. **How do you currently manage your PI across these devices?**
   Sub-questions: If you need information on one device that is stored on another device, how would you go about this process? Do you make use of a file manager or email to manage PI?
3. **What problems have you experienced with managing your PI across these different devices?**
   Sub-questions: Do you have any difficulty in managing information on different devices? Do you have any problems with the file manager you use (if any)? Do you have any problems with other methods you use to manage your PI?
4. **Do you have any ideas on how better to manage your PI across these different devices?**
   Sub-questions: Have you heard of Team Viewer or Dropbox? Do you have any ideas on managing your PI other than using removable drives for file transfer? Would you like a tool that would provide an overview of your PI across your different devices allowing access to this PI? If you would prefer such a tool, would you prefer it to be an application installed on your device or a web-based application that you would use in a browser?

3.2 Results

The transcripts of each interview were analyzed with NVivo 10 software (www.qsrinternational.com), using coding techniques to identify themes within the data. The results of the data analysis are described in this section.

Keeping (Storing): The number and combination of devices and platforms exacerbate the current PIM situation. The participants of the interview study made use of various devices to manage their PI. Most participants of the interview study used at least three devices for PIM, with a few participants using five devices for PIM.

All participants of the interview study used a desktop computer provided by the university for PIM or work-related activities. Additionally, all participants commented on using their devices for a combination of personal- and work-related activities. Nine participants made use of their mobile phone for PIM and eight participants had a desktop computer at home that they used for PIM. Six participants made use of a laptop for PIM.

Nine participants had a combination of devices which made use of different platforms. Five participants considered their desktop computer at the university as their main device used for PIM and work-related tasks. Four participants considered a combination of devices as their main devices for different purposes. Four participants mainly used a laptop when travelling and accessing their PI. Four participants travelled with their mobile phones or netbook computers.

Management: Participants make use of email for access, removable drives for additional storage and transfer, and file managers for organization. Participants mainly used different combinations of methods to manage their PI across their different devices. Participants still currently make use of email to access information on one device from another. Participants also mentioned using removable drives either continuously or at some stage for managing PI across different devices.

All participants made use of the Windows Explorer folder structure provided on their desktop computers and laptop devices to organize their PI. Six participants made use of Dropbox, either as their primary mechanism for accessing information across different devices or as a back-up tool. One participant mentioned that he used the same folder structure on each device to "not get lost":

"I have the same folder structure; it's more or less the same. So when I take something in my 'paper folder', I just put it on my 'paper folder' at home. I have that in order not to get lost, because if you have a general folder structure, you will be confused at a certain point of time."

Retrieving: Participants currently find it difficult to access PI across their devices and do not know what they will need to access in future. Participants noted various problems in managing their PI across different devices. Participants identified that the hierarchical folder structure is restrictive mainly due to being limited to categorizing a file in a single folder:

"Finding stuff can sometimes be a problem, or, 'I know I took a photo of this, but now where did I put it?' I actually had that problem the other day, because when I take the stuff off my phone, I don't always put it into the correct place."

Participants experienced problems with naming folders appropriately, especially when backing up information and organizing different versions of PI items. Other issues identified involve versioning issues, having to remember to update folders with the latest files and problems with inconsistent structures between different devices. One participant also mentioned not being able to view information in different ways:

"I organize my photos by dates, events and places, but that gets mixed up sometimes. If I want to have it by date and by location it's difficult. If, for example, for our holiday trips to Knysna, I can't remember what year it was, then it's a pain to go and search each year and check the photo. And then if you do it by location, if you want to find everything that happened last year then you have the other problem, or with people, 'view all the photos that have my little girl in it, that's my bulldog.' Then you have to go and search all the folders…"

Eight participants identified that Dropbox mainly suffers from the same limitations as the Windows Explorer folder structure. Participants identified that a problem with using various methods for organizing files across different devices is that one is required to know beforehand what information is needed to be accessed. A participant noted that he would like all his information to be available:

"…I will most certainly not know beforehand, and I would prefer everything to be available."

Participants identified that it is difficult to keep record of different versions of the same PI item. Participants found problems with the email structure and other file structures not being able to communicate, and that email items and other items cannot be linked or associated, as one participant explained:

"It would be quite a cool thing if you had this integrated view of all your information, because here you've got the email system, which is one system, and then you've got your file structure on a particular device, which is another system, and yet there could be connections between individual emails and a topic."

One participant described a problem he had recently experienced involving accessing information:

"The other day I needed a file. I was at home and my Internet was giving me hassles. I can VPN from home into this PC fairly easily, if I have the Internet connection. And there I'm sitting; I can't get hold of that file. So, I had to drive in to come and fetch a file here, to go back and work on it. So, that's the kind of problem: having no real time, online access to certain information."

Four participants mentioned the information fragmentation problem, some in terms of examples, including the following scenario:

"Here is where things get tricky, because I've got photos on my phone, I've got photos which are on the office PC, and I've got photos on my laptop. In addition, my wife's netbook has got photos on as well, and one can't easily aggregate them together. You can try and bundle them together but they are so massive, you can't really forward them easily by emailing, so you've got to use a memory stick. Even working across my WiFi network at home would be a bit slow."

Viewing: Participants need an IV tool installed on their devices which provides a single UI to facilitate an overall view of their cross-device PI. This requirement is also supported by literature [7], [8]. Four participants noted that if there was such a tool, they would like it to provide some sort of automation in organizing their PI. Participants preferred the tool to be a native application installed on each device, as a browser may provide limitations for such a tool.

Participants provided various suggestions on how better to manage PI across different devices. Participants suggested a search tool, which is capable of searching across different devices to find information. One participant suggested a tool which would intelligently "think" for the user:

"…Maybe also it would remember the things that I did the most, and pre-fetch stuff for me. Rather than saying, 'ok I'm going to give you all this stuff' but 'ok I'm rather going to give you pre-fetched stuff, that I'm watching you and seeing what you're doing', and that would be a cool thing. To be able to see what's there, which I can't

do, I can't access the stuff, and I also can't see what's there. I can't picture it. In fact…I can't visualize it."

4 Design Implications

Several design implications were identified for the IV tool suggested in Section 3.2. These design implications are categorized according to organization, visualization and interaction.

4.1 Organization

1. *Provide a virtual storage solution that aggregates PI in a single location.* There is currently limited support for accessing PI across different devices. The virtual storage solution could allow the PI to reside on the original device but aggregate references to PI items, which could be accessed when the different user devices are connected, as suggested in [10]. This could address the issues identified by participants that they are unsure beforehand which PI items will be needed.
2. *Provide support for association of linked PI items.* Supporting linking of PI items could assist with version control of PI. Support for PI item association as used in Phlat [11], could also address the issue of not being able to link items of PI collections in separate applications.
3. *Provide tagging to assist with retrieval.* Tagging PI items could also assist with version control. Manually and/or automatically tagging PI items could assist with re-finding information when searching, as used in Phlat [11].
4. *Include additional facilities other than general PI types.* In addition to the common PI types, email and contacts were regularly mentioned in the interview study. Contact management is another important task of PIM which should be supported [5].

4.2 Visualization

1. *Make use of a single UI to visualize PI across different devices.* All participants described scenarios where they experienced information fragmentation problems and nine welcomed the suggestion of a tool that provides a single UI to visualize PI across several devices. Systems, such as Phlat [11] and ZOIL [7], provide a single UI for different PI types, but do not fully support cross-device PIM.
2. *Visualize PI using suitable IV techniques.* It was identified that the list and indented-list may not be suitable for viewing PI. Suitable IV techniques could address the restrictive nature of the hierarchy and its ineffective use of screen space. Timelines have been used in various PIM systems, including VizMe [8].
3. *Provide different PI views.* Participants noted that only having one view of their PI is not sufficient and that the proposed tool should provide different PI views.

It was suggested that while time is useful for PIM, it should not be the only aspect considered [12].

4. *Consider each device's constraints.* Eight participants managed their PI across at least three devices. Thus, UI plasticity should be considered to ensure that the UI design will support several different devices. [13].

4.3 Interaction

1. *Provide intelligent searching across devices.* In addition to the IV techniques used to support browsing, intelligent searching of PI also needs to be facilitated. A combination of keyword searching, in-text searching and filters, as used in Phlat [11], could be used to allow searching PI across different devices.
2. *Provide support for file sharing and collaboration.* The current hierarchical organization method, with the exception of tools such as Dropbox, does not support file sharing or collaboration, but needs to be supported by the proposed tool [14].
3. *Provide full functionality associated with PI items.* Due to the information fragmentation problem, if an item cannot be viewed, it cannot be accessed. The full functionality of each PI item should be facilitated, as supported by ZOIL [7].
4. *Support immediate access to PI items.* The proposed tool should provide instant access to PI items if the device(s) are available, and could possibly include offline accessibility of certain PI items. Additionally, the UI delay problems experienced using Phlat should be avoided [11].

5 Conclusion

An interview study was conducted to determine how users currently manage PI across different devices and the problems that are experienced with PIM. A limitation of the study was a small sample size (n=10). Participants made use of various combinations of methods to manage their PI across multiple devices. The most popular methods included the use of Email, the folder structure, removable drives and Dropbox. The participants of the interview study identified a number of problems in managing PI, mainly attributed to not being able to view and access PI across different devices. The suggestion of an IV tool to support access to PI across multiple devices received a positive response. The results of the interview study were used to propose several design implications for organization, visualization and interaction. Future work will involve the design of an IV tool to visualize a user's PI across multiple devices, using suitable IV techniques.

Acknowledgements. Acknowledgements are due to the NMMU/Telkom Centre of Excellence for funding this work.

References

1. Jones, W., Bruce, H.: A Report on the NSF-Sponsored Workshop on Personal Information Management, Seattle, WA (2005)
2. Aires, J., Gonçalves, D.: Personal Information Dashboard - Me, At a Glance. In: PIM 2012 Workshop, pp. 1–8 (2012)
3. Golemati, M., Katifori, A., Giannopoulou, E.G., Daradimos, I., Vassilakis, C.: Evaluating the Significance of the Windows Explorer Visualization in Personal Information Management Browsing Tasks. In: Proceedings of the 11th International Conference Information Visualization (IV 2007), pp. 93–100 (2007)
4. Tungare, M.: Understanding the Evolution of Users' Personal Information Management Practices. In: Proceedings of the Eleventh IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2007 Doctoral Consortium), pp. 586–591 (2007)
5. Indratmo, J., Vassileva, J.: A Review of Organizational Structures of Personal Information Management. Journal of Digital Information 9(1), 1–19 (2008)
6. Barreau, D.K.: Context as a Factor in Personal Information Management Systems. Journal of the American Society for Information Science 46(5), 327–339 (1995)
7. Jetter, H.C., Gerken, J., König, W.A., Reiterer, H.: ZOIL - A Cross-Platform User Interface Paradigm for Personal Information Management. In: PIM 2008: Proceedings of the Workshop on Personal Information Management (CHI 2008), pp. 1–9 (2008)
8. Gomes, P., Gama, S., Gonçalves, D.: Designing a Personal Information Visualization Tool. In: Proceedings of the 6th Nordic Conference on Human Interaction (NordiCHI 2010), pp. 663–666 (2010)
9. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 3rd edn., pp. 173–202. SAGE Publications, Los Angeles (2009)
10. Karypidis, A., Lalis, S.: OmniStore: Automating Data Management in a Personal System Comprising Several Portable Devices. Pervasive and Mobile Computing 3, 512–536 (2007)
11. Cutrell, E., Robbins, D.C., Dumais, S.T., Sarin, R.: Fast, Flexible Filtering with Phlat - Personal Search and Organization Made Easy. In: Proceedings of CHI 2006, pp. 261–270 (2006)
12. Latif, K., Min Tjoa, A.: Combining Context Ontology and Landmarks for Personal Information Management. In: 2006 International Conference on Computing & Informatics (ICOCI 2006), pp. 1–6 (2006)
13. Thevenin, D., Coutaz, J.: Adaptation and Plasticity of User Interfaces. In: i3-spring99 Workshop on Adaptive Design of Interactive Multimedia Presentations for Mobile Users, pp. 1–8 (1999)
14. Collins, A., Kay, J.: Collaborative Personal Information Management with Shared, Interactive Tabletops. In: PIM 2008: Proceedings of the Workshop on Personal Information Management (CHI 2008), pp. 44–52 (2008)

Mobility Matters: Identifying Cognitive Demands That Are Sensitive to Orientation

G. Michael Poor[1], Guy Zimmerman[2], Dale S. Klopfer[3], Samuel D. Jaffee[3], Laura Marie Leventhal[2], and Julie Barnes[2]

[1] School of Engineering and Computer Science, Baylor University, Waco, TX, 76798
[2] Dept. of Computer Science, Bowling Green State University, Bowling Green, OH, 43403
[3] Dept. of Psychology, Bowling Green State University, Bowling Green, OH, 43403
Michael_Poor@baylor.edu,
{gzimmer,klopfer,jaffees,leventha,jbarnes}@bgsu.edu

Abstract. Prior studies have shown benefits of interactions on mobile devices. Device mobility itself changes the nature of the user experience; interactions on mobile devices may present better support for cognition. To better understand cognitive demands related to mobility, the current study investigated presentations on a mobile device for a three-dimensional construction task. The task imposed considerable cognitive load, particularly in demands for mental rotation; individual differences in spatial ability are known to interact with these demands. This study specifically investigated mobile device orientations and participants' spatial ability. Subjects with low spatial ability were able to complete the task more effectively when shown the presentation in a favorable orientation. Individuals who saw the presentation in an unfavorable orientation and those of low spatial ability, were differentially disadvantaged. We conclude that mobility can reduce cognitive load by limiting demands for spatial processing relating to reorientation.

Keywords: Mobility, Mental Rotation, Presentation Orientation, Spatial Ability.

1 Introduction

Recently there has been an enormous expansion in the sale and use of mobile devices as information appliances. It is reasonable to ask why, or more specifically, what is it about mobile devices that make them attractive, as compared to fixed stationary devices. There are a number of obvious answers: size, convenience or price. However, if users did not find the mobile device to be at least as useful as a fixed counterpart, would these devices be so successful? It seems likely that that users experience mental workload advantages as an outcome of the very mobility of the mobile device for some tasks. For example, extracting and interpreting instructions for construction tasks[1], tasks that impose considerable mental workload in the form of mental rotation and spatial processing on users, would fit this description.

[1] Construction tasks are ubiquitous: assembling a child's bicycle, a piece of furniture or folding a paper airplane being common examples.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 193–210, 2013.

In this paper, we first review background literature, including a discussion of the likely cognitive load issues at play in construction tasks on mobile devices, specifically mental alignment of the presentation to the built object, achieved via mental rotation. We present a study that examined the role of physical device orientation on performance on a construction task. Additionally, in the study, we explored the role of an individual difference variable, spatial ability, on performance of the construction task. Finally, we present the results and we discuss the ramifications of our findings for designers.

While the constant changes in technology make the definition of *mobile device* or *mobile interaction* moving targets, for the purposes of this paper the terms *mobile device* or *mobile interaction* will imply a handheld computing device possessing a display screen and input mechanism. This definition includes cell phones, smartphones, tablets, handheld GPS (global position) systems and PDAs (personal digital assistants), but excludes traditional desktop computers with fixed displays. The key defining feature that we focus on in this paper is the ability of the user to easily reposition the display device in any desired orientation. For clarity, we will use the term *mobile device*.

2 Background Literature

2.1 Mobility Matters

It is widely believed that different interactions engage different user capabilities and draw on different elements of human cognition [cf. 12]. More specifically, researchers have noted numerous HCI issues for mobile devices [cf. 1, 4, 6, 20]. Of interest here, [31, 32] found that people used differential strategies to varying degrees of success when performing a three-dimensional construction task using instructions presented on mobile and non-mobile devices. For these tasks, the instructional presentation included interactive 3D models. Traditionally, the instructions for completing construction tasks are presented on paper, with written directions often annotating visual representations of the assembly process. Such paper based instruction presentations are notoriously difficult to use. [30] has suggested that difficulties arise in part from task demands for mental rotation. Interactive presentations offer relief from some of the limitations of the traditional paper format. In particular, interactive presentations allow the three-dimensional displayed object to be viewed from any vantage point – giving the builder a better sense of the spatial relationships of the parts of the assembly. A number of factors potentially impact performance on construction tasks, including the nature of the presentation and the spatial ability of the participants [23]. More importantly for the current study, instructions on a mobile device allow users to physically take the instructions ‘to the object’: physically orienting the instructions by holding the device proximate to a built object. Further, the richness of construction tasks would seem to make them ideally suited to highlight mental workload differences between mobile and non-mobile devices.

[31, 32] compared performance on a construction task between a mobile device presentation and a fixed upright display presentation. They found that the mobile

device users were more efficient in building the target object than the fixed presentation users. [31] also found that at least 25% of the persons with the mobile device employed a strategy of moving and aligning the mobile device to the object being built during at least one building step and all but one participant removed the mobile device from its starting position during the building process. Interestingly, in [31] a number of participants with a stationary display brought the object being built to the display. In other words, in both conditions, participants aligned the physical device with the object being built. When the person could bring the presentation to the object instead of vice versa, performance was markedly improved. [31] concluded that the participants using the non-mobile presentation found the process of aligning the object to the screen was awkward, forcing them to mentally rotate in order to realign the images in the instructions to the constructed object. The participants *could* have interactively realigned the 3D presentation, in either device condition, at any point and it is possible that the subjects in [31] did this; it is notable that the mobile presentation users had better performance regardless.

2.2 Does Orientation Matter?

When a person is following computerized instructions that include visual presentations to construct an object, they have several choices as to how to align the spatial relations in the visual representation to those of the target object. They can physically move the presentation to the target via the mobility of the device, physically align the target to the presentation, manipulate the presentation of the digitally displayed 3D object, and/or perform any or all these operations mentally, without manipulating the object or the presentation. In other words, in a construction task, when the visuals in the presentation and the actual built object are misaligned, the user will mentally, physically, or interactively perform transformations to make the alignment. [30]'s results *suggest* that users are most successful when they choose to physically realign the device to the target and that they may be surprisingly unwilling to realign interactively.

As we consider the fact that in [31], mobile device subjects were able to move the device to realign the images in the presentation to the target while the fixed desktop subjects appeared to more often do this mentally, the next obvious question should be, *does it matter*? If desktop subjects are doing more mental rotations of the presentation, is there a cost? [25] claimed that internal (cognitive) representations share a second-order isomorphism to the world they represent. One outcome of this conjecture is that the greater the angular disparity between the starting orientation of an object and its rotated position, the more effort required for rotation of the object both in the real world and in their internal representations [26, 31]'s finding of performance advantages for mobile device users suggests that the mobility of the device may reduce user cognitive load by reducing need for mental rotations.

Some studies indicate that cognitive load increases as a person does more mental rotations [e.g.14, 15]. That some participants in [31] aligned the object that they were creating with the image on the fixed display whereas others aligned the mobile device with their built object highlights an obvious but critical difference between fixed vs.

mobile display devices: mobile devices allow the user to change the orientation of the display, which can change the frame of reference used to specify spatial features of an object, such as identifying its top or its left or right side. In the fixed display condition in [30], participants' options for rotating the presentation to align it with an external object were: mental rotation, rotation of the real object relative to the fixed presentation and/or rotation of the interactive 3D presentation. In the mobile device condition of [31], subjects had a fourth option – they could rotate the device containing the presentation.

When an observer encounters an object in the world, two frames of reference -- and the spatial relations they define -- are important to consider. First, there is an egocentric reference frame that defines spatial relations from the observer's viewpoint (e.g., up/down, left/right). The egocentric up/down axis is typically defined by gravity, with left/right defined by what's to the left and right of the viewer's midline, respectively. Because of the invariance of gravity, the up/down axis is a primary reference frame for defining the tops and bottoms of objects and whether one object is above (or below) another [cf. 27].

There is also a reference frame intrinsic to the object itself whereby spatial relations among parts of the object are specified. Object-centered reference frames can be defined by a variety of object characteristics, such as an object's focal point [5], an axis of symmetry or elongation, or surface markings [22]. The object in Figure 1(a) has an intrinsic axis of elongation, defined by the dotted line; the triangle in Figure 1(b) has an intrinsic axis of symmetry and a focal point at its upper vertex. With both objects, the intrinsic axes are aligned with the egocentric up/down axis. If these were animate objects, people would likely construe the upper portion of each object to be its head; if they were to move they would move upwards. The triangle in Figure 1(c) is probably seen as pointing up, illustrating the primacy of the egocentric up/down axis. With its three axes of symmetry, the triangle in Figure 1(c) actually points in three directions, but the tendency to see it point up is due, in part, to the viewer using the up/down axis to assign spatial relations.

When an observer encounters an object displayed on a screen, a third frame of reference comes into play: the reference frame defined by the edges of the display. Figure 2 shows the triangle from Figure 1(c) surrounded by a rectangular frame, much as how the triangle would be seen on a desktop display. Note how the two reference frames – the egocentric up/down axis and the vertical axis of the display – are aligned, and the triangle is seen as pointing up. Due to the alignment of the viewer's and the display's reference frames, a desktop display's reference frame is redundant, providing the same spatial relations as the viewer's up/down axis.

With a mobile device, the reference frame defined by the display need not be aligned with the viewer's up/down axis. Figure 3 shows the triangle from Figure 2 within a rotated rectangular frame. Here, there is a strong tendency to see the triangle pointing down to the left, although it still is possible to see it pointing up. That is, by changing the display's frame of reference, the "head" of the triangle shifts from being the upper vertex to the one in the lower left. Moreover, within the context of the rotated frame, the triangle is likely seen as heading strictly to the left instead of down and to the left.

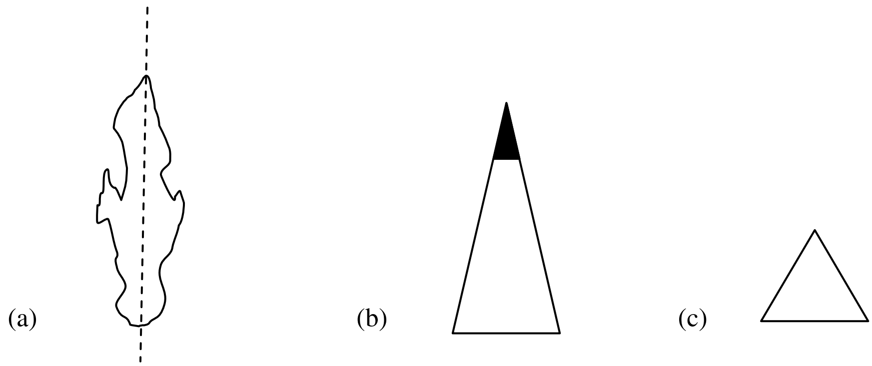


Fig. 1. (a,b,c) Objects with vertical axes of symmetry

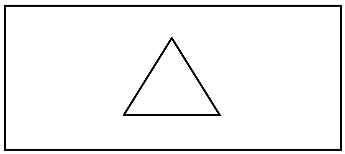


Fig. 2. Triangle object inside a display's frame of reference

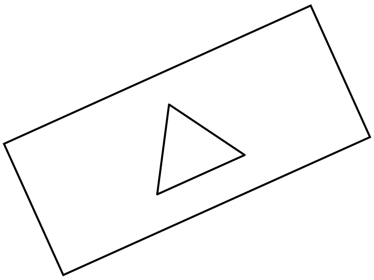


Fig. 3. Triangle inside of rotated display

As noted earlier, the mobility of mobile devices allows one to alter the display's frame of reference quite easily. Indeed, the rotated rectangular frame in Figure 3 represents just one of an infinite number of display-defined frames of reference available to the mobile device user. The interactive graphics that allow for the shape to be rotated within the display gives the user freedom to define the alignment of the presentation using whatever reference frame he or she chooses.

In the construction task in [31] there is a fourth frame of reference to consider: the one that defines the top, bottom, left, or right of the target object being constructed. In principle, there are more opportunities for the four reference frames (i.e., viewer, display, displayed object, constructed object) to be misaligned when using a mobile device than when using a fixed desktop display (i.e., viewer = display, displayed object, constructed object). Much research in cognitive psychology indicates that there is a cost in perceiving objects when viewer and object frames of reference are misaligned [cf. 13, 27]. Thus, misalignment of two frames of reference, viewer and object, can have information processing costs. To our knowledge, no research has been done that examines how users cope with the possibility of there being multiple opportunities for misalignment.

2.3 Evidence of the Importance of Presentation Orientation

Even with the possibility of physically reorienting a presentation, determining the *best* interactive realignment may in and of itself impose significant cognitive load and involve mental rotations in planning, especially for low spatial ability users. Given that many contemporary mobile devices only automatically realign in cardinal directions, should it turn out that physical orientation and spatial ability do interact on a construction task, persons of low spatial ability who cannot physically reorient to the *best* orientation will be disadvantaged unless the cognitive load for interactive realignment can be reduced.

Three older studies point to the importance of orientation in a presentation of interactive visual information. [24] found that, for map-based navigation assistance, physical rotation is the most effective form of track-up alignment on handheld mobile devices This was due to the users' difficulty to recognize a map when automatically rotated, especially when the users were not looking at the map during the time of rotation.

In addition, [28] described a comparative study of the effectiveness of four different presentations of instructions for an assembly task: printed manual, monitor-display, see-through head-mounted display, and spatially registered augmented reality (AR). Measurements were task performance (time and accuracy) and perceived mental workload. The task consisted of 56 procedural steps building an object with Duplo blocks. Participants in the spatially registered AR treatment made significantly fewer assembly errors. The authors concluded that the improvement in the AR condition was due to reduced demand for attention switching. Because the spatially registered AR appears directly on the object, it was also thought that the participants did less mental transformations between the instructions and the object.

[7] reported on a design tool to build three-dimensional, interactive and movable polyhedrons. In evaluating this tool, they found that users had a preferred orientation for the designed polyhedrons. When the figures were moved from the preferred orientation, subjects found them to be more difficult to sketch (reproduce by hand). Some participants reported elements of the preferred orientation include: 1) preference for vertical

as opposed to horizontal edges (preference for either type of edges as compared to diagonal edges), 2) bilateral symmetry, and 3) stability as indicated by the polyhedron resting on a face as opposed to resting on a vertex.

2.4 The Role of Spatial Ability

Performing mental transformations, such as those described by [14] can impose a workload on working memory. In particular, the mental rotation processes can be time-consuming and error prone, particularly as the complexity of the object being rotated increases and its familiarity decreases [3, 10]. Just how much effort the mental rotation processes require also depends upon an individual's spatial ability, i.e. the ability to generate, retain, and transform well-structured mental images [16, 17]. Individual differences in spatial ability are related to individual differences in working memory function [17, 19], with transformations such as mental rotation taking longer for users with lower spatial ability, as measured by paper-and-pencil standardized spatial ability tests [8]. Mobile devices potentially provide a means for users to align a displayed object with their own egocentric up/down, limiting the need to engage in mental rotation in order to achieve alignment. The savings would be greater for those with lower spatial ability and would potentially expand the usefulness of the device to a larger population.

2.5 Summary: Background Literature

Prior research has suggested four intersecting themes: 1) Performance on a construction task is better with a mobile device than on a fixed display device 2) Construction tasks engage mental workload, much of which is involved in mental rotation to align disparate frames of reference. The cost of mentally aligning an egocentric and presented object-centric frame of reference is known to be high; the cost of realigning those frames of reference plus others from a display and a built object are not known 3) Mental rotation requires significant mental workload and 4) People differ in their abilities to perform mental rotation; those of lower spatial ability, as measured on standardized tests, find spatial tasks like mental rotation, more difficult than those of higher spatial ability. Taking these themes together, we suggest that performance on a construction task is better with a mobile device as compared to a fixed display device because the mobile device participants are able to lessen some of the mental work of aligning the presentation to the object to be built. In [20] mobile device participants accomplished the needed rotations by a combination of mental and physical rotations rather than mental rotation alone. We hypothesize that mental rotation, interactive rotation of the presentation, or rotation of the artifact is more difficult than rotating the device itself. We suggest that when rotating physically, with the immediate visual feedback as the virtual and physical object align, the participant does less mental rotation, thus reducing their mental workload.

Recognizing that people differ in their ability to do mental rotation, those of lower spatial ability, should be differentially more impaired with a stationary device – [20] did not measure the spatial ability of their participants, so we cannot be sure of this conjecture from their results relating to spatial ability.

In the current work, we start with the assumption that part of the power of a mobile device comes from the reduction in necessary work of mental rotations. We speculate that this advantage may extend further for those who are more challenged by deficits in their ability to perform mental rotations. In our study, we seek to demonstrate that having the visual presentation for a construction task aligned in a particular way, as one would be able to do with a mobile device, would lead to superior performance, than having the presentation in other orientations.

3 Study: Impact of Orientation of Presentation and Spatial Ability on Construction Task Performance

For this study our hypothesis is:

Orientation of a mobile device, in combination with participant spatial ability will affect performance on a construction task.

We specifically hypothesize that at least one orientation will lead to better performance on the construction task by leading to fewer differences in frame of reference between the built object and the presentation by reducing mental rotations (and lowering cognitive/working memory load). However, because the built object itself is in a number of orientations during the presentation, we do not predict which orientation(s) would be favorable and which would impede subjects' performance on the task. Following prior work on the interaction between spatial ability and mental rotation, we also predicted that persons of low spatial ability would be differentially hampered in the less favorable orientations.

3.1 Experimental Design

We have two independent variables: Mobile Device Orientation and Spatial Ability, and two dependent performance variables, described in Section 4, relevant to the task. [20] found that time on the task was non-informative; it was not considered as a dependent variable.

3.2 Mobile Device Orientation

In the study, the mobile device was physically anchored in four orientations (denoted: *left*, *right*, *top*, *bottom*) as shown in Figure 4a. Figure 4b shows the experimental setup with the device in the *right* orientation. The presentation of the instructions for the construction task was symmetric relative to both vertical and horizontal orientations and we collapsed the orientations into two categories: UpDown and LeftRight. While fixing the mobile device may seem counter-intuitive (removing the “mobile” aspect of the device), we have done so in order to allow for greater experimental control in order to study the effects of the frame of reference imposed by the display of the mobile device.

The participants sat upright at the table facing the instructions and were to the extent possible, in a fixed egocentric orientation; participants' position was set so that they did not reorient the presentation by reorienting their own viewpoint.

3.3 Spatial Ability

We measured spatial ability using the Card Rotation task, a measure of two-dimensional mental rotation [8], an individual ability that should be at play, at least in part, in our construction task. Because this is not a power test (i.e., the task does not get harder at the end) the scores were calculated by subtracting the total number of wrong responses from the total number of correct responses. Our median subject score was 69. The minimum and maximum scores were -58 and 154 respectively. Using median split, participants were grouped into two categories: high and low spatial ability.

3.4 Participants

Thirty-two participants, drawn from undergraduate computer science classes, completed the task to their satisfaction with 16 persons in the UpDown condition and 16 persons in the LeftRight Condition. Two participants from each spatial ability category were dropped; they were the participants closest to the median spatial score of 69, leaving a pool of 28 participants with 14 participants in each orientation category. In terms of spatial ability, this change to the pool left 15 participants with high spatial ability and 13 with low spatial ability. Nine participants were assigned to the High spatial ability/UpDown condition, five participants to Low spatial ability/UpDown, 6 participants to High spatial ability/LeftRight and eight to Low spatial ability/LeftRight. A chi-square analysis of this frequency distribution was not significant, showing that the assignment to condition was independent from spatial ability.

3.5 Materials and Task

In previous studies we explored the effectiveness of interactive 3D graphics as a part of a system to deliver instructions for a construction task: origami paper folding [cf. 2, 30]. Paper folding does possess many representative characteristics of construction tasks: the task is non-trivial, it requires multiple manipulation steps, and it results in a 3D artifact. Researchers in multimedia learning make a distinction between single and dual presentations of instructional information [18, 21]. In single presentations, instructions are typically presented in text alone, whereas in dual presentations instructions are usually presented in text with accompanying still images or other representations [21]. A number of studies indicate that dual presentations lead to better performance [e.g. 18, 29]; this advantage has been shown with the present task [31].

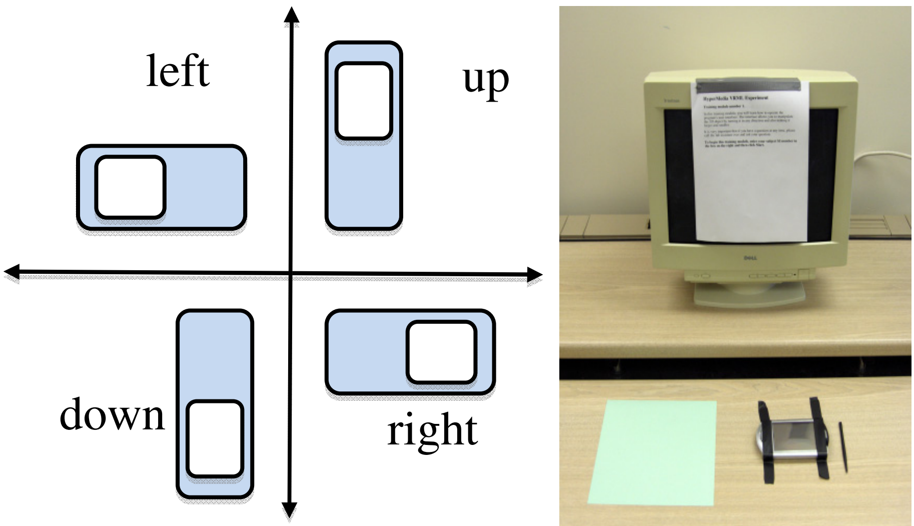


Fig. 4. a) Orientation conditions for the mobile device. b) Set up for *right* condition. Note: device is fastened to the table; paper instructions are anchored to the front of a monitor.

Our task, identical to the task used in [31], was to fold an origami whale in 25 paper folds (and unfolds), with the instructions for making the folds presented in a series of 12 steps. Approximately ½ of the steps involve 2D folds, the remainder were 3D folds, including steps to form the mouth, fins and tail. A completed whale is shown in Figure 5.

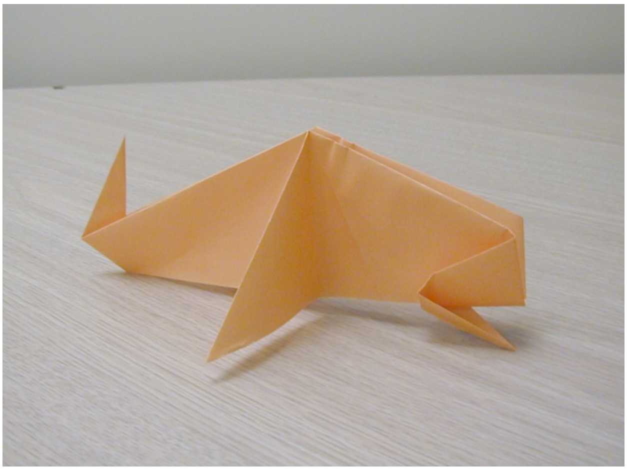


Fig. 5. The completed whale

The 3D presentation was delivered on a HP IPAQ h5455 with stylus and Microsoft Pocket PC version 3. The h5455 used the 400 MHz Intel PXA250 processor with 64 megabytes of RAM. The IPAQ and the interaction user interface were much less familiar to the participants than a smartphone or handheld GPS system; it was our thought that participants were not able to engage familiar tasks, outside of the specific experimental task, during the experiment.

The 3D interactive presentation was implemented using VRML 2.0 (Virtual Reality Modeling Language). The VRML model was rendered within Pocket Internet Explorer 5.5 using the Pocket PC Cortona VRML client plug-in. The display screen was 3.8" (diagonal) with resolution 240x320 with 16-bit color. There were two visual components to the display: the virtual sheet of origami paper (VOP/model) and the user animation/step interface. A simulated sky/horizon was also implemented to provide a spatial frame of reference for the origami paper. The user moved through the steps by clicking the forward/backward buttons on the interface. During each step the user could start/stop an animation of the desired operation (e.g., create a fold) by clicking the play/pause button. For each step that required a fold operation, the animation began by highlighting the desired fold line on the VOP; the actual fold operation was then performed on the VOP. The user also had the ability to rotate the VOP/model in any direction at any time. Technical VRML implementations details can be found in [20]. Figure 6(a) shows the 3D interactive presentation. Subjects had access to written instructions with figures printed on paper; these were anchored in front of the subject using a "flipchart" style of presentation (see Figure 6(b)).

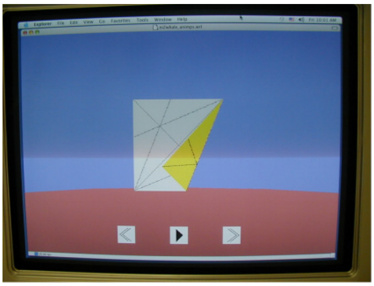


Fig. 6. a) The 3D interactive presentation b) The experimental setup; note the built whale is misaligned with the presentation (inset)

3.6 Procedure

The subjects completed the spatial ability test and a training task to familiarize them with the interactive controls. Then they completed the whale-folding task of twelve folding steps, used in [30, 31, 32].

3.7 Scoring of the Folded Whales

Each participant's whale was scored by evaluating every fold on the origami paper. Each fold was scored as Correct, Error and/or Recrease. If a subject performed a fold incorrectly and then folded correctly, these were scored as different folds (one correct and one error). In addition, a correct fold might be recreased. The range of Correct Folds was 0 - 25; of Error Folds and Recrease Folds was 0 – no maximum. Two people graded the constructed origami whales with an inter-rater reliability of 0.99.

3.8 Summary: Relating the Study to the Hypothesis

[31] posited that construction tasks in general and the one used in this study specifically impose significant mental workload as the participants must align four frames of references (themselves, the presented instructions, the display and the built object) using a combination of mental and physical rotations. In this study, physical rotation of the display and rotation of the egocentric (participant's) frame of reference were limited or fixed by the experimental setup. Realignments were possible through the interactive rotation of the presentation, physical movement of built object and mental rotation of any of the component elements. All participants were free to rotate the interactive presentation or to move the built object. Only demands for realignment by mental rotation differed by orientation. We expected to find that persons of high spatial ability would outperform those of low spatial ability across the board, simply by the fact that the cost of mental rotation is higher for individuals of low spatial ability. Our alternative orientations simulate the various positions that a mobile device could be in. If one orientation leads to better performance, especially for persons of low spatial ability, then we would have shown that physical alignment of the display to the other experimental components does reduce mental workload on our task.

4 Results

We considered the impact of the independent variables, Orientation (UpDown vs. LeftRight) and spatial ability (High vs. Low) on two dependent variables: Adjusted Number of Correct Folds (defined as Number of Correct Folds minus Number of Error Folds) and Number of Correct Recrease Folds (multiple redundant correct folds). The subjects in this study did well on the task, as indicated by our Adjusted Correct Folds measure, with an overall mean of 16.3.

As we had multiple dependent variables, we first conducted a multivariate analysis of variance (MANOVA)[2]. The dependent variables (Adjusted Number of Correct Folds, and Number of Correct Recrease Folds) were included in the MANOVA. The main effects were significant; the interaction between Orientation and Spatial Ability was not. (Wilks' Lambda = 0.714, F(2.0,,23.0) = 4.603, p < 0.021 for Orientation; Wilks' Lambda = 0.672, F(2.0,23.0) = 5.607, p < 0.010 for Spatial Ability).

[2] We performed a MANOVA because intuitively, it would make sense that our dependent variables were intercorrelated in some way. The MANOVA identifies significant intercorrelations among dependent measures such that these measures are not incorrectly identified as significant effects of the independent variables. While we recognize that our test may be somewhat lacking in power, we feel that, in light of the experimental design, these statistical procedures are the most appropriate.

A two-way ANOVA revealed significant main effects of Orientation and Spatial Ability on the Adjusted Number of Correct Folds [$F(1,24) = 9.323$, $p < 0.005$) and $F(1,24) = 10.539$, $p < 0.003$, respectively]. No univariate analyses of Number of Correct Recrease Folds were significant. The means of the dependent variables are shown in Table 1. From the means for Adjusted Number of Correct Folds, persons with high spatial ability or the Left-Right orientation performed significantly better than persons with low spatial ability or UpDown orientations.

Table 1. Mean Adjusted Number of Correct Folds and Correct Recrease Folds by Spatial Ability (High Spatial Ability vs. Low Spatial Ability) and Orientation (Updown vs. LeftRight) (standard deviations in parentheses)

	Adjusted Correct Folds		Number of Correct Recrease Folds	
	UpDown	LeftRight	UpDown	LeftRight
High spatial ability	18.00 (6.90) n=9	21.50 (3.27) n=6	4.33 (2.00) n=9	2.83 (0.75) n=6
Low spatial ability	7.00 (4.30) n=5	18.00 (4.30) n=8	3.60 (3.36) n=5	4.88 (2.96) n=8

Note: Maximum value for Adjusted Number of Correct Folds is 25.

In order to understand these findings in detail, we examined the two components of Adjusted Number of Correct Folds (viz., the Number of Correct Folds and the Number of Error Folds) separately as a function of Orientation and Spatial Ability. High spatial ability participants made more correct folds than those with lower spatial ability ($F(1,24) = 6.349$, $p < 0.019$), and participants in the LeftRight orientation made more correct folds than those in the UpDown orientation ($F(1,24) = 4.678$, $p < .041$). The Orientation X Spatial Ability interaction was not significant for Number of Correct Folds. The Number of Errors, on the other hand, was significantly greater in the UpDown than in the LeftRight orientation ($F(1, 24) = 5.410$, $p < .029$). There was also a main effect of Spatial Ability, with Low Spatial Ability participants making more errors than their High Spatial Ability counterparts ($F(1,24) = 4.566$, $p < 0.043$). The Orientation X Spatial Ability interaction was also significant for Number of Error Folds($F(1,24) = 4.424$, $p < 0.046$), showing that the performance of participants with lower spatial ability was most affected by the unfavorable UpDown orientation. The means and standard deviations are listed in Table 2.

4.1 Discussion: Study Results

Our results show significant disadvantages for participants who were in the UpDown orientation or low on a measure of two-dimensional rotational spatial ability. The disadvantages were exacerbated for persons who were both in the UpDown condition and had low spatial ability for the dependent measure, Number of Error Folds. The fact that persons of high spatial ability performed better on the construction task is not surprising – we anticipated that the construction task imposed a higher cognitive cost for persons of low spatial ability. The fact that the two orientations led to differences in performance suggests that the position of a device imposes differential demands in mental work. With a real mobile device, individuals can reduce the workload by moving the device. That the LeftRight orientation was related to improved performance suggests that the critical elements of the presentation best lined up with the object being built in this orientation. Future studies could explore the specific elements of the presentation that was influenced favorably by the LeftRight orientation.

Table 2. Mean Number of Correct Folds and Number of Error Folds by Spatial Ability and Orientation (standard deviations in parentheses)

	Number of Correct Folds		Number of Error Folds	
	UpDown	LeftRight	UpDown	LeftRight
High spatial ability	21.56 (4.50) n=9	24.33 (0.82) n=6	3.11 (2.93) n=9	2.83 (2.92) n=6
Low spatial ability	15.40 (6.54) n=5	20.88 (5.89) n=8	8.4 (4.72) n=5	2.88 (2.59) n=8

Note: Maximum value for Number of Correct Folds = 25.

5 Vertical Orientation: Does It Matter?

In our study, we manipulated the orientation only in two spatial dimensions on a flat table surface. It is possible that the flat manipulation is not ideal and that the third dimension of vertical could be key as well. In a follow up pilot study we compared the Up subjects from the UpDown group to a group of participants who saw the presentation on the mobile device in a stationary vertical position. The setup is shown in Figure 7. All other aspects of the pilot study procedure were identical to our primary study, described in Sections 3 and 4.

We chose to compare the Vertical orientation to the Up orientation from the tabletop conditions in our original study, because the Vertical presentation is also in the Up orientation but rotated 90 degrees vertically from the desk surface. We had eight subjects in the Vertical orientation (3 lows and 5 high spatial ability) and we compared this group to the original eight subjects from the Up (5 lows and 3 high spatial ability). We conducted a MANOVA with dependent variables Adjusted Number of Correct, Folds and Number of Correct Recrease Folds for the two independent variables, Orientation (Vertical vs. Up) and Spatial Ability. The only significant effect, following the MANOVA was for Spatial Ability (Wilks' lambda= 0.561, $F(2.0, 11.0) = 4.3$, $p < 0.042$). Separate univariate ANOVAs showed that Spatial Ability had a significant effect on Adjusted Number of Correct Folds ($F(1,12) = 9.377$, $p<.01$) only. So the independent effect of 3D vertical orientation did not have a significant impact on performance nor did it interact significantly with spatial ability. The means for the two groups were low spatial ability participants= 13.567, high spatial ability participants = 23.3.

Fig. 7. Mobile device in a fixed vertical orientation

5.1 Discussion: Pilot Study Results

For the whale folding task, the subjects were not folding the physical whale vertically; the pilot study results suggest the vertical orientation of the mobile device did not align effectively with the physical whale. The Vertical mobile device alignment also did not yield different results from the Up tabletop position. We posit that Up orientation placed the interactive presentation in a mostly vertical orientation like the position of the display in the Vertical orientation of the pilot. The LeftRight orientation, superior to the UpDown orientation, made up in part by the Up orientation, likely positioned the salient elements of the interactive presentation predominantly in the position that participants favored during whale construction and that limited demands for mental realignment.

6 Conclusions

We found that user performance was significantly affected by the physical orientation of the mobile device, spatial ability and their interaction on a paper folding task. We make our first conclusion – mobility does change the user experience at least for some tasks; enabling reorientation potentially reduces the need for mental rotations. While the study was not specifically designed to systematically control the orientations of the 3D interactive images shown, it appears that when making folds, performance was better when the model was aligned left and right. It is noteworthy that the subjects could have interactively changed the orientation of the 3D model at will. The fact that the orientation of the mobile device was a significant factor in performance suggests that the subjects did not, on at least some occasions, rotate the model to the more favorable orientation. This finding suggests that the frame of reference imposed by the display may have had greater power over the participants' mental representations of the task than the mental representation of the object itself. Prior studies have demonstrated that there is a cognitive cost for mental rotations effected to align disparate frames of reference. In order to select an interactive rotation would have required the subject to mentally rotate the model before interactively reorienting the presentation. Our results suggest that at least some of the time, subjects make tradeoffs between impaired performance and the mental effort required for mental rotations or for planning for interactive rotations. We conclude that mobility matters in part because mobility allows users to put presentations into favorable orientations and reduces the need for mental rotation.

Our results have implications for the design of presentations for mobile devices. We note that many contemporary mobile devices automatically alter the orientation of presentations on the screen, based on the physical position of the device. Should a person move a device to limit the mental work of rotating the presentation, only to have the device itself rotate the presentation, the automatic re-rotation could actually add to user workload.

Mobile devices and visual-spatial presentations of information are pervasive and likely to become more so, especially for tasks in which the mobile device can be moved to close proximity of the task [cf. 9]. Designers will be increasingly challenged to build user interfaces that do not inadvertently incorporate significant cognitive barriers to users in the form of memory load, especially for low spatial ability individuals. For individuals who are unable to physically reorient a device, our results suggest that they too may be potentially disadvantaged as they may be forced to rely initially on mental rotation to plan their interactive reorientation. Designers potentially may be able to expand the usefulness of their designs to broader spectra of the population by limiting the need for mental rotation via the mobile properties of the device.

Acknowledgements. We thank Ron Buchanan and Chris Glenn for their assistance on this project.

References

1. Abowd, G.D., Mynatt, E.D.: Charting Past, Present, and Future Research in Ubiquitous Computing. ACM Transactions on Computer-Human Interaction 7(1), 29–58 (2000)
2. Barnes, J., Poor, G.M., Leventhal, L., Zimmerman, G., Klopfer, D.: Look and touch: the impact of touchscreens on the delivery of instructions for inherently 3D construction tasks using web-delivered virtual reality. In: Proceedings of IPSI 2005, Amsterdam (2005)
3. Bethell-Fox, C.E., Shepard, R.N.: Mental rotation: Effects of Stimulus complexity and familiarity. Journal of Experimental Psychology: Human Perception & Performance 14, 12–23 (1988)

4. Bouri, D., Fraser, M., Fraser, D.S., Cater, K.: Augmenting Spatial Skills with Mobile Devices. In: CHI 2012, Austin, TX, May 5-10, pp. 1611–1620 (2012)
5. Braine, M.D.S.: On the Relation Between the Natural Logic of Reasoning and Standard Logic. Psychological Review 85, 1–21 (1978)
6. Dunlop, M.D., Brewster, S.A.: The challenge of mobile devices for human computer interaction. Personal and Ubiquitous Computing 6(4), 235–236 (2002)
7. Eisenberg, M., Nishioka, A., Schreiner, M.E.: Helping users think in three dimensions: steps toward incorporating spatial cognition in user modeling. In: Proceedings of the 2nd International Conference on Intelligent User Interfaces, IUI 1997, pp. 113–120 (1997)
8. Ekstrom, R.B., French, J.W., Harman, H.H., Deman, D.: Kit of Factor Referenced Cognitive Tests. Educational Testing Services, Princeton, NJ (1976)
9. Froehlich, P., Baillie, L., Simon, R.: Realizing the Vision of Mobile Spatial Interaction. Interactions, 15–18 (January–February 2008)
10. Folk, M.D., Luce, R.D.: Effects of stimulus complexity on mental rotation rate of polygons. Journal of Experimental Psychology: Human Perception & Performance 13, 395–404 (1987)
11. Instone, K., Brown, E., Leventhal, L., Teasley, B.: The challenge of effectively integrating graphics into Hypertext. In: Bass, L.J., Unger, C., Gornostaev, J. (eds.) EWHCI 1993. LNCS, vol. 753, pp. 290–297. Springer, Heidelberg (1993)
12. Jacob, R.J., Girouard, A., Hirschfield, L.M., Horn, M.S., Shaier, O., Soloway, E.T., Zigelbaum, J.: Reality-based interaction: A framework for post-WIMP interfaces. In: Proceedings of ACM Conference on Human Factors in Computing Systems, SIGCHI 2009, Boston, MA, April 4-9, pp. 201–210. ACM Press, New York (2009)
13. Jolicoeur, P.: Orientation congruency effects on the identification of disoriented shapes. Journal of Experimental Psychology: Human Perception and Performance 16(2), 351–364 (1990)
14. Just, M.A., Carpenter, P.A.: Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review 92, 137–172 (1985)
15. Klopfer, D., Athy, J., Leventhal, L.: Working Memory: Just & Carpenter (1985) Revisited. In: Proceedings of the 48th Annual Meeting of the Psychonomic Society, Long Beach, California, November 15-19 (2007)
16. Lohman, D.F.: Spatial abilities as traits, processes, and knowledge. In: Sternberg, R.J. (ed.) Advances in the Psychology of Human Intelligence, vol. 40, pp. 181–248. Erlbaum, Hillsdale (1988)
17. Lohman, D.F.: Spatial ability and G. In: Dennis, I., Tapsfield, P. (eds.) Human Abilities: Their Nature and Assessment, pp. 97–116. Erlbaum, Hillsdale (1996)
18. Mayer, R.E.: Multimedia Learning. Cambridge University Press, New York (2001)
19. Miyake, A., Friedman, N.P., Rettinger, D.A., Shah, P., Hegarty, M.: How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. Journal of Experimental Psychology: General 130, 621–640 (2001)
20. Mohamedally, D., Zaphiris, P., Petrie, H.: Recent research in mobile computing: a review and taxonomy of HCI issues. In: Proceedings of HCI International 2003, pp. 9–10 (2003)
21. Paivio, A.: Mental representations: A dual coding approach. Oxford University Press, New York (1986)
22. Palmer, S.E.: Vision science: Photons to phenomenology. MIT Press, Cambridge (1999)
23. Poor, G.M.: The effects of varying levels of reality based interaction on a subjects' ability to perform a 3D construction task. PhD dissertation. Tufts University (2008)

24. Seager, W., Stanton-Fraser, D.: Comparing Physical, Automatic and Manual Map Rotation for Pedestrian Navigation. In: ACM SIGCHI Conference on Computer-Human Interaction (CHI 2007), pp. 767–776. ACM Press (May 2007)
25. Shepard, R.N.: The mental image. American Psychologist 33, 125–137 (1978)
26. Shepard, R.N., Hurwitz, S.: Upward direction, mental rotation, and the discrimination of left and right turns in maps. Cognition 18, 161–193 (1984)
27. Shiffrar, M., Shepard, R.N.: Comparison of cube rotations around axes inclined relative to the environment or to the cube. Journal of Experimental Psychology: Human Perception and Performance 17, 44–54 (1991)
28. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative effectiveness of augmented reality in object assembly. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 73–80 (2003)
29. Teasley, B., Leventhal, L., Instone, K., Brown, E.: Effective illustrations in interactive media: what works? In: Human-Computer Interaction, INTERACT 1997, pp. 197–204 (1997)
30. Zimmerman, G., Barnes, J., Leventhal, L.: Delivering instructions for inherently-3D construction tasks: an evaluation of authoring environments for multimedia presentations. In: Proceedings of the 2000 International Computer Symposium, pp. 126–131 (2000)
31. Zimmerman, G.W., Klopfer, D., Poor, G.M., Barnes, J., Leventhal, L., Jaffee, S.D.: "How do I line up?": Reducing mental transformations to improve performance. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part I, HCII 2011. LNCS, vol. 6761, pp. 432–440. Springer, Heidelberg (2011)
32. Zimmerman, G., Barnes, J., Leventhal, L.: A comparison of the usability and effectiveness of web-based delivery of instructions for inherently-3D construction tasks on handheld and desktop computers. In: Proceeding of the Eighth International Conference on 3D Web Technology, Web3D 2003, pp. 49–54 (2003)

Ambient Timer – Unobtrusively Reminding Users of Upcoming Tasks with Ambient Light

Heiko Müller[1], Anastasia Kazakova[2], Martin Pielot[3], Wilko Heuten[1], and Susanne Boll[2]

[1] OFFIS – Institute for Information Technology, Oldenburg, Germany
{firstname.lastname}@offis.de
[2] University of Oldenburg, Germany
{firstname.lastname}@uni-oldenburg.de
[3] Telefonica Research, Barcelona, Spain
pielot@tid.es

Abstract. Daily office work is often a mix of concentrated desktop work and scheduled meetings and appointments. However, constantly checking the clock and alarming popups interrupt the flow of creative work as they require the user's focused attention. We present *Ambient Timer*, an ambient light display designed to unobtrusively remind users of upcoming events. The light display - mounted around the monitor - is designed to slowly catch the user's attention and raise awareness for an upcoming event while not distracting her from the primary creative task such as writing a paper. Our experiment compared established reminder techniques such as checking the clock or using popups against *Ambient Timer* in two different designs. One of these designs produced a reminder in which the participants felt well informed on the progress of time and experienced a better "flow" of work than with traditional reminders.

Keywords: Ambient Light Display, Reminder, Interruptions, User Studies.

1 Introduction

In many of today's office jobs, we are expected to fulfil tasks that require concentration, creativity, and time. Yet, in many professions the day of an office worker is highly fragmented [15]. We arrive at 9 for a short briefing, have a meeting from 10-11, and leave at 12 to meet a colleague for lunch. Thus, we have to squeeze other work tasks into the free time slots and handle them one step at a time. In order to handle these tasks well and efficiently we need to focus our attention on them. Preferably we enter the state of *flow*, which means that we become fully immersed and focused on the task at hand [23].

However, when fully concentrated, we are less likely to pay attention to upcoming meetings and appointments [3,4]. We have observed two strategies to approach this problem: First, people that use electronic calendars have their devices create an alarm once the next meeting approaches [5]. This, however, can disrupt suddenly and unpleasantly and leave workers with unfinished tasks, and a mental state that is not

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 211–228, 2013.

Fig. 1. Ambient Timer

ready to focus on the new topic of the upcoming meeting [1]. Second, many people keep watching the clock if a meeting is approaching soon. In this case, however, constantly switching between the primary task and watching the clock may be preventing the worker from focussing on the task at hand. In other cases, workers forget to check the time regularly and consequently do not show up in time for meetings.

What is missing is a solution that conveys the remaining time until the next meeting to workers, which allows them to focus on their primary task and enter a state of flow at the same time. Instead of sudden interruptions in form of alarms, we need a cue that gradually increases its saliency, moving slowly from subconscious into conscious perception, so that once the meeting arrives; workers have brought their primary task to a stable state and are ready to focus on the new topics.

We propose the concept of an *Ambient Timer*, which presents the time remaining for the current tasks in a way that slowly shifts from ambient to focal attention. As information display modality, we investigate the use of ambient light, which gradually increases obtrusiveness as the time for the current task runs out (Fig. 1). We envision the following usage scenario:

Alice is working on a report for the head of her department. At the same time, there is a meeting scheduled in thirty minutes, which she has to attend. The Ambient Timer is already emitting light in a low-attention state, so Alice feels confident that she will be reminded of the meeting. A few minutes before the meeting, the status of the ambient light display has changed to a more salient, intense output. While she is still working on her report, she slowly becomes aware of the nearing deadline and starts finishing the paragraph she is currently working on. One minute before the meeting *the light has become so salient that it is hard to ignore. Alice stores the document on the server, puts her computer into sleep mode, and arrives at the meeting on time.*

In this paper, we investigate how to gradually increase the saliency of an ambient light display, so that it slowly catches the attention of a person that is focusing on another task. We report from a study in which we compared the *Ambient Timer* to keeping track of time with a clock and to using a popup alarm in a typical office task. The results provide evidence that Ambient Timer gives people confidence in being able to see the progress of time, make them experience fewer interruptions, and make it subjectively easier for them to enter the state of flow.

2 Related Work

Ambient information displays received greater attention ever since Weiser introduced his work on "calm technology". He pointed out that calm technology engages and switches back and forth between periphery and centre of attention [30]. Ishii et al. provided an insight into numerous possibilities for ambient displays in a work environment when creating their "ambientROOM" [13]. Pousman and Stasko elaborated numerous definitions of ambient displays [24]. In our understanding ambient information displays - such as the Ambient Timer - provide information which is important but not critical. They make a user aware of the information but do not require her to focus on the information display.

"Human interruption", is defined as "the process of coordinating abrupt changes in people's activities" [19]. Interruptions may have internal and external sources. An internal interruption may occur when e.g. throughout the process of writing a scientific paper new demand for literature arises. Thus the person will interrupt her work in order to find the required books. Examples for external interruptions are notification systems. McCrickard and Chewar describe notification systems as "interfaces specifically designed to support user access to additional digital information from sources secondary to current activities" [17]. While notification systems may often be distracting, this behaviour is usually tolerated by users [18]. Fogarty et al. report on the development of a sensor-based system to better automatically predict interruptability of human users [9]. Ambient Timer aims to notify its users on the progression of time without distracting them from their primary task.

In their work, Matthews et al. define multiple attention levels: inattention, divided attention and focused attention [16]. Inattention defines a state, in which information is not consciously observed, while divided attention means that information is consciously perceived but does not require the fully focused attention of the user yet. The existing reminding techniques clock and popup alarm both require focused attention, as one would either have to read and understand the digits or hands of the clock, or close the popup window in order to return to the task at hand. We aim to have the *Ambient Timer* operate along inattention and divided attention transitioning from the first to the latter as time progresses. The orienting reflex as described by Sokolov [28] poses limitations on the possible design solutions. Müller et al. report on a study exploring the design space of an ambient reminding system [22].

Mankoff et al. report an ambient system, informing users on the approach of local busses at a faculty building [14]. While displaying information on an upcoming event (the arrival of the bus) the reported system does not increase obtrusiveness to make users aware of when the bus is near.

Dragicevic and Huot introduced the idea of continuously displaying the time until an event using an analogue clock metaphor on the computer screen [6]. This design aims at displaying information on time in an unobtrusive way. However, the information is displayed onscreen and not in the periphery.

Meyer et al. created a system for reminding elderly users of appointments and household events, utilising ambient light [21]. Their system however is more of an "ambient popup alarm", as it gives no information on the progression of time nor on the time remaining until the appointment starts.

Vastenburg et al. reported their work on the acceptability of notifications in home situations [29]. They found a correlation between user acceptability and intrusiveness and importance of the displayed message.

The above mentioned systems all address various aspects of ambient displays and conveying temporal information. However they do not address issues of gradual change in importance of temporal information and the need to notify the user on these changes without disrupting her.

On displaying information with increasing urgency in the periphery Birnholtz et al. report on a projected peripheral vision display [2]. Their results indicate that peripheral vision can be used to direct the user's attention. However, this system does not utilize an ambient light display, but rather projects graphical information.

What is missing is evidence that ambient light interfaces placed in the periphery allow to gradually shift the user's attention from inattention to focussed attention, without becoming disruptive.

In the following we report our experiment. We begin with a description of the apparatus.

3 Ambient Timer

As *Ambient Timer* we envision an ambient light display that notifies users unobtrusively of the progression of time in order to know when to finish the current task to be ready in time for the next appointment. *Ambient Timer* illuminates the periphery of a monitor, on which a user focuses her visual attention when working on the computer.

Focal and peripheral vision may be regarded as two separate perception channels [12]. As peripheral vision is more sensitive closer to the foveal point [20], we decided to place our ambient light display as close to the monitor as possible (Fig. 2). Placing the information in the periphery will allow the user keeping all monitor real estate for information that might not be suitable for display in the periphery.

To gradually grab the user's attention along the attention levels defined by Matthews et al. [16], the output of the *Ambient Timer* needs to gradually become more salient. In order to increase the saliency of the display to gain attention over time two forms of transition appear to be feasible: First, a sine wave change where the display alternates back and forth between two states with increasingly faster cycles. Second, we use an exponential change from the initial to the final state. Other forms of transition were ruled out after studying the related literature, e.g. linear change, which may not provide enough increased obtrusiveness due to human deficiencies in noticing slow changes [25], or cubic or saw tooth changes between initial and target state due to excessive obtrusiveness [28].

Müller et al. [22] explored the design space for displaying gradual changes with ambient light. Their results suggest that colour changes may be superior to change in brightness or saturation. While a sinusoid change may not require a sharp contrast from start to target colour, exponential changes - which appear to be more unobtrusive by design - require a sharp contrast of complementary colours such as green to red. As reported, the traffic light analogy was found helpful by participants. With regard to these results, we chose to introduce a sinusoid colour change with increasing frequency from green to orange as well as an exponential change from green to red as conditions for *Ambient Timer*, thus evaluating the patterns from each category that achieved best results overall against state-of-the-art reminding techniques.

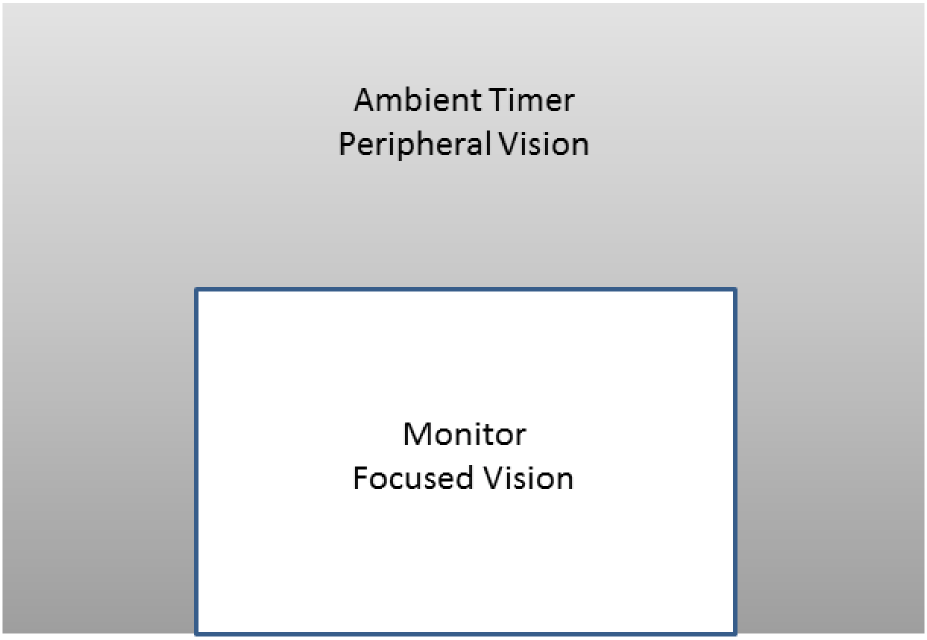


Fig. 2. Concept of Ambient Timer

Our apparatus is made of uniformly controlled LED-stripes which were fixed to a three-sided frame that had the same width and height as a 22 inch monitor used for our studies (Fig. 3). This frame is mounted on the back of the monitor (Fig. 4). We placed frame and monitor on a desk close to a matt white wall. The gap between monitor and wall was about ten centimetres. This provided us with a surface to reflect

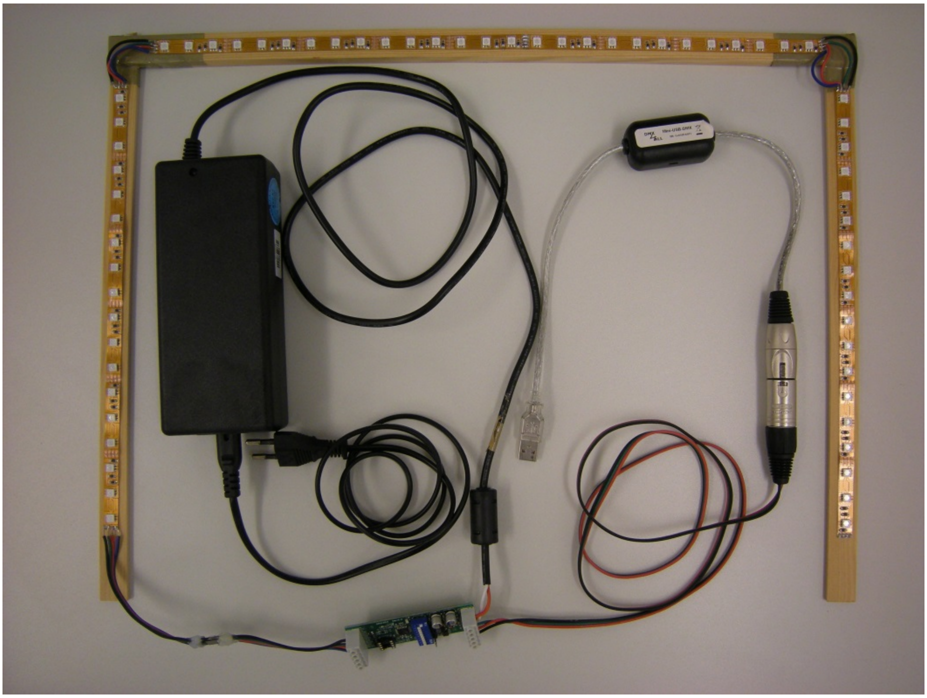


Fig. 3. LED-Frame

the light emitted by the LEDs and illuminate the immediate periphery of the monitor. While the user focuses her view on the monitor, she can perceive the ambient light display in her peripheral vision.

4 Methodology

In our experiment, we investigated if the *Ambient Timer* allows workers to work more concentrated and efficiently than when using state of the art reminders. Our general approach was to expose participants to a demanding primary task and evaluate the effectiveness of the *Ambient Timer* as reminding technique compared to state of the art baselines. As a primary task, we asked participants to copy (type) given texts and correct mistakes we placed in the original documents. Copying and correcting texts is a task requiring good concentration. Thus the impact of interruptions would be serious [10].

As secondary task, we asked the participants to be aware of the time and finish the primary task within ten minutes. We defined finishing as finalizing to copy and correct a full sentence once the participants thought that available time was running out.

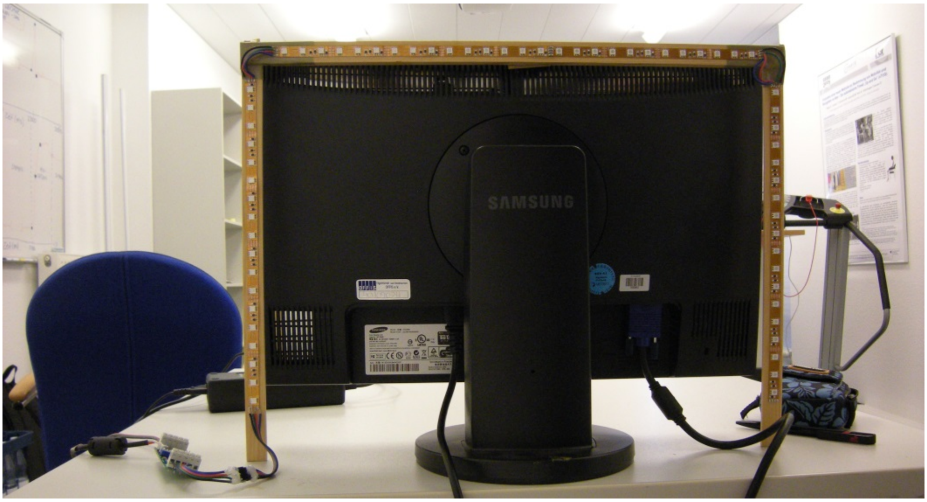


Fig. 4. LED-Frame mounted on the backside of the monitor

Participants were instructed to inform the experimenter, when they thought that they had finished the primary task. Participants were discouraged from exceeding the time limit (overshoot), i.e. to take longer than ten minutes. If participants missed the end of the time period by more than 30 seconds a popup window informed them of the end of the trial. For staying aware about the remaining time, we experimentally compared two implementations of the Ambient Timer with using a clock, i.e. participants had to monitor the progress of time on a system clock with digits, and pop-ups, i.e. the only information given was a pop-up that was shown two minutes before the time expired. The two implementations of *Ambient Timer* were:

AT Expo: remaining time was conveyed via an exponential colour change from green to red.

AT Sinus: remaining time was conveyed via a sinusoid colour change between green and orange with increasingly faster cycles.

Our hypotheses were that:

H1: With the *Ambient Timer*, participants will be or feel less interrupted than with the clock.

H2: Participants will find/correct more and make less new mistakes when using the *Ambient Timer*.

H3: Participants feel confident and well informed on the progression of time when using *Ambient Timer*.

H4: Participants will find *Ambient Timer* as the favourable reminding technique.

4.1 Design

Type of Reminder, i.e. the way that participants were informed about the time that remained to complete the current task, served as independent variable. We compared four different techniques:

- In the *Popup* condition, the participants had no information about the remaining time. Instead a popup window appeared to inform them about the end of the task two minutes before the time expired. This condition reflects the common strategy of not checking the clock and only relying on popups by the calendar system to be reminded of appointments.
- In the *Clock* condition, we showed the system clock in the task bar at the bottom of the screen. The clock showed the current time with hours and minutes. No seconds were shown. By knowing when the task had to finish, the participants could calculate how much time is left. This condition reflects the other common strategy of keeping an eye on the clock when a meeting is approaching.
- In the *AT Expo* condition, the *Ambient Timer* was used to convey the remaining time via an exponential colour change from green to red.
- In the *AT Sinus* condition, the *Ambient Timer* was used to convey the remaining time via a sinusoid colour change between green and orange. With this condition, we aimed at testing a more obtrusive pattern, in case that the exponential change proves to be too unobtrusive when the user focusses on the primary task.

Except for in the *Clock* condition, no clock was available to the participants. During the experiment, the order of these conditions was counter-balanced to cancel out sequence effects. In order to assess the unobtrusiveness of these designs, and answer the hypotheses, we logged the following measures as dependent variables:

- *Interruptions*: The number of interruptions, i.e. any occurrence of a pause between keystrokes of more than one second.
- *Keystrokes*: the total number of keystrokes per task. We measured *keystrokes* to gain insights into how well participants were able to focus on the copying task in the various conditions.
- *Keystroke Time*: the average time between two keystrokes. This measure was taken to gain insight into the ratio of total keystrokes and interruptions. As not all participants stopped exactly at the ten minute mark and as the obtrusiveness of the displays was not evenly distributed across time we did not simply normalize our measures.
- *Corrected Mistakes*: the number of grammar and spelling mistakes corrected per task. We measured *Corrected Mistakes* to quantify how concentrated our participants can work with the compared reminding techniques.
- *Newly Introduced Mistakes*: the number of grammar and spelling mistakes that are newly introduced by the participant per task. *Newly Introduced Mistakes* was also measured to estimate the participants' level of concentration and flow.
- *Overshoots*: a Boolean value indicating whether the participant did not finish the task on time. If the participant continued to work on the task for more than 30 seconds after the time ran out, we interrupted the task and counted this task as an over shoot. With this measure, we aimed at quantifying how well the different reminders helped the participants to stay aware of the remaining time.

In addition, for each of the conditions, we collected the participants' agreement (5-point Likert scale) to the following statements:

- *I think this technique is good.*
- *I felt distracted by the system.*
- *It was easy to monitor the progression of time.*
- *I felt sure about noticing the progression of time.*
- *I was able to complete the task in time.*
- *I entered the state of flow.*

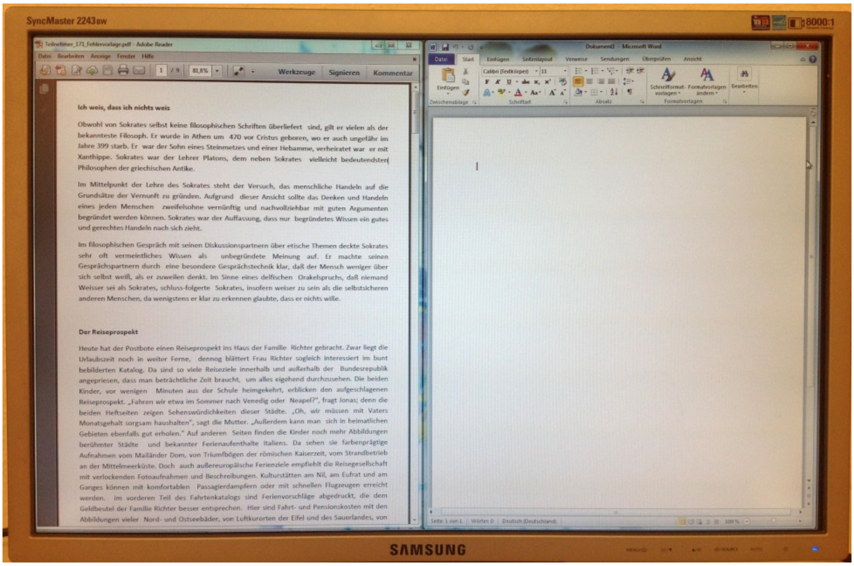


Fig. 5. Screen to copy and correct text

4.2 Apparatus

The texts to correct were taken from a collection of dictations for teachers working with tenth graders [27,7,26]. Each dictation had a special focus on certain words which we used as a guideline for placing mistakes into the text. For the copy and correct task, the screen was divided into two parallel text-fields, one containing the text, one used to copy the text into.

As work place, we used a standard desktop computer running Windows 7 with a 22-inch monitor on a standard working table. As shown in Fig. 5, the participants were provided with a text field to copy the text into. In the *Clock* Condition, we showed the current time, as shown in Fig. 6.

We conducted our studies in a room with controlled light setting thus eliminating possible effects of changing surrounding light [8]. We kept the light level constant at 420 lux at the desktop which is in compliance with rules on office workspace settings.

For tracking interruptions, we used a key logger script that measured the time between individual keystrokes.

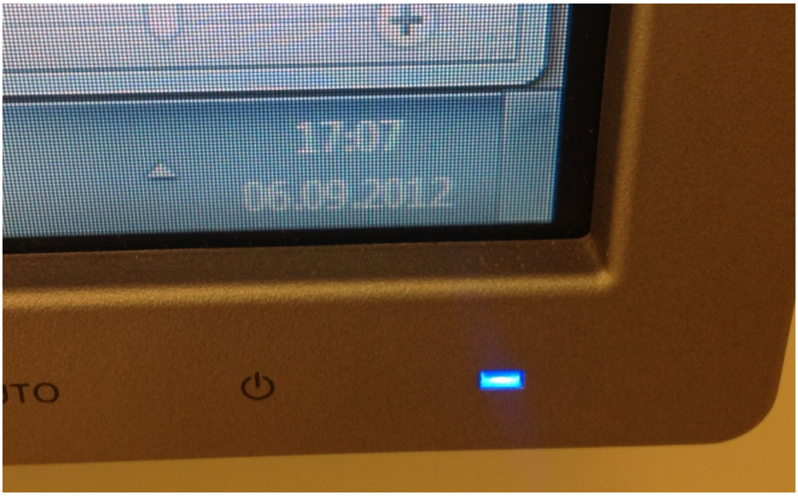


Fig. 6. On-screen system clock

4.3 Participants

12 participants (4 female, 8 male) aged 17-45 *(M = 28.3, SD = 8.8)* took part in the experiment. None of the participants reported a case of colour blindness.

All participants were experienced in writing texts on computers. All participants rated their typing speed between fast and medium, two of them used 10-finger-typing.

Asked about their method of reminding themselves of appointments, participants either used a calendar with alarms on their computer (6) and/or on their phone (5) as well as regularly checked the clock (4). When using reminders, the lead time to appointments was stated to be any time from five to 60 minutes, or even one to two days in advance for full day events. All but one participant answered that they "sometimes" to "never" missed appointments. One participant said that missing appointments was a common occurrence. Participants received no compensation for their participation.

4.4 Procedure

Before the start of the experiment, participants were introduced to the scope of the study and familiarized with both light designs of the Ambient Timer. We then conducted four trials per participant, exposing them to the four conditions in randomized order.

After each trial, participants were asked to rate their agreement to the five statements regarding the used reminding technique. Upon completion of all four trials, we collected the participants' impressions in an open interview.

5 Results

5.1 Objective Measures

Figure 7 shows the descriptive statistical summary for the objective measures. We used repeated measures one-way ANOVA and Tukey HSD to test for significant effects.

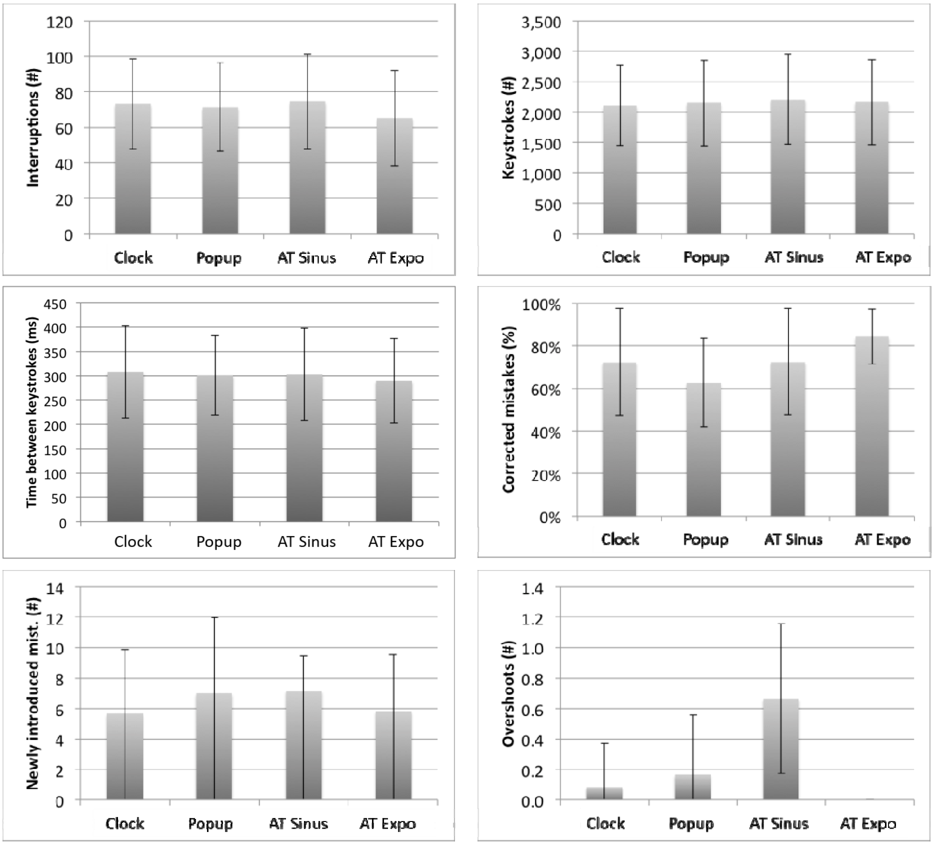


Fig. 6. Descriptive summary of objective measures (Bars show mean, error bars indicate standard deviation)

There was a significant effect on *Overshoots* $(F(3, 44) = 9.0, p < .001)$. The number of *Overshoots* was significantly higher in the *AT Sinus* condition, compared to the Popup ($p < .001$), *Clock* ($p < .001$), and *AT Expo* ($p < .001$) conditions. Thus, participants had more difficulties to finish their tasks on time in the *AT Sinus* condition; participants did not finish their tasks on time as often as in the other tasks.

There were no significant effects on *Interruptions* $(F(3,44) = 0.29, p = .83)$, *Keystrokes* $(F(3,44) = .04, p = .99)$, *Keystroke Time* $(F(3,44) = .09, p = .97)$, *Corrected*

Mistakes (F(3,44) = 2.03, p = .12), and *Newly Introduced Mistakes (F(3,44) = .47, p = .7)*. Hence, we cannot make any assumptions about the effect of *Type of Reminder* on the remaining dependent measures.

5.2 Subjective Measures

Fig. 8 shows the Median ratings of the participants' level of agreement with the six Likert-scale statements per condition. Assuming normally-distributed, interval scores, we analysed the data for significant effects by using repeated measures one-way ANOVA and post-hoc Tukey HSD tests.

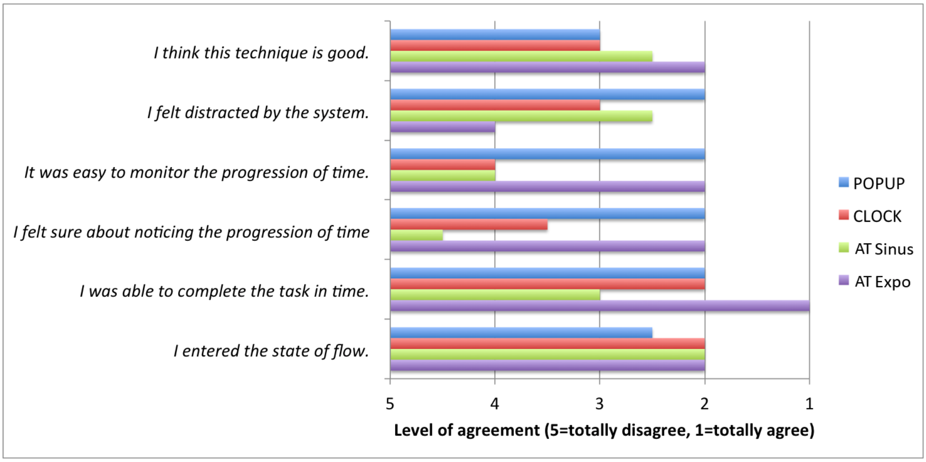


Fig. 8. Median level of agreement with the six statements per condition

There was a significant effect on the agreement to the statement *I think this technique is good (F(3, 44) = 4.5, p < .01)*. The level of agreement in the *AT Expo* condition was significantly higher than in the *Popup (p < .05)* and in the *Clock (p < .05)* conditions. Thus, participants found *AT Expo* to be a better reminding technique than *Popup* and *Clock*.

There was a significant effect on the agreement to the statement *I felt distracted by the system (F(3, 44) = 4.8, p < .01)*. The level of agreement in the *AT Expo* condition was significantly lower than in the *Popup (p < .05)*, *Clock (p < .05)*, and *AT Sinus (p = .06)* conditions. Thus, participants found *AT Expo* to be less distracting than the other reminding techniques.

There was a significant effect on the agreement to the statement *It was easy to monitor the progression of time (F(3, 44) = 8.7, p < .001)*. The level of agreement in the *Clock* condition was significantly higher than in the *Popup (p < .01)* and *AT Sinus (p < .01)* conditions. Likewise, the level of agreement in the *AT Expo* condition was significantly higher than in the *Popup (p < .001)* and *AT Sinus (p < .001)* conditions. Thus, in the *Clock* and *AT Expo* conditions, participants found it easier to monitor the progress of time than in the *Popup* and *AT Sinus* conditions.

There was a significant effect on the agreement to the statement *I felt sure about noticing the progression of time ($F(3, 44) = 14.0$, $p < .001$)*. The results are analogue to the previous statement. The level of agreement in the *Clock* condition was significantly higher than in the *Popup* ($p < .001$) and *AT Sinus* ($p < .001$) conditions. Likewise, the level of agreement in the *AT Expo* condition was significantly higher than in the *Popup* ($p < .001$) and *AT Sinus* ($p < .001$) conditions. Thus, in the *Clock* and *AT Expo* conditions, participants were more confident that they would notice the progress of time than in the *Popup* and *AT Sinus* conditions.

There was a significant effect on the agreement to the statement *I was able to complete the task in time ($F(3, 44) = 7.3$, $p < .001$)*. The level of agreement in the *AT Expo* condition was significantly higher than in the *Popup* ($p < .05$) and *AT Sinus* ($p < .05$) conditions. Thus, participants found it easier to complete the task in time in the *AT Expo* condition compared to the *Popup* and the *AT Sinus* conditions.

We could not find a significant effect on the agreement to the statement *I entered the state of flow* ($F(3, 44) = 2.6$, $p = .06$, marginally significant).

In summary, the clock and the *Ambient Timer* in the *AT Expo* condition were found to make it easy to monitor the progress of time and finish the task in time. In addition, the *Ambient Timer* in the *AT Expo* condition was found to be the better reminding technique and was rated to be least distracting.

5.3 Comments and Observations

For *Clock*, six participants answered that as keeping an eye on the clock was very common, it should not pose a problem. They felt it was reliable and that they were "in control". On the other hand ten participants answered when asked what had distracted them that switching focus from text to clock and back posed some problems in finding the correct line in the text again.

Concerning *Popup*, five participants were unhappy with the way the popup interrupted their work. Three participants commented that they liked how the system was hidden in the background until it popped up and that they could work relaxed until then.

On *AT Expo*, one participant said it was difficult to see the end of time, while another argued that it was like a clock only better. The calm exponential change was well accepted.

AT Sinus was found to be obtrusive by six participants. They felt unsure on how much time had already passed and said that they were somewhat distracted by the constant change of colours. While two participants said that the colour change was too subtle to notice without looking, one participant answered that he liked the subtle changes best. Three participants mentioned that they might be able to get used to *AT Sinus* but that it needed more time to familiarize.

Ten participants said that they can imagine using the *Ambient Timer*, while two answered that they would not want to use it. Of the ten positive answers, one participant would use the *Ambient Timer* stand-alone, while the others preferred a combination of techniques (+ Popup (6), + Clock (3)) for better control of when the time ran out.

Concerning acceptability of an ambient light display placed at the user's desks 11 participants considered using the *Ambient Timer* in a work environment as being unproblematic. Out of the positive answers, some participants argued that the system could signal other colleagues that the user is about to leave for a meeting and that a discussion should be postponed to a later time, thus adding extra value to the information of an upcoming task. However, when engaged in group meetings or talks with customers or other outsiders, participants would like an option to temporarily disable the system as they felt the use of such a system inappropriate in these situations. Only one participant could imagine using *Ambient Timer* solely in a private setting.

Asked about the display time of *Ambient Timer*, most participants would prefer a lead time of 10-15 minutes to be able to bring their current work into a stable state before having to leave for the appointment.

We asked participants if they could think of other light patterns or colours to use in *Ambient Timer*. Five participants liked the *AT Expo* the way it was, especially as they liked the green = "go", red = "stop" traffic light analogy. A couple users suggested using blue as a starting colour, arguing that blue was a calming colour. Two participants suggested an exponential pattern with a flashing part at the end of the time period so that it would be easier to see.

5.4 Discussion

The experiment compared two designs of *Ambient Timer*, one with an exponential change of colour from green to red (*AT Expo*), and one with a sinusoidal change between green and orange with increasingly faster cycles (*AT Sinus*), a clock, and pop ups as means to monitor the remaining time. Our results show that participants experience fewer interruptions when using *Ambient Timer* with an exponential change from green to red, compared to all other reminder techniques in our experiment. Their average typing speed was faster when in this condition. Participants ranked this design best, felt most confident using it and preferred it over all other techniques.

Discussing results in relation to our hypotheses, we find that we have a split between the two *Ambient Timer* designs. While the exponential change from green to red performed very well, users experienced no benefits over state-of-the-art reminding techniques from the sinusoid change between green and orange. On the other hand it is worth to mention, that the sinusoid design was by no means a dropout.

H1: Participants will experience fewer interruptions when using the Ambient Timer

No significant effects could be found in a one-way ANOVA test of key-logging data to support this hypothesis. However, our participants' answers to the question of being in the "flow" received marginally significant better results for the exponential *Ambient Timer* design than when participants were in the *Clock* condition. *Ambient Timer*'s idea of reminding users without interrupting them works well in the exponential *Ambient Timer* design.

H2: Participants will make fewer mistakes when using the Ambient Timer

Even though participants found more mistakes we had prepared in the texts when using both *Ambient Timer* designs, we could not reveal any statistical significance for these findings from the key logging data. Hence, there is no evidence to support this hypothesis.

H3: Participants feel confident and well informed on the progression of time when using Ambient Timer

This hypothesis can be confirmed for the *Ambient Timer* with the exponential design and falsified for *Ambient Timer* in the sinusoid design. The exponential design scored roughly the same results as when participants used the clock, while state-of-the-art reminding technique *Popup* and the *Ambient Timer* with the sinusoid design scored results worse than average. The data revealed significantly better results for the exponential *Ambient Timer* design compared to state-of-the-art technique *Popup* and the sinusoid *Ambient Timer* design, as well as significantly better results for *Clock* compared to *Popup* and the *Ambient Timer* with the sinusoid design. We have no statistical evidence that the *Ambient Timer* with the exponential design performed better than *Clock*. Considering the significantly better results of exponential *Ambient Timer* design over *Clock* and *Popup* concerning the statement "This pattern is good" we conclude that *Ambient Timer* when in the exponential design makes a good alternative to state-of-the-art reminding techniques.

H4: Participants will find Ambient Timer as the favourable reminding technique

This can be verified for *Ambient Timer* in the exponential design and falsified for the sinusoid *Ambient Timer* design. The exponential *Ambient Timer* design ranked best together with *Clock*, while the sinusoid design ranked last.

Overall, H1 (less interruptions) was supported for the exponential *Ambient Timer* design and *Clock*, H2 (less mistakes) was not supported, and H3 (well informed) and H4 (favourable) were supported for *Ambient Timer* in the exponential design but not for the sinusoid design.

We have recorded split results suggesting that the benefits of using ambient light for reminding users of upcoming events depends on how the design increases its saliency to grab the user's attention. In our experiment, the exponential *Ambient Timer* design receives better scores on all questions we have asked our participants and is ranked best in the order of participants' preference. Participants' feedback in our questionnaires and interviews suggest that they like the way *Ambient Timer* in the exponential design keeps them updated on the time progress without them having to take their eyes off the primary task or being interrupted unexpectedly. With the exponential *Ambient Timer* design we have introduced a true alternative to the state-of-the-art reminding techniques *Clock* and *Popup*.

Concerning acceptability, the majority of participants had no concerns about the visibility of the system in an office environment. They saw benefits such as co-workers being more aware of the individual schedules. However, some would prefer to "mute" the system in group-work situations and when meeting with customers.

One of the limitations of our experiment is the type of setting. Evaluating ambient displays in a lab condition is always difficult, as the expected benefit of blending into the environment can hardly be achieved in a lab setting [11]. When using the *Ambient Timer*, participants were using a system that was already "on", which it would not normally be in a regular office setting. Therefore we cannot make any statements on how users will react to a "warm-up/turn-on"-phase. While we aimed at creating a challenging task that could be solved by the participants, we cannot simulate tasks that are truly relevant for our participants, which would then have enabled them even further to experience flow and immerse themselves in the tasks.

6 Conclusion

We presented *Ambient Timer*, a way of unobtrusively reminding users of upcoming events and appointments using ambient light. In an experiment, we compared the exponential change of colour from green to red, a sinusoidal change between green and orange with increasingly faster cycles, a clock, and pop ups as means to monitor the remaining time. The results show that ambient light can successfully convey remaining time. In particular, the exponential change from green to red was preferred over all other approaches. The participants found it to be a non-distracting way of monitoring the remaining time, in order to know when to wrap up the primary task.

The main contribution of the work is to provide evidence that ambient light displays are a good solution for extending the information space in an office environment by displaying information outside the user's monitor. Ambient light displays can help adding important information (such as reminders on upcoming tasks) without adding clutter or interruptions to the user's workflow by requiring focused vision. In a work environment that is increasingly exposed to interruptions, such approaches may help information workers to structure their days and focus on their primary tasks.

Future directions for this work will be a long-term test in an office environment evaluating ambient light "in the wild", giving participants a chance to evaluate the system against their usually used reminder technique over the course of a week. Participants will design their “own” light patterns, thus avoiding possible pitfalls such as colour vision deficiencies. This would give insights into how it blends into the periphery and becomes noticeable when appointments are due. Further we are reworking the prototype to make it more flexible in use by adding a semi-transparent cover. This will allow users, who do not have their computer monitor against a wall, to be able to see the light display.

Another possibility for further research is on novel light patterns for, using both uniform illumination of all LEDs as well as patterns controlling LEDs individually thus creating opportunities for evaluating "moving" patterns as well.

Acknowledgments. We would like to thank our colleagues as well as and all participants in our pilot study and the experiment for the valuable feedback on our design and all the time they invested.

References

1. Adamczyk, P.D., Bailey, B.P.: If not now, when?: the effects of interruption at dfferent moments within task execution. In: Proc. CHI 2004, pp. 271–278. ACM, New York (2004)
2. Birnholtz, J., Reynolds, L., Luxenberg, E., Gutwin, C., Mustafa, M.: Awareness beyond the desktop: exploring attention and distraction with a projected peripheral-vision display. In: Proceedings of Graphics Interface, GI 2010, pp. 55–62. Canadian Information Processing Society (2010)
3. Block, R.: Prospective and retrospective duration judgment: The role of information processing and memory. In: Time, Action and Cognition (1992)
4. Casini, L., Macar, F.: Multiple approaches to investigate the existence of an internal clock using attentional resources. Behavioural Processes 45, 73–85 (1999)
5. Czerwinski, M., Horvitz, E., Wilhite, S.: A diary study of task switching and interruptions. In: Proc. CHI 2004, pp. 175–182. ACM, New York (2004)
6. Dragicevic, P., Huot, S.: Spiraclock: a continuous and non-intrusive display for upcoming events. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2002, pp. 604–605. ACM, New York (2002)
7. Fehlbusch, C.: 150 Diktate, 5. bis 10. Klasse: Regeln und Texte zum Üben. Dudenverl, Mannheim (2006)
8. Ferwerda, J.A., Pattanaik, S.N., Shirley, P., Greenberg, D.P.: A model of visual adaptation for realistic image synthesis. In: Proc. 23rd Annual Conference on Computer Graphics and Interactive Techniques (1996)
9. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., Yang, J.: Predicting human interruptibility with sensors. ACM Trans. Comput.-Hum. Interact. 12(1), 119–146 (2005)
10. Gillie, T., Broadbent, D.: What makes interruptions disruptive? a study of length, similarity, and complexity. Psychological Research 50, 243–250 (1989)
11. Hazlewood, W.R., Stolterman, E., Connelly, K.: Issues in evaluating ambient displays in the wild: two case studies. In: Proc. CHI 2011. ACM, New York (2011)
12. Horrey, W.J., Wickens, C.D., Consalus, K.P.: Modeling drivers visual attention allocation while interacting with in-vehicle technologies. Journal of Experimental Psychology 12, 67–78 (2006)
13. Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ulmer, B., Yarin, P.: ambientroom: Integrating ambient media with architectural space. In: Proc. CHI 1998, pp. 173–174 (1998)
14. Mankoff, J., Dey, A., Hsieh, G., Kientz, J., Lederer, S., Ames, M.: Heuristic evaluation of ambient displays. In: Proc. CHI 2003, pp. 169–176. ACM, New York (2003)
15. Mark, G., Gonzalez, V.M., Harris, J.: No task left behind?: examining the nature of fragmented work. In: Proc. CHI 2005, pp. 321–330. ACM, New York (2005)
16. Matthews, T., Dey, A.K., Mankoff, J., Carter, S., Rattenbury, T.: A toolkit for managing user attention in peripheral displays. In: Proc. UIST, pp. 247–256 (2004)
17. McCrickard, D.S., Chewar, C.M.: Attuning notification design to user goals and attention costs. Commun. ACM 46, 67–72 (2003)
18. McCrickard, D.S., Chewar, C.M., Somervell, J.P., Ndiwalana, A.: A model for notification systems evaluation assessing user goals for multitasking activity. ACM Trans. Comput.-Hum. Interact. 10, 312–338 (2003)
19. McFarlane, D.C., Latorella, K.A.: The scope and importance of human interruption in human-computer interaction design. Hum.-Comput. Interact. 17, 1–61 (2002)

20. Mckee, S., Nakayama, K.: The detection of motion in the peripheral visual field. Vision Resources 24, 25–32 (1984)
21. Meyer, E.M., Heuten, W., Meis, M., Boll, S.: Multimodal presentation of ambient reminders for older adults. In: Proc. Ambient Assisted Living 2010 (2010)
22. Müller, H., Kazakova, A., Pielot, M., Heuten, W., Boll, S.: Unobtrusively reminding users of upcoming tasks with ambient light: Ambient timer. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, NordiCHI 2012, pp. 801–802 (October 2012)
23. Nakamura, J., Csikszentmihalyi, M.: Flow Theory and Research. In: Oxford Handbook of Positive Psychology, pp. 195–206. Oxford University Press (2009)
24. Pousman, Z., Stasko, J.: A taxonomy of ambient information systems: Four patterns of design. In: Proc. Advanced Visual Interfaces, pp. 67–74 (2006)
25. Rensink, R.: Change detection. Annu. Revs. of Psychology 53, 245–277 (2002)
26. Sämann, C.: 200 Diktate wie in der Schule: Deutsch, 5. - 10. Schuljahr. Klett-Lerntraining, Stuttgart (2009)
27. Sczyrba, K.: 100 Diktate für das 8. - 10. Schuljahr: mit Übungen zur Rechtschreibung und Grammatik. Bange, Hollfeld (2007)
28. Sokolov, E.: Higher nervous functions: The orienting reflex. Annu. Revs. of Physiology 25, 545–580 (1963)
29. Vastenburg, M., Keyson, D., de Ridder, H.: Considerate home notification systems: A user study of acceptability of notifications in a living-room laboratory. International Journal of Human-Computer Studies 67, 814–826 (2009)
30. Weiser, M., Brown, J.S.: Designing calm technology. Powergrid Journal 1 (1996)

Novel Modalities for Bimanual Scrolling on Tablet Devices

Ross McLachlan and Stephen A. Brewster

Glasgow Interactive Systems Group,
School of Computing Science, University of Glasgow, Glasgow, G12 8QQ
r.mclachlan.1@research.gla.ac.uk, stephen.brewster@glasgow.ac.uk

Abstract. This paper presents two studies investigating the use of novel modalities for bimanual vertical scrolling on tablet devices. Several bimanual interaction techniques are presented, using a combination of physical dial, touch and pressure input, which split the control of scrolling speed and scrolling direction across two hands. The new interaction techniques are compared to equivalent unimanual techniques in a controlled linear targeting task. The results suggest that participants can select targets significantly faster and with a lower subjective workload using the bimanual techniques.

Keywords: Bimanual interaction, scrolling, tablets.

1 Introduction

Touchscreen tablet devices present an interesting challenge to interaction design: they are not quite handheld like their smartphone cousins, though their form factor affords usage away from the desktop and other surfaces. This means that users will often have to dedicate one hand to holding the device, constraining their ability to use two hands on the touchscreen. This work explores the possibility of using novel input modalities mounted on the tablet (such as pressure sensors and physical dials) to enable simultaneous two-handed input while the user is holding the device.

The form factor of tablet devices requires a user to support a larger weight and navigate more screen space than a phone. Thus, while the tasks being performed may be similar, the form factor of tablet devices dictates that either two hands or a supporting surface are required for interaction. Users may interact with the device while it is perched on a desk, worktop or even while the device rests on their lap. In these cases, it is possible to interact on the screen directly with both hands, which is not easy on a small phone screen as it could create 'fat finger' problems with screen occlusion [1], [2].

In other instances, users may need to interact with the tablet while holding it in their hands. To do this, users will need to dedicate one hand to holding the device, while interacting with the other [3]. Thus, the repertoire of touch gestures is reduced. From this, it is not clear how to design interaction techniques for tablet devices that work in all of the usage scenarios afforded by their form factor.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 229–246, 2013.

Previous studies suggest there are bimanual interaction techniques that offer benefits over equivalent unimanual ones [4], though these studies have assumed static interactions in a desktop environment or have been limited to using only the touchscreen on a tablet device [3]. The challenge of designing bimanual interaction techniques for tablet devices is to allow simultaneous two-handed input while still allowing the user to hold the device comfortably. This paper presents two studies that aim to provide insight towards the design and future research of techniques that could allow such bimanual input.

1.1 Why Bimanual?

As human beings, we have natural bimanual motor skills that we have been using and perfecting our entire lives. This is not to say, however, that all two-handed action is equivalent and certainly not all tasks are performed best using two hands [5]. It cannot be said, for instance, that writing with a pen in each hand improves the efficiency of writing. Human beings have natural bimanual motor skills, but only when each hand adopts an appropriate role. Depending on the task being carried out, the two hands can cooperate symmetrically or asymmetrically. A useful and well tested characterisation of asymmetric bimanual action is Guiard's Kinematic Chain (KC) model [6]. Central to the KC Model is the cooperative and asymmetrical nature of bimanual action, meaning that when human beings perform tasks with both of their hands, they adopt different and complementary roles in order to do so. Guiard argues that the relationship between the dominant hand (DH) and non-dominant hand (NDH) is analogous to the relationship between proximal and distal elements in a kinematic chain (a series of abstract motors, a common example of which is an arm). The implication of this is that the dominant hand will act in relation to the non-dominant hand.

Tablet interaction, as it currently exists, conforms to the KC Model insofar as the user's NDH sets the frame of reference for the action of the DH by holding the device. Though, in much the same way as writing on paper - where the NDH holds the page (sets the frame of reference) for the DH to write on (the primary action of the task) - the NDH in tablet interactions primarily takes a passive role. Designing tablet interactions that offer the user's NDH a more active role in the interaction, while still properly supporting the device, has the potential to enable the user to use both hands to complete tasks in a richer way in a wider range of circumstances.

Previous work has suggested that interactions designed using the KC model can out-perform equivalent unimanual ones in a desktop environment [7]. There has also been early evidence suggesting that multitouch screen gestures that are based on human body movements that are not well documented or studied, can increase the risk of musculoskeletal disorders [4]. From this, the designers of tablet interactions could benefit from a better understanding of the ways in which human beings have evolved to use both of their hands to complete tasks. In doing so, interaction designers can take advantage of the natural abilities of human beings in order to create more effective ways to use devices.

Fig. 1. Bang & Olufsen BeoSound 5. The side-mounted dial is used to scroll the on-screen content.

Unimanual and Bimanual Scrolling. On tablet devices, unimanual scrolling involves controlling speed and direction using flick and drag gestures on the touch-screen of a multitouch device or by performing similar rotational movements (flicking and dragging) on a physical dial (such as on the Bang & Olufsen BeoSound 5 (see Figure 1). The physical dial on the Bang & Olufsen BeoSound 5 is used to scroll through a music library displayed as a circular list on the screen. This kind of scrolling behaviour is also exhibited on scroll-wheels on mice and keyboards). While this technique is straightforward to learn, it only offers very coarse control over scrolling speed and scrolling long lists can be time consuming. There are several alternative strategies users can adopt to find items in long lists such as searching or filtering the list using text input, or by jumping directly to a letter in an alphabetically ordered list (such as on Apple iOS devices), or by employing a separate fast scrolling slider (as on Google Android devices), though these techniques often require the user to know what s/he is looking for in advance, which is not always the case. While the need for scrolling through large collections can be mitigated by finding better ways to provide good recommendations or by improving search, there is always a need to have an efficient and appealing way to access 'your stuff' in its entirety.

Scrolling is composed of two variables: the scrolling speed and the scrolling direction. The purpose of this paper is to establish whether there is a potential benefit in splitting the control of scrolling speed and scrolling direction over two hands. By allowing the user's NDH to set the scrolling speed while their DH controls direction, it may be possible to give the user more control over the interaction. In terms of the KC Model [6] we can say that the NDH is setting the frame of reference (the speed) for the action of the DH (the scrolling). In this paper, we describe a number of scrolling techniques whereby we augment existing scrolling methods (drag and flick gestures on a touchscreen and on a dial) with a speed control mechanism. Control of the scrolling speed is given to the user's NDH using either pressure input or an on screen slider, and the control of direction is given to the user's DH using on screen drag gestures or a rear mounted dial.

2 Background

2.1 Bimanual Interaction on Touchscreen Devices

Multitouch devices are, by definition, capable of accepting bimanual input. By being able to sense multiple points of contact on the screen, a user can use either multiple fingers from one or multiple hands to interact. Studies have shown that touchscreen bimanual interaction techniques can improve performance [7], [8] and selection accuracy [9]. However, these studies assume that both hands are free to interact. There is no evidence to suggest that they would be beneficial in contexts where one hand is constrained by holding the device.

Despite the fact that one hand is often required to hold the device, it can do so in a variety of ways. As the hand may be in contact with the bezel and back of the device, these areas could be augmented with additional hardware to enable interaction. For example, RearType [10] includes a physical keyboard on the back of a tablet PC. Users hold it with both hands while entering text, thus avoiding an on-screen keyboard and graphical occlusion by the fingers. Lucid Touch [11] is a proof-of-concept see-through tablet that supports simultaneous touch input on the front and on the back of the device. Users hold the device with both hands, with thumbs on the front and remaining fingers on the back. The device is small enough that users can reach the entire back allowing multitouch interaction with both hands while fully supporting the device. However, the arm-mounted camera currently makes this approach impractical. Gummi [12] is a prototype "bendable" tablet that allows bimanual interaction by deforming the device by gripping its edges.

Wagner *et al.* [3] designed BiPad, a user interface toolkit to introduce bimanual interaction on tablets. It is designed to work on existing touchscreen tablets, without any additional hardware. The users' NDH can execute commands on special regions of the screen that are accessible while they are holding the tablet. For example, users can activate contextual menus to control the zooming and rotation of maps by tapping, gesturing or making chords with their NDH, while their DH selects items from the menus, or controls the position of the zooming and rotation, simultaneously. They found that the bimanual techniques did improve performance over unimanual techniques. Their aim was to provide a general way to provide bimanual interaction on tablet devices and actual behaviour of the NDH would vary from application to application.

2.2 Models of Bimanual Action

Early work on bimanual HCI assumed that users would be sat at a desktop interacting through various peripheral devices placed on the desk. Leganchuk, Zhai and Buxton [4] give an overview and valuable insight into the early work. In surveying the literature on bimanual HCI, they observe that there are contrasting views on whether bimanual interaction techniques actually provide any benefit when applied to desktop interactions. By analysing the interaction techniques from the early experiments with

Fig. 2. Hardware setup for our study: Griffin Technologies PowerMate with extended radius (right), the dial affixed to the rear of the tablet (centre) and the pressure sensor affixed to the top-left of the bezel of the tablet (left)

respect to Guiard's KC model [6], they observed that bimanual techniques which conformed to the model showed advantage over unimanual equivalents, while those that did not showed little or no advantage over equivalent unimanual techniques. From this, they concluded that two hands are not always better than one, and that when designing bimanual interaction techniques, it is important to do so using the KC model.

While Guiard's KC model is a useful and well tested characterisation, it only models a particular class of bimanual action [6]: asymmetric bimanual action. The cooperative and asymmetrical nature of the KC model describes that when human beings perform tasks with both of their hands they adopt different and complementary roles in order to do so. Guiard argues that vast majority of real life human manual acts belong to the bimanual asymmetric class and that asymmetry in action is the rule and symmetry the exception. Meaning that not only are there a set of tasks, such as opening a bottle or slicing food, that are obviously bimanual and asymmetric, but that even supposed unimanual tasks, such as throwing a dart or brushing your teeth, are essentially bimanual actions (where the NDH plays a supportive, postural role) and tasks where both hands perform essentially the same role either in phase (such as rope skipping or lifting) or out of phase (such as typing or rope climbing) are the exception to the rule.

Guiard argues that the relationship between the dominant and non-dominant hand is analogous to the relationship between a proximal and distal element in a kinematic chain. The implication of which is that the DH will act in relation to the action of the NDH, the granularity of action of the NDH is much coarser than the DH hand (i.e. the movement of the NDH is macrometic while the movement of the DH is micrometric) and the sequence of motion is NDH followed by DH. However, Latulipe and others [13–15] have demonstrated that there is a class of common HCI tasks that can be modelled as symmetric bimanual actions. Particularly, geometric translations are more effectively performed symmetrically than asymmetrically. Latulipe [15] describes a model of symmetric bimanual interaction in which tasks can be thought of and broken down into symmetric components that can be distributed over two hands.

However, one of the caveats of the model is that in order to perform symmetric interaction effectively, a user requires device symmetry. Therefore, using both hands on the touchscreen, symmetric bimanual input is possible (as is demonstrated in the ‘pinch-to-zoom’ and ‘rotate’ touch gestures on many touchscreen devices).

Since the goal of this paper is to explore ways to enable simultaneous two-handed input while the user is comfortably holding the device, we must conclude that in delegating one hand to holding the device, both will not be able to gain full access to the touchscreen and so asymmetric bimanual input should be used.

3 Bimanual Scrolling – Experiment 1

This study was based on the premise that the control of scrolling speed and vertical scrolling direction can be thought of as separate tasks and that the current *status quo* of combining both into a single unimanual gesture on a touchscreen or on physical dial can be improved upon. The experiment sought to determine whether splitting the control of scrolling speed and scrolling direction over two hands, in accordance with the KC Model [6], could improve user performance in a one-dimensional scrolling task on a touchscreen tablet device.

In this paper we control both the way the user holds the tablet and the amount they have to support it in order to determine whether these techniques have any value in and of themselves without having to deal with the numerous different ways people choose to hold tablets [3], which we saw as a confounding factor.

3.1 Input Methods

For direction control, we chose to use two existing scrolling methods for our input modalities: drag gestures on a touchscreen and a free rotating physical dial. Therefore, our direction control modalities were Touch and Dial.

A pressure sensor was chosen for one of the speed control modalities, since pressure has been demonstrated to be a useful modality for the control of speed (for rate based cursor control) [16]. A pressure sensor can be mapped well to the control of speed using an accelerator metaphor, where increasing the force will increase the speed and *vice versa*. Furthermore, isometric force input is useful as an input modality on mobile devices [16–18] and as an augmentation of finger/stylus input on touch-screens [19] (although not tested in the NDH). It can be detected using force sensing resistors (FSRs) that are flat and can be added to different locations on a device without changing its form factor.

Since we have a combination of physical and touch interactions in the direction control, we included a touch-based slider control for speed control as well. It did not require the NDH to perform a precise task (just one dimensional movement) and that it could be mapped well to speed control: up to increase the speed, down to decrease.

Fig. 3. Experimental Setup – Participant sat at a desk with the tablet supported on a stand

3.2 Interaction Techniques

There were six interaction techniques used in the study: two unimanual and four bimanual.

Unimanual Techniques. The two unimanual techniques were Unimanual Touch and Unimanual Dial, in which scrolling direction and speed were combined. The Unimanual Touch technique was the same as that found on current tablets and was used as a control condition for the experiment. To scroll through the list used in the study, participants would either drag on the screen or to perform a flick gesture on the screen that would cause the menu to scroll quickly in the direction of the flick. Flicking faster increased the velocity of the scrolling. The Unimanual Dial technique was similar insofar as participants could drag the dial to scroll through the list, or 'flick' the dial to scroll quickly in the direction of the flick in a way similar to that on the BeoSound 5 (See Figure 1).

Bimanual Techniques. The bimanual techniques used the same scrolling direction devices as the unimanual techniques, but two additional methods were used to control speed. The speed could be controlled dynamically using either a force sensing resistor (FSR) mounted on the top left front of the device's bezel or a software slider bar that appeared on the top left of the screen. Participants controlled speed by applying force to the FSR using the thumb of their NDH on the sensor and their other fingers behind the device, in a pinching gesture. Increasing the pressure dynamically increased the speed at which the direction control methods would scroll the menu. Releasing the pressure from the sensor would decrease the speed. A pressure space (amount of pressure that has to be applied to reach the maximum speed) of 9N was used for speed control. This was chosen because pilot tests revealed that with smaller pressure spaces, the speed control became binary, with the pinch pushing right through the pressure space.

The software slider bar was also controlled by the participants' NDH. Pushing the slider bar upwards increased the speed of scrolling and *vice versa*.

In the bimanual techniques, the speed control was completely separated from the direction control and so it was no longer possible to perform 'flick' gestures on either the dial or the touchscreen to increase scrolling speed. All permutations of these bimanual techniques were used: Bimanual Touch and Pressure, Bimanual Touch and Slider, Bimanual Dial and Pressure, and Bimanual Dial and Slider.

3.3 Participants

Eighteen participants (4 female, 14 male) ranging from 19-55 years of age (M=23) took part in the study, all of whom were right handed. They were paid £6 for participating.

3.4 Hypothesis

H1: Bimanual techniques designed with the KC Model will outperform equivalent unimanual techniques, measured by faster movement times, fewer target overshoots and lower subjective workload.

H2: The bimanual techniques will provide more benefits as the distance to the target increases, measured by faster movement times, fewer target overshoots and lower subjective workload.

H3: Within the bimanual techniques, pressure will outperform the touch slider as a speed control method, measured by faster movement times, fewer target overshoots and lower subjective workload.

H4: Within the bimanual techniques, the dial will outperform touch drag as a direction control method, measured by faster movement times, fewer target overshoots and lower subjective workload.

3.5 Experimental Design and Procedure

The study aimed to answer two research questions. Firstly, whether the bimanual techniques were better than the unimanual ones and secondly which combination of bimanual modalities were most effective. For the former, we simply compared each of the techniques, resulting in the variable *Interaction Technique* with six levels: Unimanual Touch, Unimanual Dial, Bimanual Touch + Pressure, Bimanual Touch + Slider, Bimanual Dial + Pressure, Bimanual Dial + Slider. These variables were used to test H1 and H2.

However, in doing this we cannot say anything about the different speed and direction control techniques that are being used. Since it is not possible to compare the bimanual speed and direction controls with the unimanual techniques (the control of speed or direction cannot be isolated in the unimanual techniques), an additional set of independent variables was required. Therefore, the variables *Scroll Method* and *Speed Method* were used to compare the bimanual techniques to one another, excluding the unimanual techniques. Each of these had two levels: Scroll Method (Touch or Dial) and Speed Method (Pressure or Slider). These variables were used to test H3 and H4.

Across both the research questions, we considered the effect of target distance on performance. Target distance was a useful measure as it allowed us to assess whether having a greater control of scrolling speed was useful when moving different distances. Therefore, the independent variables in the study were *Interaction Technique* and *Target Distance*, or *Scroll Method, Speed Method* and *Target Distance*. The dependent variables were *Movement Time* and Number of *Target Overshoots.* After each condition participants completed a NASA TLX [20], a six item questionnaire that assesses subjective workload. Movement Time was a measure of how long it took to complete a selection, from the first scrolling movement to the last scrolling movement before selection. Movement Time encapsulated the entire time to scroll though did not include any additional time taken to select an item (when, for instance, a participant had to move his or her hand from the dial to the touchscreen). Number of Target Overshoots was defined as the number of times a target disappeared from view after being visible. This meant we could measure how many times a participant overshot a target before selecting it, which served as a measure of control; fewer overshoots meant that the technique allowed greater control. Finally, Subjective Workload was measured using the NASA TLX [20], which gave a measure of how hard a participant though s/he had to work using each technique.

Procedure. The interaction techniques were implemented on a Viewsonic Viewpad 10" touchscreen tablet running custom software on Windows 7. A Griffin Technologies PowerMate Dial, with an extended radius (using the lid from a jar of fruit so that it could be easily reached at the side of the tablet), was used for the dial conditions and a single Force Sensing Resistor connected through a SAMH Engineering SK7-ExtGPI01 input/output module (which handled A-D conversion and sensor linerisation [21]) was used for pressure sensing (See Figure 2). Users applied pressure by performing a pinch gesture with the thumb and forefinger of their NDH on the left-hand bezel of the device.

The experimental task involved participants scrolling to and selecting an item from an alphabetically ordered list of 312 musical artists, which is similar to the task of selecting an artist to listen to from a long list within a music library on a tablet. In each condition, participants performed 19 tasks in total (the first 6 being training tasks). Target names would appear automatically on screen and after a selection had been made (whether correct or incorrect) the next task would begin automatically with the user being returned to the top of the list. This continued until all tasks had been completed. There were 6 unique data sets used in the study to avoid learning effects, and each participant used a different data set in each condition. Each set contained 312 alphabetically ordered musical artists.

The tasks were defined in terms of how far the participant would have to scroll from the very top of the list to the target. There were 13 experimental tasks (excluding the 6 training tasks) and the target to be selected in each task was different in each condition since there was a different data set for each condition. The distances associated with the tasks were 10, 35, 60, 85, 110, 135, 160, 185, 210, 235, 260, 285 and 310 items from the top. By defining the tasks in this way, and using a different dataset in each condition, we could compare the performance of each interaction technique over distance while mitigating any learning effect that might have occurred if participants were asked to select the same items in every condition. Conditions were counterbalanced using a Latin Square to mitigate any order effects.

4 Results

4.1 Overall Results – Interaction Technique and Distance

This section presents an overall analysis of the bimanual and unimanual conditions in which we compare each of the techniques to each other whole. In doing so, we can compare the performance of each technique to one another and test H1.

Movement Time. A two-way, repeated measures ANOVA showed a main effect for Interaction Technique, $F (5, 85) = 23.555$, $p < .001$, a main effect for Distance, $F (12, 204) = 47.653$, $p < .001$. The Interaction Technique x Distance interaction was not significant, $F (60, 1020) = 1.638$, $p = .099$.

Post hoc pairwise comparisons with Bonferroni corrections revealed that the combination of Dial and Slider was significantly faster than all other interaction techniques ($p < .001$). Touch and Slider was significantly faster than both the Unimanual Touch and Touch and Pressure techniques ($p <.001$). Unimanual Dial was significantly faster than the Unimanual Touch ($p < .001$) and the Dial and Pressure technique was significantly faster than the Touch and Pressure Technique ($p < .001$) and the Unimanual Touch Technique ($p<.001$).

Number of Target Overshoots. A two-way, repeated measures ANOVA showed no main effect for Interaction Type, $F(5,85) = 2.245, p = .057$. There was a main effect for Distance, $F (12, 204) = 1.516$, $p < .001$. There was no interaction between the two.

Subjective Workload. A one-way repeated measures ANOVA on the overall workload scores for each condition showed a significant main effect for Interaction Technique, $F(5, 85)$, $p <.001$. *Post hoc* pairwise comparisons with Bonferroni corrections revealed that the combination of Dial and Pressure had a significantly lower workload score than Dial and Slider ($p<.05$), Unimanual Touch ($p<.001$) and Touch and Pressure ($p<.05$).

Discussion. With this analysis we were interested in trying to ascertain whether there were any benefits of bimanual techniques over some equivalent unimanual techniques. The hypothesis that bimanual techniques designed to conform to the KC model will outperform equivalent unimanual techniques (H1) was generally borne out. The Dial and Slider technique was superior to the others in terms of Movement Time (see Figure 4a) with the Dial and Pressure technique superior in terms of Subjective Workload (See Figure 4b). These results suggest that the bimanual techniques do have advantages over unimanual equivalents.

In general, it took participants longer to select targets that were further away, which explains the main effect for Distance, but since there was no interaction effect for Interaction Technique x Distance there is no evidence to suggest that any of the bimanual techniques provide additional benefit as the distance from the target increases, and thus there is no evidence to support H2.

A more detailed analysis of the bimanual techniques in the following section will attempt to explain why participants performed better with the Dial and Slider combination, though had a lower subjective workload with the Dial and Pressure combination. A fully balanced experiment was not possible with these six interaction techniques (we could not isolate speed control for the unimanual conditions), we cannot make any concrete inferences about whether the Dial is a superior modality to Touch for controlling scroll direction either unimanually or bimanually.

4.2 Detailed Results – Scroll Method/Speed Method and Distance

A second analysis was carried out to test H3 and H4 which are concerned with the particular modalities used in the bimanual techniques and aim to test which, if any, resulted in better performance. The Independent Variables in this analysis were Scroll Method (Dial or Touch), Speed Method (Pressure or Slider) as well as Target Distance. The dependent variables were the same as the previous analysis: Movement Time, Number of Target Overshoots and Subjective Workload.

Movement Time. A three-way, repeated measures ANOVA on the movement times showed a main effect for Scroll Method, $F(1, 17) = 44.262$, $p < .001$, a main effect for Speed Method, $F(1,17) = 35.747$, $p < .001$ and a main effect for Distance $F(12, 204) = 27.898$, $p < .001$. There were no significant interactions between these three.

In general, it took longer to select items that were further away in the list. In addition, the movement times for the techniques that used the slider as a Speed Method

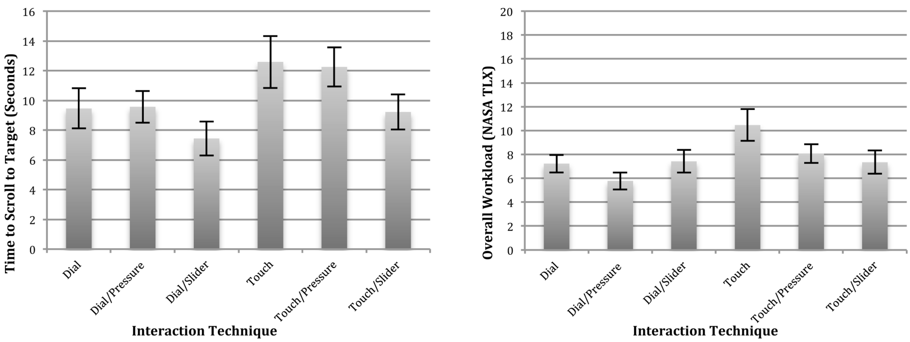


Fig. 4. (a) Average Movement Times for each Interaction Technique (b) Overall Subjective Workload for each Interaction Technique

were faster (M=8330ms, SD=5005ms) than the techniques that used the pressure as a speed control method (M=10911ms, SD=5290ms).

Number of Target Overshoots. A three-way, repeated measures ANOVA on the number of target overshoots showed no significant main effect for Scroll Method, $F(1,17) = .592, p= .452$, nor for Speed Method, $F (1, 17) = .426, p = .523$. There was a the main effect for Distance $F(12,204)= 3.381, p < .001$. Only the Scroll Method x Speed Method x Distance interaction was significant $F (12, 204)=2.778, p <.05$.

Subjective Workload. A two-way repeated measure ANOVA on the overall workload scores for each condition showed no significant main effect for Scroll Method $F(1,17) = 3.373, p = .084$, no significant main effect for Speed Method $F(1,17) = 1.255, p=.278$ though there was a significant interaction between Speed Method and Scroll Method $F(1,17)=10.111, p<.05$. *Post hoc* pairwise comparisons with Bonferroni corrections revealed that the combination of Dial and Pressure had a significantly lower workload score than Dial and Slider (p<.05), and Touch and Pressure (p<.05).

Discussion. The results from the second analysis reveal that, as a direction control method, participants performed tasks faster using the dial than with on screen touch gestures, though there was no evidence to suggest that Scroll Method has any effect on number of target overshoots, lending some support to the hypothesis that participants would perform better using the dial than the touch gestures (H4). Repetitive flick and or drag gestures make it difficult to get anything but very coarse control over the scrolling speed and cause the interaction to become slow and staggered. We believe that the reason the dial turned out to be faster was because it provided more continuous control during scrolling.

As before, it took participants longer to select targets that were further away, which explains the main effect for Distance. However, since there was no interaction effect between Scroll Method, Speed Method or Distance there is no evidence to suggest that any of the bimanual techniques provide additional benefit over any of the other techniques as the distance from the target increases.

As a speed control method, the on screen slider was better than the pressure sensor. Participants performed tasks faster when using the on screen slider than when using the pressure sensor. However, the pressure sensor was favoured in the subjective workload metrics, implying that people found it easier to use. There was no evidence to suggest that either of the modalities had an effect on the number of target overshoots. The hypothesis that pressure will outperform the touch slider as a speed control method (H4), then, was not supported. If we examine the differences in the levels of speed achieved using each of the techniques we can begin to explain these differences. Figure 5 shows the variation in the speed values for each technique. It can be seen that the distribution of speed values was skewed toward to lower end of the scale for the techniques that used pressure for speed control and is distributed across the centre for techniques that used the on screen slider for speed control.

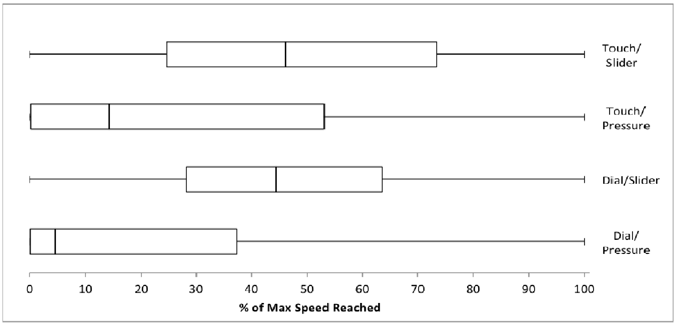


Fig. 5. Distribution of Speed Levels

During the experiment debrief, a number of participants commented that they found that they had to apply too much force in order to get to maximum speed; they therefore only applied a comfortable amount of force, and did not reach maximum speed. To scroll the list at the maximum speed, the participant would have to apply a force of 9N continuously on the pressure sensor. Whereas while with the touch slider it was possible to reach maximum speed by simply moving the slider knob to the top of the slider bar, participants rarely set is to the maximum value. Rather, participants commented that they only moved the slider in small increments or decrements because it was awkward to change. It was easier and more comfortable to travel at slower speeds with the pressure sensor, which could explain why participants took longer to complete the tasks with the pressure sensor. From this, we could hypothesise that by reducing the amount of force that is required to reach the maximum speed, we could reduce the amount of time needed to carry out the task while maintaining the subjective workload benefits of the pressure sensor.

5 Bimanual Scrolling – Experiment 2

The goal of the second study was to investigate further the use of pressure as a speed control method for bimanual scrolling interactions. In the previous study, the techniques that used pressure control did not out perform others, though participants perceived them to have a lower subjective workload. From this, it seems possible that the pressure space was too large. In this study, we tested whether decreasing the pressure space could improve objective performance while maintaining perceived workload. We were also interested in whether there is an optimal pressure space for bimanual scrolling.

5.1 Interaction Technique

There were eight interaction techniques used in the study, all of which were bimanual and all of which used pressure as a speed control method. They differed from the techniques used in the previous study in the following ways. Firstly, two different, and smaller 4N and 6N pressure spaces were used. These were smaller than the pressure space used in the last study (9N), which had been chosen because pilot testing had suggested that with too small a pressure space the control of speed became binary. Thus, while the aim of this study is to evaluate whether decreasing the pressure space improves performance, we do not expect that the smallest pressure space will necessarily be the best.

In addition to varying the pressure space, the way in which the pressure sensor controlled the speed was varied as well. In the first study, an accelerator metaphor was used, but we encountered the problem that participants struggled to reach maximum speed. While reducing the pressure space, as discussed above, could solve this it could also be solved by applying a "brake" metaphor instead. A brake metaphor could be useful because by default the scrolling speed would be set to maximum, and then could be decreased by applying pressure. In doing this it makes it easier to achieve higher speeds, though could also make it more difficult to scroll at slower speeds, reducing target selection accuracy. By including this variation here we hope to characterise this trade-off as well.

5.2 Participants

Sixteen new participants (9 female, 7 male) ranging from 19-31 years of age (M=21) took part in the study, all of whom were right handed. They were paid £6 for participating.

5.3 Hypothesis

H5: The larger (6N) pressure space will outperform the smaller (4N) pressure space, measured by faster movement times, fewer target overshoots and lower subjective workload.

H6: As a method of speed control, Brake will provide better performance over longer distances than Accelerator, measured by faster movement times, fewer target overshoots and lower subjective workload.

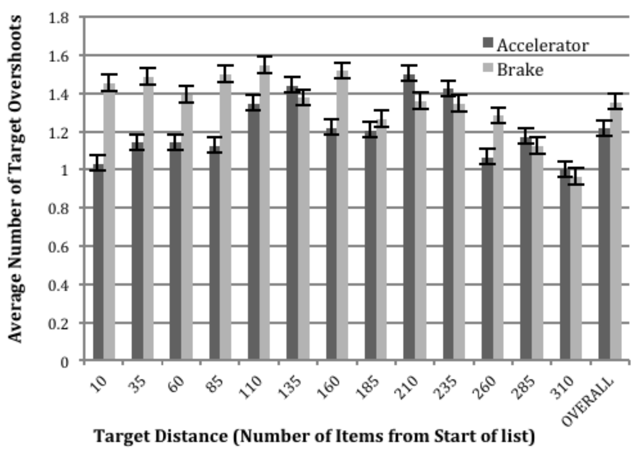


Fig. 6. Number of Target Overshoots

5.4 Experimental Design and Procedure

The experimental task was identical to the one in the previous study and involved participants scrolling to and selecting an item from an alphabetically ordered list of 312 musical artists. Conditions were counterbalanced using a Latin Square to mitigate any order effects.

The independent variables were: Scroll Method (Dial, Touch Drag), Pressure Space (4N, 6N), Pressure Mode (Accelerator, Brake) and Distance (10, 35, 60, 85, 110, 135, 160, 185, 210, 235, 260, 285 and 310). The dependent variables were Movement Time, Number of Target Overshoots and Subjective Workload.

5.5 Results

Movement Time. A four-way, repeated measures ANOVA showed a significant main effect for Scroll Method $F(1, 14) = 5.426$, $p < .05$ and for Distance $F(12, 168) = 11.413$, $p < .001$. There was no evidence that either Pressure Mode (Accelerator or Brake) or Pressure Space (4N or 6N) had any effect on Movement Time. There were no significant interactions. In general, participants took longer to scroll to targets that were further away from the start of the list and participants took longer to scroll to targets when using the Touch Scroll method (M=13708ms, SD=7135ms) than when using the Dial Scroll method (M=12429ms, SD=7922ms).

Number of Target Overshoots. A four-way, repeated measures ANOVA showed a significant main effect for Pressure Mode, $F(1,14) = 7.583$,$p < .05$, and a significant main effect for Distance $F(12, 168) = 2.985$, $p < .001$ as well as a significant interaction effect for Mode x Distance $F(12, 168) = 1.973$, $p < .05$ and for Scroll Method x Pressure Space x Pressure Mode x Distance $F(12, 168) = 2.246$, $p < .05$.

In general, there were significantly more target overshoots in the conditions in which the pressure sensor was used as a brake (M=1.35, SD=0.9) than in the conditions in which it was used as an accelerator (M=1.21, SD=0.58). *Post hoc* pairwise comparisons of the number of target overshoots across all 13 Distances revealed that when selecting the target at position 310 (the distance furthest away from the top, which was on the last page of targets and could not be overshot) participants had significantly fewer target overshoots than with any of the other distances ($p < .001$). As can be seen in Figure 6 the Pressure Mode x Distance interaction can be explained by the fact that for the targets closer to the start of the list, the Brake mode had a much larger number of target overshoots than the Accelerator mode, though as the target distance increases, the difference between the two modes decreases.

Subjective Workload. A four-way repeated measure ANOVA on the overall workload scores for each condition showed no significant main effect for Scroll Method $F(1,15) = 2.475$, $p = .137$, no significant main effect for Pressure Space $F(1,15) = 2.524$, $p=.133$ and no significant main effect for Pressure Mode $F(1,15)=3.750$,$p=.072$. There were no significant interactions.

5.6 Discussion

The results of the second experiment were not conclusive. There was no evidence to suggest that the differences in Pressure Space (4N or 6N) or Pressure mode (Accelerator or Brake) had any effect on Movement time. However, the data do suggest that participants could perform faster with the Dial over Touch as a method to control scrolling direction, supporting H4. This mirrors the results obtained in the first study and suggests that the dial is better suited as a direction control device for these interactions.

The data also suggests that the Brake mode resulted in less accurate performance for targets that were closer to the start position. Since, by definition, the Brake mode moves very quickly for small movements when no pressure is applied, and the starting state for the condition was to have no pressure applied, then it is conceivable that for targets that are a short distance away participants are more likely to overshoot. The potential advantage of the Brake mode lies in the fact that it requires less effort to achieve higher speeds, though it comes with the trade-off of more effort to reach lower speeds. Thus, it is not clear whether this potential advantage has any merit in realistic situations due to the extra effort involved in travelling short distances. In addition there was no evidence that the Brake mode actually improved performance or reduced subjective workload for larger distances, leading us to reject H6. When using the Accelerator mode, people can always navigate to a target, albeit slowly, without needing to apply a great deal of pressure, which for short distances seems to result in more accurate performance.

6 General Discussion

6.1 Dial vs. Touch for Direction Control

With the prevalence of touchscreens, physical dials are not particularly common on modern devices. However, the results in this paper suggest that, in terms of movement time and subjective workload, they are superior to flick and drag gestures on a touchscreen for the control of scrolling direction. Numerous keyboards and mice contain small dials that are used for scrolling through content on a desktop machine, and the first generation Apple iPod featured a front mounted touch sensitive dial that was the main source of input on the device.

A dial provides the opportunity for continuous control during a scrolling task, unlike the flick and drag gestures that are used on touchscreen devices, which may be part of an explanation as to why it performed better in the studies described in this paper. However, the dial used in this study was cumbersome when mounted on the device. Future work will consider less obtrusive ways to incorporate a dial into the form factor of a tablet, such as with a flat touch sensitive dial.

6.2 Pressure Space

There were three ‘pressure spaces’ used across the studies presented in this paper: 9N in the first study and 4N and 6N in the second. It was observed that the 9N pressure space might have been too large for participants to comfortably apply the force required to reach the maximum speed. In response to this, the second study contained two smaller pressure spaces (4N and 6N) as well as introducing a 'brake' metaphor for speed control (alongside the 'accelerator' metaphor that was used in the first study) in an attempt to make it more comfortable to control the scrolling at higher speeds. We hypothesised that the 6N pressure space would give rise to better performance than the 4N pressure space since it would reduce the amount of force participants had to apply, while still giving a wide enough range to allow expressive use of the speed control. However, there was no evidence to suggest that the differences in pressure space had any effect on performance. It is possible that the distances travelled during the experiment were too small to allow for truly expressive use of the speed control. For some target distances, it may not have been possible to achieve maximum speed before the target was reached (or overshot). If this were the case, the effect of pressure space would be masked because the task did not require it to be fully utilised. Future work will explore this issue by evaluating the interaction techniques using tasks that are longer and more involved than targeting tasks, such as a browsing task. This will mean that the use of the techniques can be studied when the user has to navigate the collection in more detail, thus giving more opportunity to make use of the input strategies available.

7 Conclusions and Future Work

The studies presented here suggest that the bimanual scrolling techniques are better than the status-quo unimanual techniques in terms of both performance and preference, lending support to the body of evidence in HCI that the KC Model [6] is a useful tool to inform the design of bimanual interactions that allow people to carry out tasks more effectively than with unimanual equivalents. The studies also suggest that, as a method of scrolling control, the physical dial is better than conventional touchscreen gestures in both the bimanual and unimanual techniques.

As for speed control methods, the evidence suggests that using touch slider resulted in faster performance than the pressure sensor, however participants favoured the pressure sensor in terms of subjective workload, implying they found it easier to use. There was no evidence to suggest that either had an effect on target overshoots. We proceeded by investigating how different configurations might improve performance with the pressure sensor with no clear results. No particular configuration came out as better than the rest, and the average movement time across all conditions was higher (M=13s, SD=7.5) than the first experiment (M=10s, SD=5.5s). However, since each study used different participants and the second had more conditions, we cannot compare the results of the two studies directly. In future work we will look more closely at the different pressure configurations and in such a way that allows us to compare them to the touch speed control method we used.

In conclusion, the studies presented in this paper suggest that splitting the control of scrolling speed and scrolling direction across two hands is a viable way to scroll on a tablet, which could support simultaneous two-handed input while the user is holding the device.

Acknowledgments. This work was partly funded by Bang & Olufsen.

References

[1] Vogel, D., Cudmore, M., Casiez, G., Balakrishnan, R., Keliher, L.: Hand occlusion with tablet-sized direct pen input. In: Proceedings ACM CHI 2009, pp. 557–566 (2009)
[2] Vogel, D., Casiez, G.: Hand occlusion on a multi-touch tabletop. In: Proceedings of ACM CHI 2012, pp. 2307–2316 (2012)
[3] Wagner, J., Huot, S., Mackay, W.: BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets. In: Proceedings of ACM CHI 2012, pp. 2317–2326 (2012)
[4] Leganchuk, A., Zhai, S.: Manual and cognitive benefits of two-handed input: an experimental study. ACM Transactions on Computer-Human Interaction 5(4), 329–346 (1998)
[5] Deml, B., Egermeier, H.: Are Two Hands Always Better Than One? A Study on Bimanual Input Control. In: Prococeedings of IEEE Humanoids (2003)
[6] Guiard, Y.: Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. Journal of Motor Behavior 19(4), 486–517 (1987)
[7] Kin, K., Hartmann, B., Agrawala, M.: Two-handed marking menus for multitouch devices. ACM Transactions on Computer-Human Interaction 18(3), 1–23 (2011)
[8] Brandl, P., Forlines, C., Wigdor, D.: Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 154–161 (2008)
[9] Benko, H., Wilson, A.D., Baudisch, P.: Precise selection techniques for multi-touch screens. In: Proceedings of ACM CHI 2006, pp. 1263–1272 (2006)
[10] Scott, J., Izadi, S., Rezai, L., Ruszkowski, D.: RearType: text entry using keys on the back of a device. In: Proceedings of MobileHCI 2010, pp. 171–179 (2010)
[11] Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., Shen, C.: LucidTouch: A See-Through Mobile Device. In: Proceedings of ACM UIST 2007, pp. 269–278 (2007)
[12] Schwesig, C., Poupyrev, I.: Gummi: a bendable computer. In: Proceedings of ACM CHI 2004, pp. 263–270 (2004)
[13] Latulipe, C., Mann, S., Clarke, C.L.A., Kaplan, C.S.: symSpline: Symmetric Two-Handed Spline Manipulation. In: Proceedings of ACM CHI 2006, pp. 349–358 (2006)
[14] Balakrishnan, R., Hinckley, K.: Symmetric bimanual interaction. In: Proceedings of ACM CHI 2000, pp. 33–40 (2000)
[15] Latulipe, C.: A symmetric interaction model for bimanual input. University of Waterloo (2007)
[16] Wilson, G., Brewster, S.A., Halvey, M., Crossan, A., Stewart, C.: The effects of walking, feedback and control method on pressure-based interaction. In: Proceedings of MobileHCI 2011, pp. 147–156 (2011)
[17] Wilson, G., Stewart, C., Brewster, S.A.: Pressure-based menu selection for mobile devices. In: Proceedings of MobileHCI 2010, pp. 181–190 (2010)
[18] Brewster, S.A., Hughes, M.: Pressure-based text entry for mobile devices. In: Proceedings of MobileHCI 2009, p. 9 (2009)
[19] Ramos, G., Boulos, M., Balakrishnan, R.: Pressure widgets. In: Proceedings of ACM CHI 2004, vol. 6(1), pp. 487–494 (2004)
[20] Hart, L.E., Staveland, S.G.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload (1988)
[21] Stewart, C., Rohs, M., Kratz, S., Essl, G.: Characteristics of pressure-based input for mobile devices. In: Proceedings ACM CHI 2010, pp. 801–810 (2010)

Public Information System Interface Design Research

Ning Zhang, Junliang Chen, Zhengjie Liu, and Jun Zhang

Computer Science and Technology College, Dalian Maritime University,
116026 Dalian, China
{jenny,liuzhj,chenjl,william}@dlmu.edu.cn

Abstract. The diversity of users' cognitive skills remains the challenge of public information system interface design. In this paper, we focus on the universal interaction design method for public information systems like kiosks. We have developed a method with six steps based on the resources model. The method we proposed aims at reducing users' cognitive load and enabling designers to optimize interface information. To validate this method, two prototypes were designed based on the method and a usability test was conducted to compare users' cognitive load, performance and satisfaction between the designed prototypes and the current referencing system. Results show that, in contrast with the current reference system, prototypes we designed based on the proposed method can reduce user's cognitive load, and enhance user's performance and satisfaction.

Keywords: Universal usability, Cognitive load, Public information system.

1 Introduction

With the widespread of public information systems (PIS) in various service industries in China, users with various backgrounds become a challenge for PIS's interface design [1]. PIS designers and interface researchers have been exploring different ways to provide PIS users with more user-friendly and understandable interaction platform, including developing multi-model interface [2], setting up accessibility guidelines [3], and providing interface solutions based on usability test [4]. However, the lack of interaction design theory and design methods concerning PIS users 'cognitive levels remains the bottleneck for PIS interaction design, resulting in the loss of users and the low use rate of current PIS in China.

In this paper, a PIS interface design model based on distributed cognition theory was built to provide foundations and guide for the appropriate PIS interface design method, so that users' cognitive load can be minimized during the PIS interface design process. With hotel self-service kiosk interface design as an example, two prototypes were designed based on the proposed PIS interface design method. To validate whether PIS interface design method can reduce users' cognitive load, a comparative usability test was carried out to evaluate users' performance, satisfaction and cognitive load. Results confirm the validity and effectiveness of the proposed PIS interface design method.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 247–259, 2013.

2 Relevant Research

Resources model of distributed cognition theory and universal design method are main supportive research area for this study.

2.1 Resources Model

Resources model of distributed cognition theory was first introduced into human-computer interaction (HCI) design by Wright [5], to describe how information is distributed between users and computer systems, how to design appropriate and reasonable outer representations to minimize users' cognitive load. Resources model is mainly applied to describe human computer interaction activities and evaluate interfaces, and remains one of the most influential models in HCI field.

Resources refer to a collection of information structures that can be defined for each step in an interaction and which can be used to inform action. Information structure and interaction strategies are two components of resources model. Plan following, plan construction, goal matching and history-based selection are four strategies defined by Wright to analyze, design and evaluate interfaces. The basic information structure of resources model consists of six elements, including plan, goal, affordance, history, state and action-effect relations. Detailed definitions of the six elements are listed in figure 1. As an emerging theory in HCI, resources model is mainly applied in HCI research by Smith [6], Cheng [7] and Wang [8].

Table 1. Elements definitions of information struction of resources model

Element	Definition
Plan	A sequence of actions, events and states that could be carried out.
Goal	A required state of the system
Affordance	A set of possible next actions can be taken by the user for a give state of the system
History	Actions, events or states already achieved in the interaction
Sate	The collection of relevant values of the objects that feature in the interaction at a given point in the interaction
Action-effect relation	A causal relation between an action or event and state which represents the effect that executing the action or event will have on the interaction

2.2 Universal Usability

Universal usability was first put forward in interface interaction design area by Ben Schneiderman [9], emphasizing the design of information products and services should be usable for every citizen.

Universal usability design (UUD) method includes three steps. First, it requires designers to classify different user groups that need special considerations during design process; second, build universal usability design matrix for the classified user groups, listing the specific requirements of different user groups and relevant design items; third, propose design solutions according to design matrix and evaluate the design solutions.

Since PIS users significantly differ in terms of age, education background, and other variables impacting cognitive load, it is feasible to adopt universal usability design method to guide PIS design process. As a newly introduced design method in HCI study, the application of this method in PIS interaction design is still limited, some representative exploring studies include Kouroupetroglou [10], Carbonell [11], and Takeo [12].

3 Interaction Model for PIS

There are two common ways to reduce users' cognitive load in interaction process. One is to simplify interaction flow; the other is to provide information easily understood for majority users. Resources model and UUD method can provide necessary support for the realization of the above mentioned ways. By applying resources model into the development of PIS interface, interaction flow can be simplified during design process, and cognitive resources can be allocated to systems as much as possible. Combing with UUD method, designers can provide appropriate information easily understood for majority users. Based on resources model and UUD method, PIS interface interaction model and design method can be built to help designers make reasonable interface information allocation decisions.

3.1 PIS Interface Information Structure

PIS interface differs from traditional desk interface in its full screen display, and users can only complete tasks in a certain fixed interaction flow. Besides that, public context has impact on users' interactions with PIS. Therefore, we adopt flow and context as two key elements into PIS interface information structure. Flow is defined as the set of actions during interaction process. Describing flow can help interface

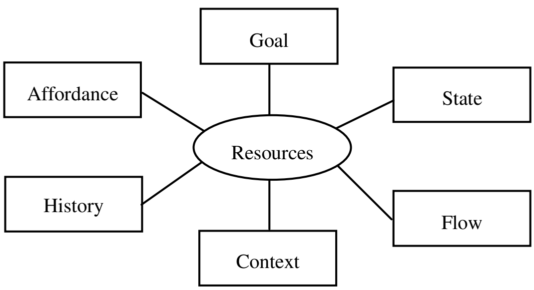


Fig. 1. PIS interface information structure

designers and researchers follow users' real time goals. Context is defined as users' action-related conditions and variables in human computer interaction. The illustrations of PIS interface information structure is showed in figure 1.

The PIS interface information structure in fig 1 is aimed at describing general information a user need to complete a certain task through PIS. Based on this information structure, a set of interaction strategies can be built to guide the design and evaluations of PIS interaction process.

3.2 PIS Interface Interaction Strategy

We propose three basic interaction strategies for PIS interface, including goal matching and optimization, flow construction and evaluations, and affordance identification and obtaining. These interaction strategies can describe users' possible actions and correspondent operations a system provides in interaction process.

(1) Goal matching and optimization

According to goal matching and optimization strategy, users need goal, context, flow, state, and affordance to complete a task. This is a dynamic process. The state of resources changes with users' operations. For users with different cognitive skills, the matching extent of goals and resources differs. Designers have to analyze users' cognitive skills and evaluate the resources a system can provide to improve design and realize the optimized matching of users' goals and resources a system can provide. Detailed goal matching and optimization strategy is showed in figure 2.

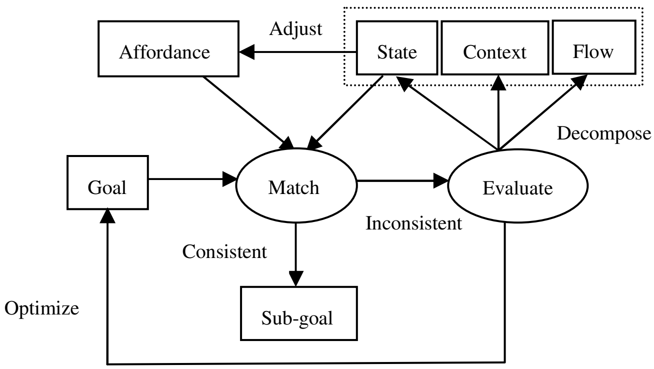


Fig. 2. Goal matching and optimization strategy

This strategy requires designers investigating users' cognitive skills before allo cating resources during interface design process. Methods like card sorting and participatory design can help designers bridge the gap between users' mental model and designers' mental model, and realize goal matching in an optimum condition.

(2) Flow construction and implementation

Users need information like affordance, current system state to decide next possible actions when interacting with computers. Basic operation flow of a task is composed of necessary interaction steps. Efficient and concise operation flow will facilitate the realization of goals. Detailed flow construction and implementation strategy is showed in figure 3. In figure 3, internal flow means the task completion order users perceive according to knowledge and system current state, external flow is the real task completion order of the designed system. Designers need analyze users' understanding about the system's working flow and decide the optimum flow for system development.

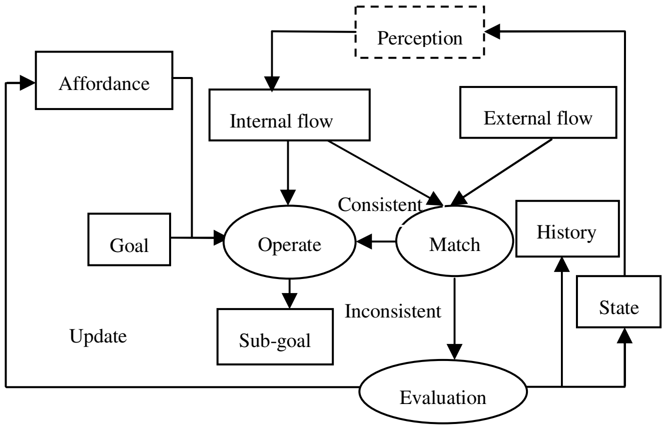


Fig. 3. Flow construction and implementation

This strategy requires designers building the appropriate task completion order and interface layout based on interaction task analysis. Universal usability design method can be taken into account to design interface layout. The design matrix from UUD method includes all the affordance users of different groups need to take actions according to task completion order. The set of affordance in the design matrix lists necessary points designers need consider in interface design process. PIS interface designers can make use of UUD method to propose specific design proposals for users with different background.

(3) Affordance identification and obtaining

Whether or not obtaining the right affordance is crucial for users to complete an interaction task. Too much affordance will increase users' cognitive load, while the lack of affordance prevents users completing tasks. Due to the diversity of PIS users' background, different users need different affordance. Therefore, affordance identification and obtaining strategy requires designers evaluating users' cognitive load and adjusting affordance based on evaluation results. Besides interface layout, context is also an important factor influence users' affordance identification and obtaining. If users' cognitive load is tested in usability lab, designers can use tapping test to simulate real operation situations, and to enhance the reliability of evaluation results.

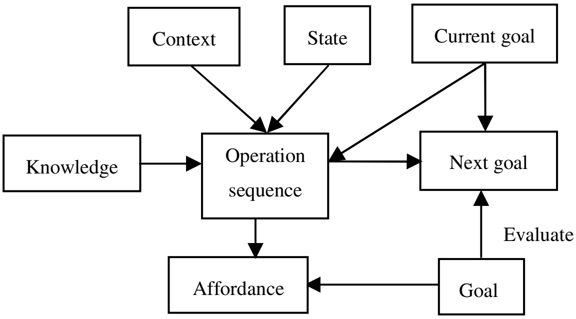


Fig. 4. Affordance identification and obtaining

The above mentioned PIS interaction model reflects the relationship between actions and information presentations. However, only models are not enough for interface designers to carry out specific design. PIS interaction methods based on PIS interaction model are necessary tools for designers to operate during design process.

4 PIS Interaction Design Method

Goal matching and optimization strategy requires designers focusing on different user groups' cognition analysis; flow construction and implementation strategy requires designers implementing information architecture under the help of UUD method; affordance identification and obtaining strategy requires designers evaluating users' cognitive load to make sure if the design proposal is reasonable enough for users with diversified background. These strategies need to be implemented throughout the whole design process. The common user-centered design methods include user needs analysis, functional analysis, information architecture, prototype design and usability evaluation. Correspondently, the PIS interaction design method we proposed is divided into six stages, that is, users' cognitive analysis, interaction task analysis, interface information layout, design project framing, prototype design, and cognitive load evaluation. Figure 5 shows the outline of the PIS interaction design method.

4.1 Users' Cognitive Analysis

Most of PIS users are either beginners or occasionally users, named "forever middle users" by Alan Cooper [13], not to mention those without computer skills. Relevant research [1] shows that age, education and experience are main factors influencing users' attitudes towards PIS products. Therefore, PIS users can be grouped according to their age, education and PIS product experience.

After dividing users into different groups, designers need contextual inquiry and interview to analyze different users' cognition towards PIS products, watching or interviewing users about their interaction experience with PIS products, recording the problems they meet, and the reasons users' goals are not matched, including problems from contextual influence, information overload, information misleading, and information deficiency. According to ISO25062 [14], each user group need at least 8 participants. Findings from users' cognition analysis should be edited into table form, as checklists for further evaluation reference.

4.2 Interaction Task Analysis

In interaction task analysis stage, designers need solve two problems. One is task categorization; another is to ensure task completion orders. The aim of this stage is to settle the general structure of the system being designed, and to describe the relationship between users' input and system's output.

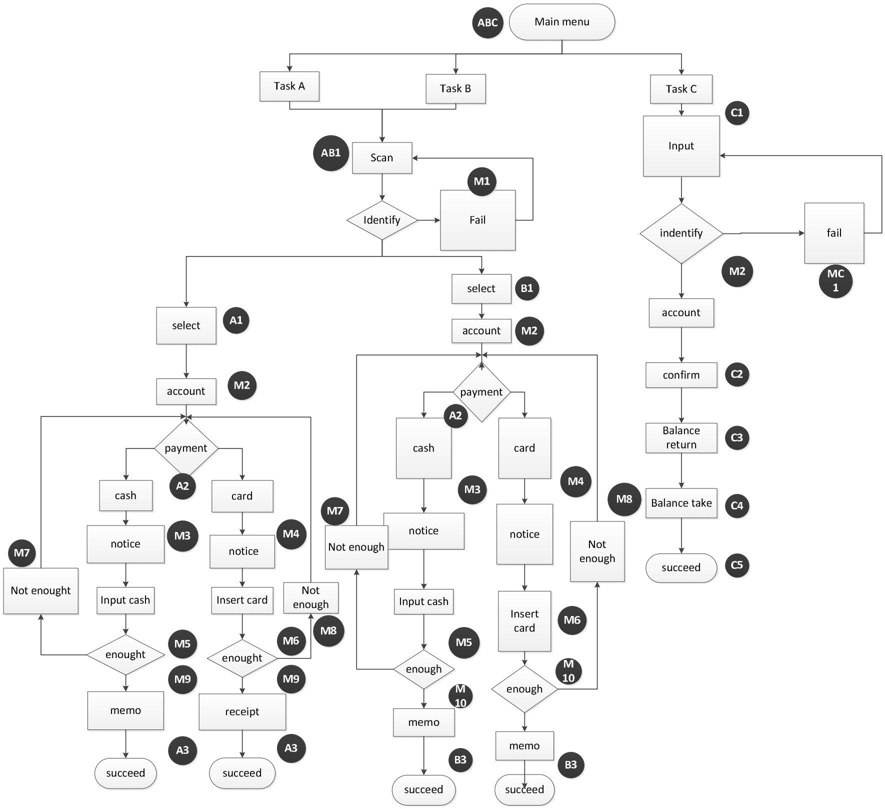


Fig. 5. Illustrations of task completion order

Designers can categorize different tasks need to be realized in the target system by card sorting, and then build completion order for each specific task through participatory design method, each step of the task flow has to be numbered according to the fixed order, as it is showed in figure 5. In this way, the basic task flow of the system can be drawn. In figure 5, the work flow of the three tasks of the main menu is listed.

For example, C1 in figure 5 represents the first subtask of task C, and M2 represents the second system information provided for users to complete certain task.

4.3 Interface Information Layout

Based on the results of users' cognition analysis and interaction task analysis, PIS interface design matrix can be built through UUD method. All the points listed in the matrix can help designers do specific design to reduce users' cognitive load, as it is showed in table 2.

Table 2. Examples of design matrix

user / task	Senior users	Fresh users	Low educated users
Step 1	××	××	××
Step 2	××	××	××
Step 3	××	××	××
:	××	××	××
:	××	××	××
Step n	××	××	××

With user groups as the horizontal axis of the matrix, and task completion order as the vertical axis, designers can input the rest intersectional content of the matrix according to users' cognition analysis results, and current PIS interface design standards [3], including various specific requirements of different user groups, the solutions of their problems need to be considered. Each solution corresponds to the problem a group of user have in a certain stage of a task flow. Usually there is no appropriate data that can be used directly by designers in a specific design process, especially quantitative data. Users' self-reported data obtained by interview, observation and contextual inquiry is the main source for designing PIS interface for users with different cognitive levels.

4.4 Design Project Framing

The output of design project framing is design document. Interaction designers need to draw up specific design documents based on the design matrix, so that system developers can read in details. A basic design document is consisted of information architecture, interface layout, interface elements, task completion flow and interface order.

4.5 Prototype Design

In prototype design stage, designers can use Flash, Axure RP and other tools to realize the information visualization based on the design documents. The designed prototypes should be revised according to evaluation results. Formative evaluation methods like cognitive walkthrough and heuristic evaluation can be applied in this stage to save design and development cost.

4.6 Cognitive Load Evaluation

User's cognitive load level correlates the extent of system affordance identification and obtaining. Research has found that users' performance is not directly correlated with cognitive load level, some system are tested with high performance but low users' satisfaction [15]. The main reason is the existence of cognitive load during human computer interaction process. Therefore, it is necessary to evaluate the cognitive load level of PIS users.

Tapping test proves an effective and efficient method to test users' cognitive load and system's usability [16]. Comparison test is also necessary for designers to compare users cognitive load levels when interacting with different prototypes. Cognitive load can be evaluated by NASA-TLX [17], Paas Scale [18], and user performance.

5 Case Study

In order to test the validity and feasibility of the proposed PIS interaction method, we recruited two interaction designers to design hotel self-service system interface according to the requirements of PIS interaction design method. We chose PIS in hotel industry because of the increasing tendency of PIS applications in hotels, and the diversity of hotel clients.

Fig. 6. Prototype A

Fig. 7. Prototype B

Prototype system A and B designed by PIS interaction design method are showed in figure 6 and figure 7. We compare the prototypes with a current running hotel self-service system in Shandong, China, as the reference system showed in figure 8. Prototype systems and the reference system will be compared in terms of users' cognitive load, performance and satisfaction.

The comparison test is conducted in the usability lab of Dalian Maritime University, China. We recruited 32 users to participate in our test, and divide these users in two groups. One is mainstream user group, with 16 users using PIS at least once a month, aged from 19 to 28, all the users in mainstream user group are university students; the other is non-mainstream user group, with 16 users no PIS product experience, aged from 45to 78, all the users in testing group have no bachelor degree.

Fig. 8. Reference System

All users are required to complete three tasks on touchscreens, including self-service check-in, self-service check-out, and room renewal, using prototype system A and B, as well as reference system, When interacting with systems, users are asked to tap with the rhythm of the music playing in the usability lab. Users' interaction process is recorded to calculate task completion time and error rate. Users are also asked to complete NASA-TLX scale and SUS (system usability scale) to calculate their cognitive load and satisfactions towards different systems. Test results are listed below.

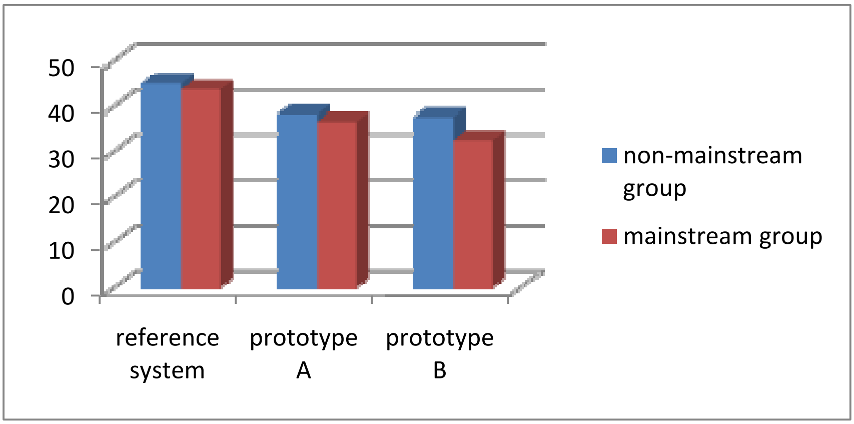


Fig. 9. Comparison of cognitive load means

Figure 9 shows that the cognitive load of the mainstream user group is lower than the cognitive load of the testing group when they complete the designed tasks through the three systems. Both of the groups' cognitive load is lower when using prototype systems than using the reference system.

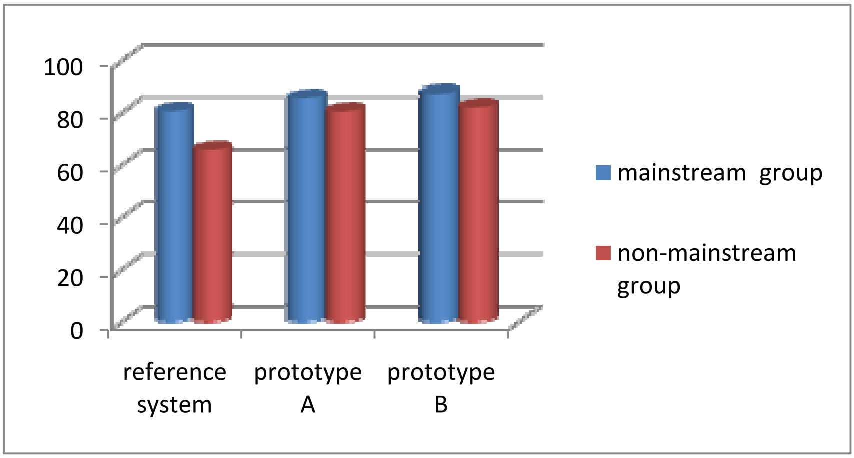


Fig. 10. Comparison of satisfaction means

Figure 10 shows that the mainstream users' satisfaction towards the three systems is higher than non-mainstream group's satisfaction. Both of the groups' satisfaction towards prototype systems is higher than towards reference system.

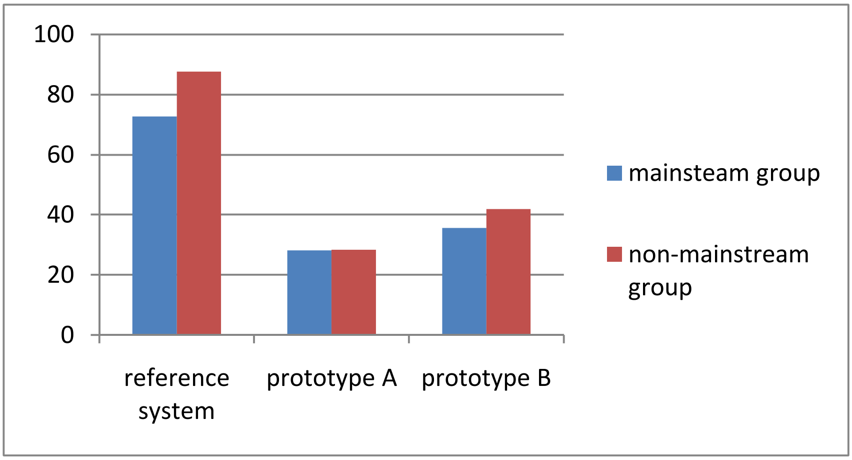


Fig. 11. Comparison of task completion time means

Figure 11 shows that main stream users spend less time completing the designed tasks through three systems than non-mainstream users. Both of the groups spend less time completing the designed tasks through prototype systems than the reference system.

Tapping test results show that the ratios of mainstream users and non-mainstream users' missing the rhythm of the music when completing the designed task through the three systems are8:6, 4:3, and 1:1 respectively, as it is showed in figure 12.

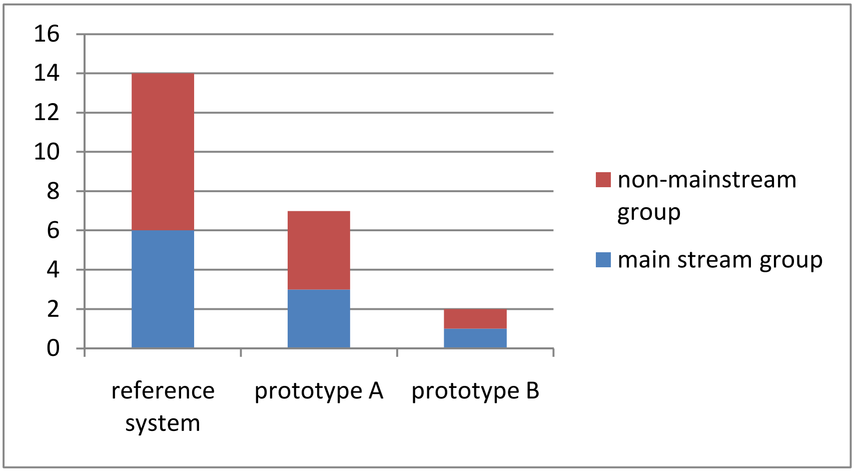


Fig. 12. Tapping missing times

Independent T test results show that there is significant difference between users' satisfaction towards reference system and towards prototype systems. There is also significant difference between users' performance on the reference system and on the prototype systems. However, there is no significant difference between users' cognitive load on the reference system and on prototype systems, and this result needs further validation among broader and more representative sample.

Data analysis results mentioned above, to some extent, support the validity and feasibility of the PIS interaction method.

6 Limitations and Recommendations

Due to the limitations of the research sample, the results of the study still need further validation. The comparison test was conducted in usability lab, although tapping test was applied to simulate the real interaction environment, there is still difference between lab environment and the real world situation.

References

1. Cremers, A.H.M., de Jong, J.G.M., van Balken, J.S.: User-Centered Design with Illiterate Persons: The Case of the ATM User Interface. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 713–720. Springer, Heidelberg (2008)
2. Mark, T., Fredrik, E., Annika, W.: Adaptive User Interfaces for Public Kiosks. Journal of User Modeling and User-Adapted Interaction 14(2) (2004)
3. Maguire, M.C.: A Review of User-Interface Design Guidelines for Public Information Kiosk Systems. International Journal of Human-Computer Studies 50, 263–286 (1999)
4. Farooq, A.: Haptics in Kiosks and ATMs for the Disabled. University of Tampere, Tampere (2009)
5. Wright, P.C., Fields, R.E., Harrison, M.D.: Analyzing Human-Computer Interaction as Distributed Cognition: the Resources Model. Journal of Human Computer Interaction 51, 1–41 (2001)
6. Smith, S., Duke, D., Wright, P.C.: Using the Resources Model in Virtual Environment Design. In: King's Manor Workshop on user-Centered Design and Implementation of Virtual Environments. University of York (1996)
7. Cheng, S.W., Sun, S.Q.: Human-Computer Interaction Resource Model based on Distributed Cognition. Journal of Computer Integrated Manufacturing System 14, 1683–1703 (2008)
8. Wang, C.Q., Deng, C.Z., Ma, C.X., et al.: an Extended Resources Model based on Distributed Cognition. Journal of Software 16, 1717–1725 (2005)
9. Shneiderman, B., Hochheise, H.: Universal Usability as a Stimulus to Advanced Interaction Design. Behaviour & Information Technology 20, 367–376 (2001)
10. Kouroupetroglou, G.: Universal Access in Pubic Terminals: Information Kiosks and Automated Teller Machines. In: The Universal Access Handbook. CRC Press (2008)
11. Carbonell, N., Stephanidis, C.: Universal Access: Theoretical Perspectives, Practice and Experience. In: 7th ERCIM International Workshop on User Interfaces for All, pp. 63–67 (2002)
12. Takeo, A., Akinori, O., Taku, A.: Promoting Universal Design of Automated Teller Machines. FUJITSU Scientific and Technical Journal 41, 86–96 (2005)
13. Cooper, A., Ding, C., Dream, D.: The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity. Publishing House of Electronics Industry, Beijing (2006)
14. ISO25062.Software Product Quality Requirements and Evaluation (SQuaRE)-Common Industry Format (CIF) for Usability Test Reports. International Standards Organization. Geneva, Switzerland (2006)
15. Fisher, C.D.: Why do Lay People Believe that Satisfaction and Performance are Correlated? Possible Sources of a Commonsense Theory. Journal of Organizational Behavior 24, 753–777 (2003)
16. Albers, M.: Tapping as a Measure of Cognitive Load and Website Usability. In: Proceedings of the 29th ACM International Conference on Design of Communication, pp. 25–32 (2011)
17. Hart, S.G.: NASA- Task Load Index (NASA-TLX):20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 904–908 (2006)
18. Paas, F., Tuovinen, J.E., Tabbers, H., et al.: Cognitive Load Measurements as Means to Advance Cognitive Load theory. Educational Psychologist 38, 63–71 (2003)

Comparison of User Performance in Mixed 2D-3D Multi-Display Environments

Abhijit Karnik[1], Tovi Grossman[2], and Sriram Subramanian[1]

[1] Department of Computer Science, University of Bristol, Bristol, UK
{karnik,sriram}@cs.bris.ac.uk
[2] Autodesk Research, 210 King St. E., Toronto, Ontario, Canada
firstname.lastname@autodesk.com

Abstract. Stereoscopic displays and volumetric 3D displays capable of delivering 3D views have in use for many years. These standalone displays have been investigated in detail for their impact on users' viewing experiences. Effects like aesthenopia and nausea are well-known for flat-screen based stereoscopic displays. However, these devices have not been tested in the context of multi-display environments (MDEs). The performance cost of repetitive switching between a 3D (stereo or volumetric) display and a standard 2D display are not known. In this paper, we perform a thorough user study where we investigate the effects of using such 3D displays within the context of a MDE. We report on our findings and discuss the implications of the same on designs involving such hybrid setups. Our experiments show that in the condition involving two 2D displays which allow for motion parallax and perspective correction, the participants performed the task the fastest.

Keywords: stereoscopic display, autostereoscopic display, volumetric display, zone of comfort, multi-display environment, performance, mental load.

1 Introduction

Multi-display environments (MDEs) combine multiple display elements into a single coherent system. Such systems have been explored in different combinations of tabletop, wall and hand-held setups delivering 2D content as well as non-stereoscopic content. Of late, with the increase of availability of stereoscopic 3D displays (stereo-3D), such devices are also becoming part of MDEs [14, 16, 23]. Another class of 3D displays, which we term as spatial 3D displays, display true depth and are inherently multi-view and autostereoscopic [10, 25, 38]. Both stereo-3D and spatial 3D displays allow an interesting case for MDEs which allow mixed content delivery.

There are demonstrable advantages of stereo-3D and spatial 3D displays in terms of perception of 3D digital content [13, 34] which been studied in detail. However, there is an associated cost with the use of these devices especially stereo-3D. Prolonged use of stereo-3D has been associated with aesthenopia [3] as a combination of blurred vision, headaches, fatigue, nausea and pain. These symptoms are associated with visual-vestibular conflict and vergence-accommodation conflict.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 260–277, 2013.

Visual-vestibular conflict arises when stereoscopic content is meant to simulate great depth and movement, such as in cinemas. This triggers the brain into assuming motion of the body. However the vestibular organs (in the inner ear) which detect physical motion indicate that the body is still. This results in the effect termed as visual-vestibular conflict. This conflict is less pronounced in desktop and office environments which are physically smaller (than cinema screens) and also afford other environmental clues pointing to the lack of motion. Thus this effect is not considered within scope of this paper.

Vergence-accommodation conflict is more important with the display sizes relevant to MDEs. Normally, human eyes accommodate (rotate inwards or outwards) such that the lines of sight intersect on an object of interest and the focus is adjusted to the same location. However with stereo-3D displays, there is a disparity between the focal plane and the perceived location of the object. This results in the effect termed as vergence-accommodation conflict.

Vergence-accommodation conflict is relevant to MDEs using stereo-3D displays. While fatigue is well-reported for continuous use of stereo-3D displays, a relatively unexplored area is what impact vergence-accommodation conflict has on performance. As expected in a task spanning across a standard 2D display and stereo-3D display, the user would have to switch context between the two devices on a regular basis. Would this context switch aggravate symptoms resulting from vergence-accommodation conflict?

While stereo-3D displays are known to have issues with vergence-accommodation conflict, spatial 3D displays usually don't suffer from such problems. This makes them ideal for tasks involving localized 3D content while delivering realistic 3D views. However, as a part of MDEs, these devices can also impose a performance penalty as the user has to switch between a virtual 2D view and a realistic 3D view.

In the often cited example of using 3D visualization for air traffic control, the controller may be forced to switch between a 3D visualization of the air traffic to a 2D view listing weather conditions or information about inbound aircrafts. If the hybrid nature of the setup affects the performance of the controller in any form, such effects need to be studied. Thus an evaluation of performance becomes critical if such MDEs are to become part of day to day use. Motivated by this, we performed an experiment involving a 3D task involving three scenarios. The experiment aims to answer the question: "What is the effect of repetitive switching between a standalone 3D display and a 2D display during a task involving content spread across both?"

We perform a study that uses a mental rotation task to investigate the effects of using a 2D display in conjunction with either a spatial 3D or a stereo-3D display. Three conditions are studied. In the first condition, we pair a 2D display with another 2D display that supports motion parallax and perspective correction. The second condition involves a 2D display and a stereo-3D display. The last condition uses a swept volumetric display (a type of spatial 3D) paired with the 2D display.

Thus the main contribution of this paper is a systematic investigation of the effects of a hybrid MDE on user performance for a 3D data intensive task. Our experiments show that in the condition involving two 2D displays which allow for motion parallax and perspective correction, the participants performed the task the fastest.

However they also achieved higher accuracy when using the stereo-3D display. Finally, the spatial 3D condition had lowest scores for time as well as accuracy. We conclude by offering some possible explanations for these outcomes.

2 Related Work

We consider three different aspects of related research in this section. We wish to explore 3D displays in the context of MDEs, so it is pertinent to explore the MDE literature. Since spatial 3D display based MDEs are not commonly known, we also look at the literature around standalone 3D displays separately. Finally we explore literature related to the cognitive effects that 3D displays have on users.

2.1 Multi-Display Environments

Multi-display environments (MDEs) that combine interactive tables with wall-mounted displays provide users with enhanced visualization and interaction capabilities. Such setups have been around for a while now. Earlier examples like VIP [1] and ImmersaDesk [7] have demonstrated that multiple views of a task on different projection planes enhances user experience. Similarly, MDEs have been shown to be useful for a range of tasks such as geospatial applications [11], biomolecular modeling [2] and astronomy [40]. Most of these MDEs only explored the combination of planar 2D displays making them 2D display-based MDEs.

With these setups, 3D data is displayed by 3D rendering on a 2D surface [4, 11, 18]. The '3D content' is non-stereoscopic in nature. With user tracking systems (e.g. Kinect) it is possible to provide motion parallax as well as perspective correction [29] for the 3D content. While planar devices are capable of delivering a rich rendering of 3D views, they lack the realism delivered by displaying true 3D objects in terms of accuracy of depth estimation [13] and orientation [15].

More recently, MDEs with inbuilt stereoscopic capabilities have been demonstrated. Systems such as Toucheo [14], Holodesk [16] and PiVOT [23] are capable of generating stereoscopic views collocated with 2D views thus providing a composite MDE that is capable of delivering mixed content. In case of Toucheo and Holodesk, the 2D content appears spatially below the 3D content. With PiVOT the content is collocated but accessed by leaning forward or back. While these are special examples of such MDEs, a simpler example would be one that involves a desktop setup where one display is a 2DD and another is either a stereo-3D or a spatial 3D display. These desktop setups are currently feasible given the availability of 3D monitors and could present the mixed content side by side.

2.2 Standalone 3D Devices

Standalone 3D devices fall into two broad categories: Planar stereo-3D displays and spatial 3D displays. With stereo-3D displays, the binocular disparity is generated by delivering two different views to the user's eyes looking at a static planar screen.

Examples of stereo-3D displays include ones using lenticular arrays [28], microlens arrays, parallax barriers [33] or a hybrid combination of these [26]. While there are some glasses-free stereo-3D displays [22], the commercially available state of the art relies mainly on shutter-glass based systems. All these devices work by providing different views to each eye of the user. Thus vergence-accommodation conflict affects all these displays.

On the other hand spatial 3D displays generate views such that the visualized object has real spatial depth and dimensions. To achieve real spatial depth, the relevant points in the volumetric space are turned into point sources of light. The relevant points are representative of the reflective surface(s) of the object(s) allowing perception by the eye. Different methods have been utilized to solve the problem of lighting the volumetric pixel points (voxels). A stack of static but sequentially switched diffusers achieve the true 3D effect in DepthCube [38]. Other approaches apply different physical properties like plasma bubbles generated by a pulsed laser as shown by AIST, Japan, laser-triggered fluorescence [8] and laser-induced damage glass [30]. The swept diffuser technique used by LightField [21], Vermeer [5], Perspecta [10] and its anisotropic implementation [6] with a view-point driven autostereoscopic view have also been demonstrated.

2.3 Visual Comfort in 3D Setups

It is necessary to first make a case in support of 3D displays (both stereo-3D and spatial 3D) as standalone devices. Price and Lee [34] have shown that performance in spatial cognitive tasks for students improves with stereoscopic imagery. Also, Jin et al. [19] found that stereo-3D provided an advantage when presenting complex structures and spatial relationships. A study involving volumetric displays by Grossman and Balakrishnan [13] showed that volumetric displays (i.e. spatial 3D displays) can provide better depth perception in some tasks, in comparison to stereo-3D displays.

However, since visual discomfort arising from vergence-accommodation conflict is well known for stereo-3D displays, extensive research work has investigated these effects in a purely single display context. Kooi and Toet [24] explored how binocular disparity affects viewing comfort while Tam et al. [39] explored the visual discomfort with respect to a 3D TV setting. The dynamic accommodative response to stimuli corresponding to stereo-3D display was studied by Oliveira et al. [31] showing effects due to vergence-accommodation conflict. Emoto et al. [9] showed that repeated vergence adaptation leads to decline in visual functions. Lastly, work by Shibata et al. [36] and Hoffman et al. [17] have explored as to how the visual performance degrades while working with stereo-3D displays.

While these visual fatigue effects associated with stereo-3D displays have been shown in detail by prior research, there is little literature regarding effects on cognitive load when a 3D display is used in tandem with a 2D display. Paas et al. [32] describes that cognitive load can be measured through three properties: mental load, mental effort, and performance. They also mention that mental load and mental effort are more difficult to measure since they involve the use of secondary tasks. However, performance can be measured in terms of item accuracy and completion time. Thus, a task can be designed to measure performance and extrapolate it to cognitive load resulting from a particular setup. This leads us to our experiment.

3 Experiment

Our experiment is used to determine the impact that integrating 3D displays into MDE environments has on a user's viewing experience, visual comfort, and task performance. While there are multiple options of configuring a 3D MDE, we consider a desktop configuration with two displays side by side; as such systems could be easily adopted into existing workplaces today. We studied three possible display combinations, as described in the experiment conditions below.

3.1 Conditions

It has already been established that 3D images, irrespective of display type, are better for performing shape understanding tasks [37]. Instead, our goal is to determine the impact of hybrid 2D-3D display environments. The following three display combinations were used:

- **2D-2D:** In this condition the first display was a static 2D display. The second display was also a 2D display, but head tracking was used to present perspective corrected views that would also respond to motion parallax. Thus the view would be regenerated based on the head-position of the user and would seem three dimensional whenever the user moved their head.
- **2D-3D:** In this condition, we combined a static 2D display (as described above), and a stereo-3D display. The stereo-3D display also used head tracking to provide perspective corrected views.
- **2D-VO:** This condition combined the static 2D display with a 3D volumetric display.

While an additional condition 3D-VO would be possible, it was not considered since our assumption is that in the MDE setup, the task always has a 2D display element.

3.2 Task

We wished to identify a task which presented a significant amount of cognitive load on the participant and also required frequent switching between the two displays. We chose a modified form of the Shepard-Metzler Mental Rotation test (SMT) [35] as the experimental task. The ability to rotate two and three-dimensional objects in the mind is known as mental rotation. The SMT is used to test the ability of a participant to accurately and rapidly mentally rotate three dimensional objects. The original SMT used two images each containing a 3D shape made up of cubes connected at the face. The shape in the second image is either a) the same 3D shape but rotated along one of the 3 axes or b) the mirrored (along one of the 3 axis) version of the 3D shape and then rotated.

In our task, the shape is made up of 1 unit diameter spheres connected to each other as a chain. The shapes that were select satisfied the criteria that they were three dimensional (having 3-4 non-coplanar 90° bends) and fit inside a 5×5×5 grid. The number of spheres per shape ranged from eleven to seventeen. Like the original SMT, the task was to identify if the two shapes presented on the displays were same or mirrored. Our task differs from the original SMT in terms of rotation. While the original SMT rotates the shapes along one axis, in our task, the rotation can happen along all three axes at the same time. We defined two difficulty levels, easy and hard. With the 'easy' level, the one shape is rotated by <30° along a random axis as compared to the other shape. With the 'hard' level, the second shape was rotated along all three axes such that the sum of the rotations was >60°. Thus, based on *match* state (*same* or *mirror*) and *difficulty level* (*easy* or *hard*), four different combinations (as shown in Figure 1) were possible.

The participants had to identify if the shapes were same or mirrored and indicate their answer via key presses. The detailed process is described in Section 3.5 (Procedures).

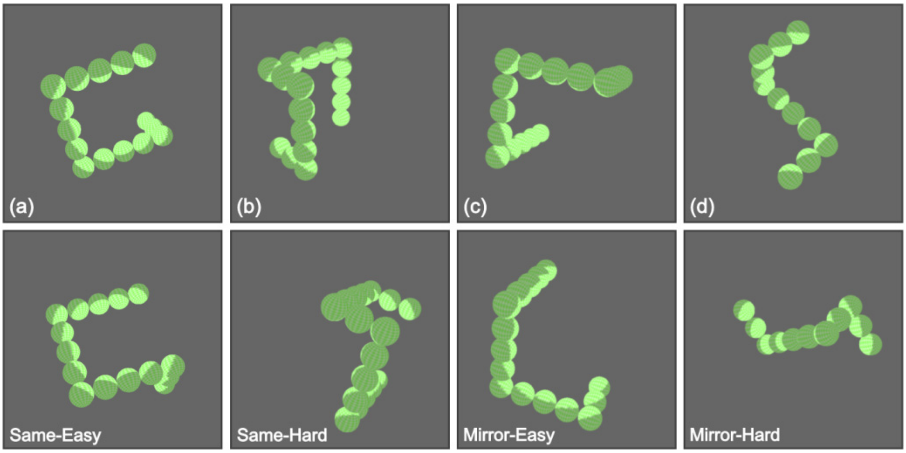


Fig. 1. Shapes used for the task. Top row shows the four shapes used. Bottom row shows the paired shape with *difficult level* (*easy* or *hard*) and *match* state (*same* or *mirrored*).

3.3 Apparatus

The setup consisted of three displays placed next to each other in front of a plain background.

1. **2D display:** This was a Dell 21" 1920×1080 pixel monitor. Usually, the shutter glasses used for the 3D TV interact with some 2D monitors. This causes the screens to appear black through the glasses. The monitor we used did not get affected in the same way.
2. **Stereo-3D display:** A 40" (1920×1080 pixel) Samsung UN40ES6500 3D TV was used as the stereo-3D display. The display operates in a side-by-side mode for 3D thus allowing an effective resolution of 960×1080 pixels for 3D mode using active shutter glasses.

3. **Volumetric 3D display:** A Perspecta display [10] from Actuality Systems was used as the volumetric display. Perspecta is a swept volumetric display with a resolution of 100 million voxels and a 10" spatial display diameter. It is one of the few spatial 3D displays that was ever available commercially, and has been used for numerous other experiments [6, 12-13].

The first two displays were connected to an Intel Core i7 machine running Windows 7. The Perspecta display was connected to a standalone Windows XP machine. The two machines were networked together for exchanging experiment state information. For tracking the user's head-pose, we used NaturalPoint's Optitrack Duo. The Optitrack Duo uses a marker constellation to provide spatial position of the tracked object along with its orientation in space with sub-millimeter accuracy. The head-pose information was used to present motion parallax such that the users could 'look around the corners' of the 3D shapes used for the experiment. Head-pose information was received by the master program on the first machine via VRPN. The master program intercepted the user's inputs and communicated updates to the slave program running on the second machine via OSC messages. The setup is shown in Figure 2.

The three displays have very different physical dimensions. Thus it was necessary to ensure that the field of view (FOV) coverage of voxels of Perspecta should be comparable or similar to the FOV coverage of the pixels of both the stereo-3D and the 2D display. The actual positions of the shape and the pixel dimensions of the shapes were adjusted on the stereo-3D and the 2D displays to match that of Perspecta. In each case the effective physical dimension of the shapes was 9" and they were all aligned horizontally, spaced 25" apart. The shapes shown on the stereo-3D display were always within the Percival zone such that the experiment was run within the specifications of Shibata's results [36]. Also since the Perspecta requires low light operation, the whole experiment was run in a darkened room.

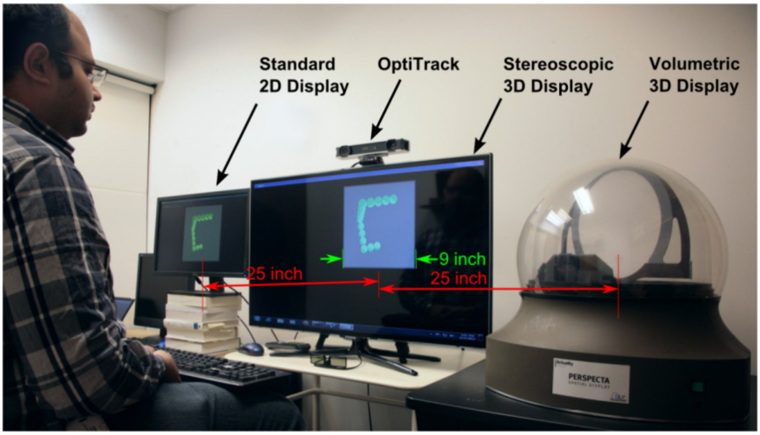


Fig. 2. Experimental setup: All three shapes (one on each display) were aligned horizontally and had the same physical dimensions

3.4 Participants

A total of twelve participants (7 male, 5 female) were recruited from a local university and via Craigslist. Participants' ages ranged from 18 to 40 years (nine participants from the 21-30 years age-group). All had normal stereo-acuity which was verified through the Titmus-Wirt Fly test. Participants who regularly wore optical correction (5) wore the correction during the experiment (2 wore contact lenses, 3 wore spectacles). All subjects had no prior experience with the task. Also none of them had any experience with Perspecta and had never used a stereo-3D display in a work setting. The participants were compensated for their time with a gift voucher.

3.5 Procedure

For the 2D-2D condition, the 2D display displayed one of the shapes as a fixed 2D shape. This shape did not respond to the user's head movement. The stereo-3D display was operated as a 2D display and it showed the second shape as a single flat 2D shape. This shape responded to the user's head movement allowing motion parallax based viewing. For the 2D-3D condition, the second shape was displayed as a stereo-3D shape while still allowing for motion parallax. The participant had to wear the shutter glasses to view the shape correctly. For both conditions, the program reoriented the shape to a perspective correct orientation thus giving a sensation of 3D. For the third condition (2D-VO), given the physical arrangement of the three displays, the stereo-3D display was operated as a normal 2D display and the first immobile 3D shape was displayed on it. The second shape was displayed on Perspecta. Since Perspecta is autostereoscopic, there was no need to use the head-pose information to reorient the shape.

Before performing the experiment, the participant's stereo-acuity was confirmed by the Titmus-Wirt Fly test. We intended to reject participants who failed the test, however all the participants passed the test. After the test, the participants were acclimatized with setup (especially Perspecta) and then the task was explained. The participants were seated 30" from the display plane. They were encouraged to move their head right and left but asked to limit the motion towards or away from the displays. The actual experiment consisted of 2 phases. During the initial phase, consisting of 20 trials, the users were allowed to get used to the experimental procedure. Data from this phase was discarded. The second phase was the actual experiment. The participants were asked to perform the task as quickly as possible and were made aware that the accuracy of their answers was also being recorded.

For a single trial, two shapes would be displayed on the two displays (relevant to the test condition). To clearly demarcate switching of context between the two devices, only one display would should its shape at any given time. To switch to the other shape, the participant had to press the spacebar. The participant had to indicate the match state (same or mirror) through a single key-press. Once an answer was given, the experiment moved onto the next trial. Since a minimum of one switch would be required to see both the shapes, the program would not allow the experiment to move to the next trial if the answer was indicated without a single switch. This prevented participants from accidentally or intentionally skipping through the trials.

The experiment consisted of a total of 300 trials divided into blocks of 100 trials per condition per participant. For a single block of 100 trials, there were 50 *hard* and 50 *easy* trials. Also, the block had 50 *same* and 50 *mirror* shape pairs. The order of *easy* and *hard* trials was randomized to prevent monotony. The participants were not allowed to pause between trials but were allowed to take a 10 minute break between conditions. The participants filled out questionnaires before and after each block and at the end of the experiment. The learning and order effects were counterbalanced by changing the order of the conditions per participants using a Latin square design.

3.6 Measures

The experiment was run as a within-subjects design for the three conditions being compared. For the task metrics, *difficulty level* was used as an additional variable. The following data was collected from the participants:

Task Metrics.
We measured four details per trial. The number of *switches* between the two displays was logged along with the answer given by the participant. *Accuracy* of the answer was binary, either right or wrong. The *total time* taken to perform each trial was also logged. Since it was possible to differentiate between the time spent on one display versus the other (only one display was active at a time), the time spent viewing the 2D shape and the time spent viewing the 3D shape were logged separately. We wished to investigate if the participants spent more time on either the 2D display or the other display. So *time non-2D* metric was computed as the percentage time spent on the non-2D display (100×Time spent on non-2D/Total time). The time metrics were recorded in milliseconds with the accuracy derived from the system clock.

Head Pose.
The head-position information was being used by the program to render perspective correct views for the 2D-2D and the 2D-3D conditions. However, we also recorded the head-pose (as to where the head was pointed) as another parameter. Given the distance between the two displayed shapes (25”), simple saccadic motion of the eyes was not enough and the participants resorted to turning their heads to view the displayed shapes. This turning of the head could be detected by the tracking system and could be sampled 60 times a second. We used this to generate heat maps of where the participant was looking and for how long.

Questionnaires.
We used three different questionnaires during the experiment. Similar to Shibata et al. [36] we wished to record any occurrence of symptoms usually associated with stereo-3D displays. We used a *symptom questionnaire* adapted from their study as shown in

Figure 3 left. Since each participant could have a different starting symptom state, we administered a pre-trial questionnaire and a post-trial questionnaire for each condition. The questionnaires were answered on a digital form requiring the users to click on the desired answers. We also administered a NASA TLX questionnaire after the end of each block. The NASA TLX questions gauged the mental demand, physical demand, pace of task, perceived success, effort and irritation for each block.

The last questionnaire was a *ranking questionnaire* (as shown in Figure 3 right). We asked the participants to rank the three conditions as per their preference for each question. The participant answered this five-question questionnaire after the completion of the experiment. The first three questions measured the participants' perception about the three conditions as compared to each other for the symptoms of fatigue, eye irritation and headache. The last two questions measured the participants' preference for a particular condition and the associated easiness enabled by the condition.

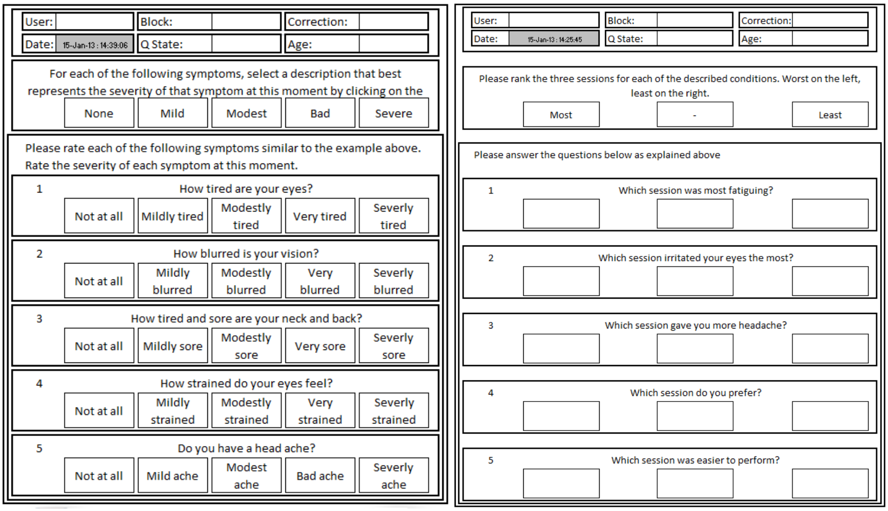


Fig. 3. Questionnaires used for evaluation of symptoms and task load. (left) Symptom questionnaire (right) Final ranking questionnaire. Both tests were administered digitally.

3.7 Results

The data logged for the test and collected from the questionnaires was analyzed with SPSS 19. These are presented below.

Task Metrics.
For the task metrics, the experiment presented as a within-subjects repeated measures design with two independent variables as the *conditions* (3 groups: 2D-2D, 2D-3D and 2D-VO) and *difficulty level* (2 groups: *Easy* and *Hard*). We averaged the results (except for accuracy, which was summed) per user for each *condition* and then

analyzed the results through two-way repeated measures ANOVA with Bonferroni correction. We looked at the main effects of *condition* and *difficulty level* as well as interaction between the two. We expected that the main effect of *difficulty level* would present as lower accuracy and higher time as well as switches for the *hard* trials.

1. *Accuracy*: The accuracy metric measured if the participants correctly identified the trail pair of shapes to be mirrored or the same. A maximum score of 50 was possible for each combination of *condition* versus *difficulty level*. There was a significant main effect of *condition*, $F_{(2, 22)} = 8.51$, $p<0.005$. The 2D-3D condition had highest average accuracy while the 2D-VO condition fared the worst. As expected, there was a significant main effect of *difficulty level*, $F_{(1, 11)} = 6.34$, $p<0.05$, with *easy* trials having higher average accuracy. Also, there was a significant interaction between *condition* and *difficulty level*, $F_{(2, 22)} = 5.84$, $p<0.05$. The results are shown in Figure 4 left.
2. *Switches*: The metrics were calculated for the average number of switches performed by user per combination of *condition* and *difficulty level*. There was no significant main effect of *condition* on the number of switches. However there was a significant main effect of *difficulty level*, $F_{(1, 11)} = 9.28$, $p<0.05$. The participants performed more switches for *hard* trials as compared to *easy* trials. There was a significant interaction between *condition* and *difficulty level*, $F_{(1.24, 13.66)} = 8.2$, $p<0.05$. The results are shown in Figure 4 right.

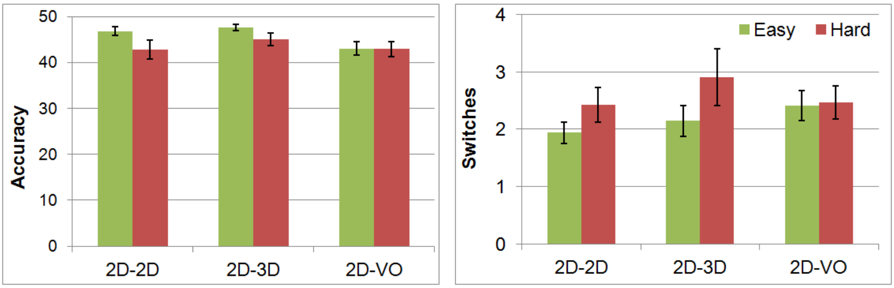


Fig. 4. Task metrics analysis with standard error-bars. The green bars for each condition correspond to the Easy task and the red for Hard task. (left) Average accuracy achieved by participants with standard error-bars. Maximum possible score was 50. (right) Average number of switches performed before arriving at the answer.

3. *Total time*: The average of total time taken per trial was used for analysis across the six combinations. We did not find any significant difference between *conditions* for total time thus implying a lack of significant main effect. For *difficulty levels*, however, we found that there was a significant main effect, $F_{(1, 11)} = 7.37$, $p<0.05$. We also found a significant interaction between *condition* and *difficulty level*, $F_{(1.33, 14.61)} = 5.12$, $p<0.05$. Testing for within-subjects contrasts, we found significant difference between 2D-VO and 2D-2D *conditions*, $F_{(1, 11)} = 10.54$, $p<0.05$. The results are shown in Figure 5 (right).

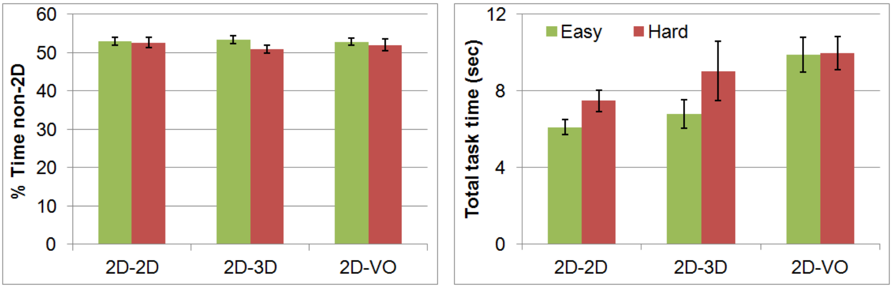


Fig. 5. Task metrics analysis with standard error-bars. The left bars for each condition correspond to the Easy task and the right for Hard task. (left) Average of percentage of time spent by participants looking at the non-2D display. (right) Average of total time in Seconds taken by participants per trial.

4. *Time non-2D*: We found no significant difference between the three conditions for the amount of time spent on non-2D display. Similarly, no significant difference was observed for main effect of *difficulty levels* as well as interaction between *condition* and *difficulty levels*. The results shown in Figure 5 (left) show near similar averages across all combinations.

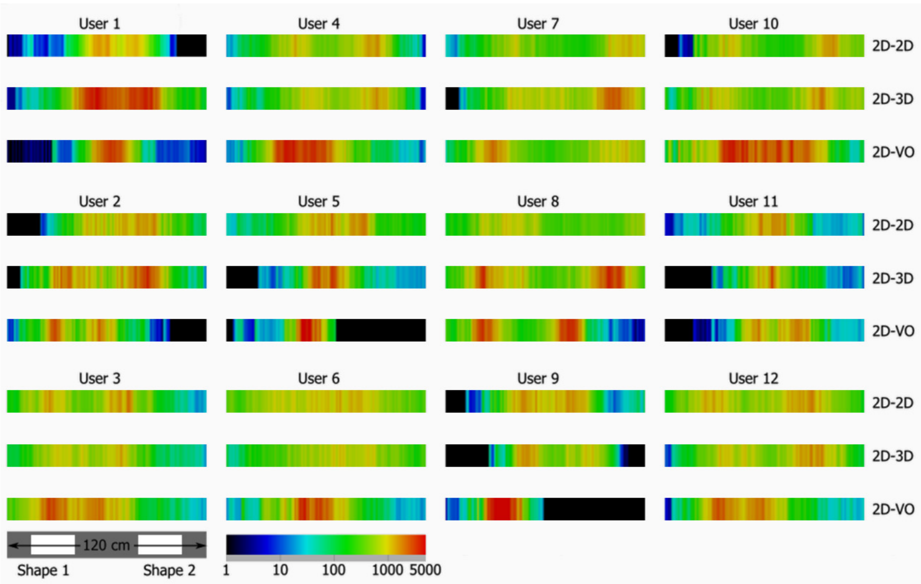


Fig. 6. Heat maps showing how much time a participant spent looking at what part of the screens. Color scale is logarithmic and clamped at 5000 samples.

Heat Maps.
The head-pose was available as a 3D spatial coordinate of the tracker and a local rotation of the tracker. Since the tracker was mounted on the participant's forehead, it gave a good representation of where the user was looking. The participants did look down at the keyboard when they had to enter their answer. The head-pose records related to these events (which showed the tracker orientation almost parallel to the ground) were removed from analysis. The remaining records were used to generate the heat maps. To compare the head-pose characteristics of different participants, we chose to only represent the horizontal location of the head-pose for drawing the heat map as the x-axis of the map. The tracker provided 60 samples/second and the number of samples at each specific location were used to draw the map. As shown in Figure 6, the x-axis shows the 120 cm region centered on the two shapes. The color scale is a logarithmic scale and is clamped at 5000 samples.

Questionnaires.

NASA TLX.
The NASA TLX questionnaires were administered for each condition after the trial block ended. Hence, we did not look for differences due to *difficulty levels* and were looking for effects of the *conditions* only. The questionnaire data was analysed using Kruskal-Wallis test. For all questions excluding the first one (How mentally demanding did you find the task?), we found no significant difference between the three conditions. For *Mental Demand*, there was a significant difference, $H(2) = 6.87$, $p < 0.05$. Mann-Whitney tests were used to follow up this finding (3 pairs) with Bonferroni correction, (significance level at *0.0167*). The *Mental Demand* was not significantly different between 2D-2D and 2D-3D ($U = 71$, $r = -0.01$) as well as 2D-3D and 2D-VO ($U = 37.5$, $r = -0.41$). However, *Mental Demand* for 2D-VO as compared to 2D-2D was significantly higher ($U = 29$, $r = -0.51$).

Symptom Questionnaire.
The symptom questionnaire results were analyzed separately for each condition and question using Wilcoxon signed-rank test. We found no significant change in symptoms for *Vision clarity* and *Headache* for any of the conditions. Also, for *Eye Tiredness* there was no significant change for the conditions 2D-2D and 2D-3D. However, we found significant increase in *Eye Tiredness* for 2D-VO, $z = -2.236$, $p < .05$, $r = -0.46$. Similarly, we found that the participants reported a significant increase for *Neck and Backache* for 2D-3D, $z = -2.0$, $p < .05$, $r = -0.41$ and 2D-VO, $z = -2.236$, $p < .05$, $r = -0.48$. Lastly, we found significant increase in *Eye Strain* for all three conditions: for 2D-2D, $z = -2.24$, $p < .05$, $r = -0.48$; for 2D-3D, $z = -2.65$, $p < .05$, $r = -0.54$ and for 2D-VO, $z = -2.07$, $p < .05$, $r = -0.42$.

Ranking Questionnaire.
The ranking questionnaire results are presented in Figure 7 (right). For *headache*, the participants consistently ranked 2D-VO condition as the worst (a lower average rank). This does not tie in well with the symptom questionnaire where we found no significant

difference between the three conditions as well as the verbal feedback of the participants stating that they didn't get a headache from any of the conditions. For overall *fatigue*, the 2D-VO condition again was ranked the worst while 2D-3D condition was ranked the best. For the remaining questions (*Eye Irritation*, *Easier* and *Preferred*), the results were tied. We did not see a clear trend of preference for any of these questions. All participants ranked *Preferred* for the three conditions in exactly the same order as they ranked the conditions with respect to *Easier*.

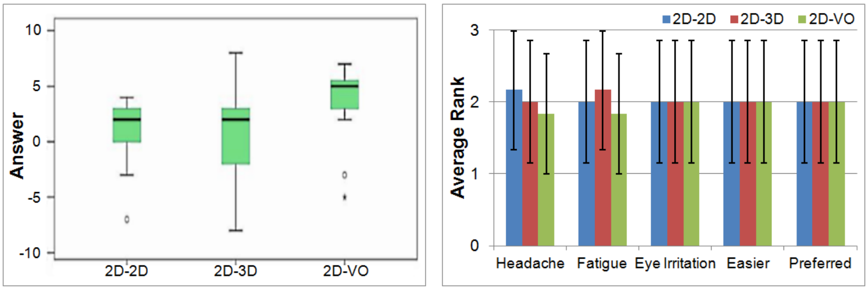


Fig. 7. Questionnaire analysis. (Left) Boxplot for NASA TLX *Mental Demand* shows higher indications for 2D-VO condition. (Right) Ranking questionnaire analysis.

4 Discussion

4.1 Interpretation of Results

2D-2D Condition.
On average, participants performed less switches within the 2D-2D condition. When factored for *difficulty level*, the 2D condition had higher accuracy than the 2D-VO condition. The participants also took less time to complete the task.

The 2D-2D condition also presented the least change in symptoms. Only the *Eye Strain* symptom was aggravated by the condition, but again a similar change was seen in the remaining two conditions. *Mental demand,* as measured by NASA TLX, was significantly lower than 2D-VO. In the overall ranking, 2D-2D was ranked lower for *headache* and *fatigue.*

While preference scores were tied, the 2D-2D condition performed better and caused fewer symptoms. Thus, it is possible that 3D cues provided by simple motion parallax are somewhat better suited than those afforded by stereo-3D. This is comparable to the results of Johnston et al. [20] where multi-frame motion parallax is shown to work better than stereopsis.

2D-3D Condition.
The 2D-3D condition has the best accuracy as compared to all the other conditions. As per the *final questionnaire*, the 2D-3D condition was ranked the least in terms of *fatigue*. However on average, participants performed more switches and took longer

than the 2D-2D condition. The condition also aggravated the symptoms of *Eye strain* and *Neck & Backache*.

Thus we can say that the 2D-3D condition has an advantage in case of tasks where accuracy is crucial. Our participants reported it to be less fatiguing but also agreed with the previously observed results indicating *eye strain* resulting from vergence-accommodation conflict. It is important to note that in our setup the objects did not present extreme negative or positive parallax. This could be one of the reasons as to why other symptoms were not reported. Our results are in line with Shibata et al.'s [36] findings. We can conclude that for the 2D-3D condition, if we limit extreme parallax we gain on accuracy with lesser fatigue.

2D-VO Condition.
The outcomes of the experiment suggest that 2D-VO condition fares worst in terms of accuracy and average task time. Even if we only consider the *hard difficulty level*, we see that the results are just marginally better than the 2D-2D condition for accuracy and still worse for average task time. From the NTLX scores for *mental demand*, we find that it has higher values than the other two conditions. These results are unexpected. We assumed that a more realistic representation of objects would help recover more information and hence help accuracy.

4.2 Implications

A general and direct implication of these results is that an MDE consisting of a spatial 3D display and a 2D display should be avoided for spanned tasks involving high cognitive load. This was contrary to our initial expectation that 2D-VO condition would be significantly better than the rest. However, as pointed out by Grossman and Balakrishnan [13], there are a few mitigating factors for poor performance of a spatial 3D display. Even as of today, the display quality of such devices is not at par with that of 2D and stereo-3D displays. There are artefacts in the display (for e.g. the central spine of the display cannot show any information and colour quality) which can influence the results. However, it is also important to note that in our case, the experimental setup was designed such that these effects were minimized. The visual size of the output of all three displays was matched and there were no extra visual cues provided by the shapes shown on the 2D display or the stereo-3D display.

When compared to the 2D-2D condition, it is possible that there are other focus based factors affecting the performance for 2D-VO condition. For the side-by-side 2D displays, the user has a fixed reference to a focal point on the plane of the display. However with a spatial 3D display, there is no central plane and thus no central point of focus. This can add to mental load when there is switching between the displays.

Also as the visualization of the shape in true 3D makes it look more real, it is possible that the switching process becomes one where the user has to switch contexts (from virtual world to real world) and they potentially do not view the two displays as a part of the same system. Such a situation also arises when the user has to switch focus between a physical object and a virtual object. Surprisingly, we could not find any research that investigates performance effects while comparing a purely virtual context to one with mixed context. The closest work is in the tangible literature by Marshall et al. [27] where they cautiously suggest that in a single user instance, a tangible interface is not necessarily better. Thus further investigation in this regard is warranted.

It is also possible that users find it difficult to compare a true 3D shape with a 2D shape that the first display shows. We refrained from using perspective correction on the 2D shape for the 2D-VO condition as the static 2D served as a common control shape to all three experimental conditions.

Lastly, as a recommendation for selection of 3D display elements for MDEs, we feel that there is a possible benefit of amalgamation of the 2D-2D and 2D-3D condition. The 2D-2D condition allowed perspective corrected views and motion parallax for the second shape. Comparing its accuracy results with 2D-3D, the overall results for symptoms and NTLX we can argue that this may be an ideal configuration for prolonged use tasks. With the availability of low cost desktop based head tracking systems, it might be beneficial to have a stereo-3D display which is operated mainly in 2D mode but allow perspective corrected views for presenting 3D. Only when the task involves high density of 3D elements, the device can switch to stereoscopic mode thus adding binocular disparity as another cue.

4.3 Future Work

We do not believe that a spatial 3D display is unsuitable for MDE setups. For tasks similar to our experimental task, our results hold true. However there may be other tasks with minimal cross-device contexts wherein a spatial 3D display may prove to be more beneficial. Future work could be used to explore the impact which the task has on our results. Furthermore, effects like aesthenopia and nausea were not a major factor in any of our conditions. However, with prolonged use, the effects may become most prevalent in the 2D-3D condition. Another factor of our tested MDE setups is that they all contained only 2 displays. In the future, it may be interesting to understand the impact of having a greater number of displays, or including displays with larger form factors and with arrangements that favor the type of 3D display used.

5 Conclusion

We have investigated the performance cost of repetitive switching between a 3D (stereo or spatial) display and a standard 2D display in context of a MDE. The experimental results prove that there is a cost involved with the scenarios involving a 2DD and a spatial 3D display which is higher than other scenarios. The results should provide a guideline for the design of MDEs utilizing either spatial 3D or stereo-3D elements.

Acknowledgments. We thank Daniel Wigdor and DGP Toronto for helping out with the participant recruitment. We also thank Anne Roudaut for her inputs. This work is supported by the European Research Council (Starting Grant Agreement 278576) under the Seventh Framework Programme.

References

1. Aliakseyeu, D., Martens, J.-B., Sriram Subramanian, M., Vroubel, W.W.: Visual Interaction Platform. In: Proc. INTERACT 2001, pp. 232–239. IOS Press (2001)
2. Bailly, G., Nigay, L., Auber, D.: NAVRNA: visualization - exploration - editing of RNA. In: Proc. AVI 2006, pp. 504–507. ACM, New York (2006)
3. Berezin, O.: Digital cinema in Russia: Is 3D still a driver for the development of the cinema market. 3D Media 2010 (2010)
4. Bowman, D.A., North, C., Chen, J., Polys, N.F., Pyla, P.S., Yilmaz, U.: Information-rich virtual environments: theory, tools, and research agenda. In: Proc. VRST 2003, pp. 81–90. ACM, New York (2003)
5. Butler, A., Hilliges, O., Izadi, S., Hodges, S., Molyneaux, D., Kim, D., Kong, D.: Vermeer: direct interaction with a 360° viewable 3D display. In: Proc. UIST 2011, pp. 569–576. ACM, New York (2011)
6. Cossairt, O.S., Napoli, J., Hill, S.L., Dorval, R.K., Favalora, G.E.: Occlusion-capable multiview volumetric three-dimensional display. Appl. Opt. 46, 1244–1250 (2007)
7. Czernuszenko, M., Pape, D., Sandin, D., DeFanti, T., Dawe, G.L., Brown, M.D.: The ImmersaDesk and Infinity Wall projection-based virtual reality displays. SIGGRAPH Comput. Graph. 31, 46–49 (1997)
8. Ebert, D., Bedwell, E., Maher, S., Smoliar, L., Downing, E.: Realizing 3D visualization using crossed-beam volumetric displays. Commun. ACM 42, 100–107 (1999)
9. Emoto, M., Niida, T., Okano, F.: Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television. Display Technology 1, 328–340 (2005)
10. Favalora, G.E.: Volumetric 3D Displays and Application Infrastructure. Computer 38, 37–44 (2005)
11. Forlines, C., Esenther, A., Shen, C., Wigdor, D., Ryall, K.: Multi-user, multi-display interaction with a single-user, single-display geospatial application. In: Proc. UIST 2006, pp. 273–276. ACM, New York (2006)
12. Grossman, T., Balakrishnan, R.: The design and evaluation of selection techniques for 3D volumetric displays. In: Proc. UIST 2006, pp. 3–12. ACM, New York (2006)
13. Grossman, T., Balakrishnan, R.: An evaluation of depth perception on volumetric displays. In: Proc. AVI 2006, pp. 193–200. ACM, New York (2006)
14. Hachet, M., Bossavit, B., Cohé, A., Rivière, J.-B.D.L.: Toucheo: multitouch and stereo combined in a seamless workspace. In: Proc. UIST 2011, pp. 587–592. ACM, NY (2011)
15. Hancock, M., Nacenta, M., Gutwin, C., Carpendale, S.: The effects of changing projection geometry on the interpretation of 3D orientation on tabletops. In: Proc. ITS 2009, pp. 157–164. ACM, New York (2009)
16. Hilliges, O., Kim, D., Izadi, S., Weiss, M., Wilson, A.: HoloDesk: direct 3d interactions with a situated see-through display. In: Proc. CHI 2012, pp. 2421–2430. ACM, NY (2012)
17. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. Vision 8 (2008)
18. Jiang, H., Wigdor, D., Forlines, C., Borkin, M., Kauffmann, J., Shen, C.: LivOlay: interactive ad-hoc registration and overlapping of applications for collaborative visual exploration. In: Proc. CHI 2008, pp. 1357–1360. ACM (2008)
19. Jin, Z.X., Zhang, Y.J., Wang, X., Plocher, T.: Evaluating the Usability of an Auto-stereoscopic Display. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 605–614. Springer, Heidelberg (2007)
20. Johnston, E.B., Cumming, B.G., Landy, M.S.: Integration of stereopsis and motion shape cues. Vision Research 34, 2259–2275 (1994)

21. Jones, A., McDowall, I., Yamada, H., Bolas, M., Debevec, P.: An interactive 360° light field display. In: Proc. SIGGRAPH 2007 ETech, p. 13. ACM, New York (2007)
22. Karnik, A., Mayol-Cuevas, W., Subramanian, S.: MUSTARD: a multi user see through AR display. In: Proc. CHI 2012, pp. 2541–2550. ACM, New York (2012)
23. Karnik, A., Plasencia, D.M., Mayol-Cuevas, W., Subramanian, S.: PiVOT: personalized view-overlays for tabletops. In: Proc. UIST 2012, pp. 271–280. ACM, New York (2012)
24. Kooi, F.L., Toet, A.: Visual comfort of binocular and 3D displays. Displays 25, 99–108 (2004)
25. Langhans, K., Bahr, D., Bezecny, D., Homann, D., Oltmann, K., Oltmann, K., Guill, C., Rieper, E., Ardey, G.: FELIX 3D display: an interactive tool for volumetric imaging. In: Proc. Stereoscopic Displays and Virtual Reality Systems IX, pp. 176–190. SPIE (2002)
26. Lin, Q., Qiong-Hua, W., Jiang-Yong, L., Wu-Xiang, Z., Cheng-Qun, S.: An Autostereoscopic 3D Projection Display Based on a Lenticular Sheet and a Parallax Barrier. J. Display Technology 8, 397–400 (2012)
27. Marshall, P., Rogers, Y., Hornecker, E.: Are tangible interfaces really any better than other kinds of interfaces? In: Proc. Tangible User Interfaces in Context and Theory Workshop, CHI 2007. ACM, New York (2007)
28. Matusik, W., Pfister, H.: 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. ACM Trans. Graph. 23, 814–824 (2004)
29. Nacenta, M.A., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., Gutwin, C.: E-conic: a perspective-aware interface for multi-display environments. In: Proc. UIST 2007, pp. 279–288. ACM, New York (2007)
30. Nayar, S.K., Anand, V.N.: 3D Display Using Passive Optical Scatterers. Computer 40, 54–63 (2007)
31. Oliveira, S., Jorge, J., González-Méijome, J.M.: Dynamic accommodative response to different visual stimuli (2D vs 3D) while watching television and while playing Nintendo 3DS Console. Ophthalmic and Physiological Optics 32, 383–389 (2012)
32. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychol. 38, 63–71 (2003)
33. Perlin, K., Paxia, S., Kollin, J.S.: An autostereoscopic display. In: Proc. SIGGRAPH 2000, pp. 319–326. ACM, New York (2000)
34. Price, A., Lee, H.-S.: The Effect of Two-dimensional and Stereoscopic Presentation on Middle School Students' Performance of Spatial Cognition Tasks. J. Sci. Educ. Technol. 19, 90–103 (2010)
35. Shepard, R.N., Metzler, J.: Mental Rotation of Three-Dimensional Objects. Science 171, 701–703 (1971)
36. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: Predicting visual discomfort with stereo displays. Vision 11 (2011)
37. John, M., Cowen, M.B., Smallman, H.S., Oonk, H.M.: The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks. Human Factors 43, 79–98 (2001)
38. Sullivan, A.: DepthCube solid-state 3D volumetric display. In: Proc. Stereoscopic Displays and Virtual Reality Systems XI, pp. 279–284. SPIE (2004)
39. Tam, W.J., Speranza, F., Yano, S., Shimono, K., Ono, H.: Stereoscopic 3D-TV: Visual Comfort. IEEE Trans. Broadcasting 57, 335–346 (2011)
40. Wigdor, D., Jiang, H., Forlines, C., Borkin, M., Shen, C.: WeSpace: the design development and deployment of a walk-up and share multi-surface visual collaboration system. In: Proc. CHI 2009, pp. 1237–1246. ACM, New York (2009)

Touching the Void Revisited: Analyses of Touch Behavior on and above Tabletop Surfaces

Gerd Bruder[1], Frank Steinicke[1], and Wolfgang Stuerzlinger[2]

[1] Department of Computer Science
University of Würzburg
{gerd.bruder,frank.steinicke}@uni-wuerzburg.de
[2] Department of Computer Science and Engineering
York University
wolfgang@cse.yorku.ca

Abstract. Recent developments in touch and display technologies made it possible to integrate touch-sensitive surfaces into stereoscopic three-dimensional (3D) displays. Although this combination provides a compelling user experience, interaction with stereoscopically displayed objects poses some fundamental challenges. If a user aims to select a 3D object, each eye sees a different perspective of the same scene. This results in two distinct projections on the display surface, which raises the question where users would touch in 3D or on the two-dimensional (2D) surface to indicate the selection. In this paper we analyze the relation between the 3D positions of stereoscopically displayed objects and the on- as well as off-surface touch areas. The results show that 2D touch interaction works better close to the screen but also that 3D interaction is more suitable beyond 10cm from the screen. Finally, we discuss implications for the development of future touch-sensitive interfaces with stereoscopic display.

Keywords: Touch-sensitive systems, stereoscopic displays, 3D interaction.

1 Introduction

Recent exhibitions and the entertainment market have been dominated by two different technologies: (i) (multi-)touch-sensitive surfaces and (ii) stereoscopic three-dimensional (3D) displays. Interestingly, these two technologies are orthogonal, as (multi-)touch is about input, whereas 3D stereoscopic display about output. Both technologies have the potential to provide more intuitive and natural interaction with a wide range of applications, including urban planning, architectural design, collaborative tabletops, or geo-spatial applications. First commercial hardware systems have recently been launched, e.g., [9], and interdisciplinary research projects explore interaction with stereoscopic content on 2D touch surfaces, e.g., [1, 2]. Moreover, an increasing number of hardware solutions provide the means to sense human gestures and postures not only on surfaces, but also in 3D space, e.g., the Kinect, the ThreeGear system, or Leap Motion [3]. The combination of these novel technologies provides enormous potential for a variety of new interaction concepts.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 278–296, 2013.

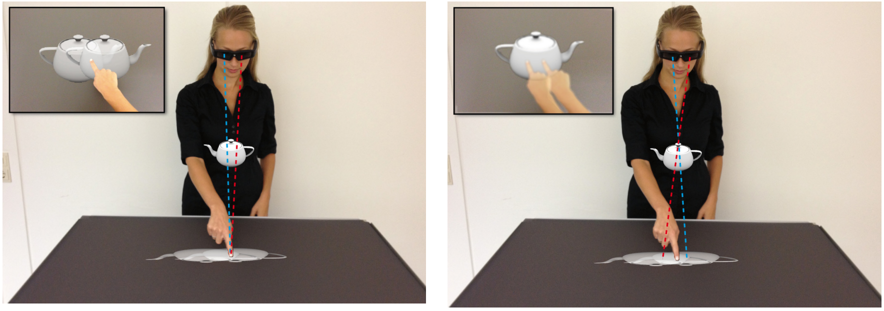


Fig. 1. Illustration of the main problem of touch interaction with stereoscopically displayed 3D data: (left) the user is either focused on her finger, which makes the selection ambiguous, or (right) on the object, which disturbs the visual perception of the finger

Until recently, research in the area of (multi-)touch interaction was mostly focused on monoscopically displayed data. For this, the ability to directly touch elements without additional input devices has been shown to be very appealing for novice as well as expert users. Also, passive haptics and multi-touch displays have both shown their potential to considerably improve the user experience [6]. Touch surfaces build a consistent and pervasive illusion in perceptual and motor space that the two-dimensional graphical elements on the surface can be touched. Yet, three-dimensional data limits this illusion of place and plausibility [33]. Such 3D data sets are either displayed monoscopically, which has been shown to impair spatial perception and performance in common 3D tasks, or stereoscopically, which can cause objects to appear detached from the touch surface [24, 31,7].

Stereoscopic display technology has been available for decades. Recently, it was revived due to the rise of 3D cinema, upcoming 3D televisions and 3D games. With stereoscopic displays, each eye sees a different perspective of the same scene through appropriate technology. This requires rendering of two distinct images on the display surface. When using stereoscopic technology to display each projection to only one eye, objects may be displayed with *negative*, *zero*, or *positive parallax*, corresponding to their appearance in front, at, or behind the screen. Objects with zero parallax appear attached to the projection screen and are perfectly suited for touch interaction. In contrast, it is more difficult to apply direct-touch interaction techniques to objects that appear in front of or behind the screen [15, 26, 29]. In this paper we focus on the major challenge of touching objects that appear in front of the projection screen. Two methodologies can be used for touching such stereoscopic objects on a tabletop display:

1. If the touch-sensitive surface captures only direct contacts, the user has to penetrate the stereoscopically displayed object to touch the 2D surface behind it [38, 39].
2. Alternatively, if the system can capture finger movements in front of the screen, the user may virtually “touch” the object in *mid-air*, i.e., in 3D space [3].

Due to the discrepancy between perceptual and motor space and the missing passive haptic feedback, both approaches provide natural feedback only for objects

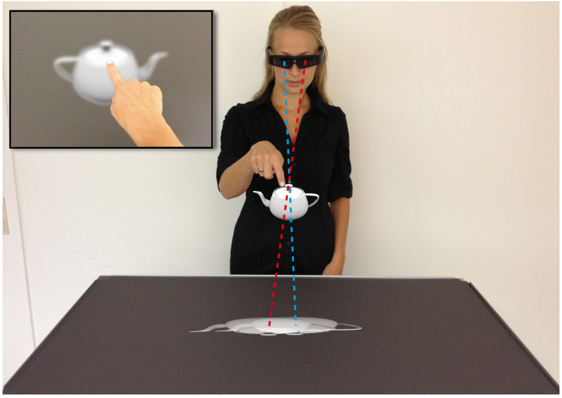


Fig. 2. Illustration of the main problem of 3D mid-air "touch" interaction with stereoscopically displayed 3D data: If a 3D tracking system is used, the user can see a stereoscopic image while converging to her finger. Due to the vegence-accommodation conflict, the virtual object appears blurred in comparison to the finger [7].

rendered with zero parallax. This poses the questions where users "touch" a stereoscopically displayed object in 3D space. Here, one issue is the well-documented issue of misperception of distances in virtual 3D scenes [20]. Another problem arises from potential touch locations on the 2D display surface, as there are two distinct projections, one for each eye. If the user penetrates the object while focusing on her finger, the stereoscopic impression of the object is disturbed, since the user's eyes are not accommodated and converged to the projection screen's surface anymore. Thus, the left and right stereoscopic images of the object's projection appear blurred and can usually not be merged as illustrated in Figure 1 (left). However, focusing on the virtual object causes a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged on the object's 3D position, see Figure 1 (right). If a 3D tracking system is used, the user can see a stereoscopic image while converging her eyes to her finger. Yet, due the vergence-accommodation conflict [7,8], the virtual object will appear blurred in comparison to the real finger (see Figure 2).

In this paper we address the challenge of how to interact with stereoscopic content in front of a touch-sensitive tabletop surface. Towards this, we also analyze touch behavior when touch sensing is constrained to the 2D screen surface. In order to allow the user to select arbitrary objects, a certain area of the touch surface, which we refer to as *on-surface target*, must be assigned to each object. In the monoscopic case the mapping between an on-surface touch area and the intended object point in the virtual scene is straightforward. Yet, with stereoscopic projection this mapping is more problematical. In particular, since there are different projections for each eye, the question arises *where* users touch the surface when they try to "touch" a stereoscopic object. In principle, the user may touch anywhere on the surface to select a stereoscopically displayed object. However, according to our previous work [39], the most likely alternatives that users try to touch are the (see Figure 4):

- *midpoint (M)* between the projections for both eyes,
- projection for *the dominant eye (D)*, or
- projection for the *non-dominant eye (N).*

A precise approach to this mapping is important to ensure efficient interaction and correct selections, in particular in a densely populated virtual scene. First, we determine a precise *on-display target area* where users touch the screen to select a 3D object. Second, we compare this approach with systems where the user's finger can be tracked in 3D space, and where users virtually touch objects in mid-air 3D space. The results of this experiment provide guidelines for the choice of touch technologies, as well as the optimal placement and parallax of interactive elements in stereoscopic touch environments.

In summary, our contributions are:

- An analysis of on-display touch areas for 3D target objects in stereoscopic touch-sensitive tabletop setups.
- A direct comparison of 2D touch and 3D mid-air selection precision.
- Guidelines for designing user interfaces for stereoscopic touch-sensitive tabletops.

The remainder of this paper is structured as follows. Section 2 summarizes related work in touch interaction and stereoscopic display. Section 3 describes the experiments we conducted to identify 2D/3D touch behavior. Section 4 presents the results, which are discussed in Section 5. An example application using the derived guidelines is described in Section 6. Section 7 concludes the paper.

2 Background

Recently, many approaches for extending multi-touch interaction techniques to 3D applications with monoscopic display have been proposed [15, 23, 28, 29]. For instance, Hancock et al. [15] presented the concept of shallow-depth 3D, i.e., 3D interaction within a limited range, to extend interaction possibilities with digital 2D surfaces. However, direct touch interaction with stereoscopically displayed scenes introduces new challenges [31], since the displayed objects can float in front of or behind the interactive display surface. Müller-Tomfelde et al. presented anaglyph- or passive polarization-based stereoscopic visualization combined with FTIR-based touch detection on a multi-touch enabled wall [25], and discussed approaches based on mobile devices for addressing the formulated parallax problems. The parallax problem described in the introduction is known from the two-dimensional representation of the mouse cursor within a stereoscopic image [31]. While the mouse cursor can be displayed stereoscopically on top of objects [31] or monoscopically only for the dominant eye [36], movements of real objects in the physical space, e.g., the user's hands, cannot be constrained such that they appear only on top of virtual objects. Grossman and Wigdor [11] provided an extensive review of the existing work on interactive surfaces and developed a taxonomy for this research. This framework takes the perceived and actual display space, the input space and the physical properties of an interactive surface into account. As shown in their work, 3D volumetric visualizations are rarely considered in combination with 2D direct surface input.

Even on monoscopic touch surfaces, the size of the human fingers and the lack of sensing precision can make precise touch screen interactions difficult [14, 40]. Some approaches have addressed this issue, for example, by providing an adjustable [6] or fixed cursor offset [27], by scaling the cursor motion [6] or by extracting the orientation of the user's finger [16].

2.1 Kinematics of Touch

The kinematics of point and grasp gestures and the underlying cognitive functions have been studied by many research groups [13, 21, 41]. For instance, it has been shown that total arm movement during grasping consists of two distinct component phases:

1. an initial, *ballistic phase* during which the user's attention is focused on the object to be grasped (or touched) and the motion is basically controlled by proprioceptive senses, and
2. a subsequent *correction phase* that reflects refinement and error-correction of the movement, incorporating visual feedback in order to minimize the error between the hand or finger, respectively, and the target [18].

Furthermore, MacKenzie et al. [19] have investigated the real-time kinematics of limb movements in a Fitts' task and have shown that, while Fitts' law holds for the total limb-movement time, humans usually start sooner decelerating the overall motion, if the target seems to require more precision in the end phase. The changes of the kinematics and control of the reaching tasks within virtual environments have also been investigated [7, 9, 29]. Valkov et al. [38] showed that users are, within some range, insensitive to small misalignments between visually perceived stereoscopic positions and the sensed haptic feedback when touching a virtual object. They proposed to manipulate the stereoscopically displayed scene in such a way that the objects are moved towards the screen when the user reaches for them [37, 38]. However, the problem is that objects have to be shifted in space, which may lead to a disturbed perception of the virtual scene for larger manipulations.

2.2 3D Touch for 3D Objects

To enable direct "touch" selection of stereoscopically displayed 3D objects in space, 3D tracking technologies can capture a user's hand or finger motions in front of the display surface. Hilliges et al. [16] investigated an extension of the interaction space beyond the touch surface. They tested two depth-sensing approaches to enrich multi-touch interaction on a tabletop setup. Although 3D mid-air touch provides an intuitive interaction technique, touching an intangible object, i.e., *touching the void* [8], leads to potential confusion and a significant number of overshoot errors. This is due to a combination of three factors: depth perception is less accurate in virtual scenes than in the real world, see e.g., [32], the introduced double vision, and also vergence-accommodation conflicts. A few devices, such as the CyberGrasp, support haptic feedback when touching objects in space, but require extensive user instrumentation. A similar option for direct touch interaction with stereoscopically rendered 3D objects is to separate the interactive surface from the projection screen, as proposed by Schmalstieg et al. [30]. In their approach, the user is provided with a physical *transparent prop*, which can be moved in front of the object of interest. This object can then be manipulated via single- or multi-touch gestures, since it has almost zero parallax with respect to the prop.

2.3 2D Touch for 3D Objects

Recently, multi-touch devices with non-planar surfaces, such as cubic [10] or spherical ones [5], were proposed. Other approaches are based on controlling the 3D position of a cursor through multiple touch points [4, 34]. These can specify 3D axes or points for indirect object manipulation. Interaction with objects with negative parallax on a multi-touch tabletop setup was addressed by Benko et al.'s balloon selection [4], as well as Strothoff et al.'s triangle cursor [34], which use 2D touch gestures to specify height above the surface. Valkov et al. [39] performed a user study, in which they displayed 3D objects stereoscopically either in front of or behind a large vertical projection screen. They recorded user behavior when instructed to touch the virtual 3D objects on the display surface. They identified that users tend to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. However, the results suffered from a large variance between subjects. Hence, it is unclear how far these results can be applied to different setups, such as mobile screens or tabletops, where users have an easy frame of reference due to the bezel. Also, they may engage in different touch behavior due to physical support and gravity.

So far, no comparative analysis exists for 2D and 3D touch interaction in stereoscopic tabletop setups. Thus, it remains unclear if 2D touch is a viable alternative to 3D mid-air touch.

3 Experiments

Here we describe our experiments in which we analyzed the touch behavior as well as the precision of 2D touch and 3D mid-air touches. We used a standard ISO 9241-9 selection task setup [19] on a tabletop surface with 3D targets displayed at different heights above the surface, i.e., with different negative stereoscopic parallaxes.

3.1 Participants

Ten male and five female subjects (ages 20-35, M=27.1, heights 158-193cm, M=178.3cm) participated in the experiment. Subjects were students or members of the Departments of computer science, media communication or human computer-interaction. Three subjects received class credit for participating in the experiment.

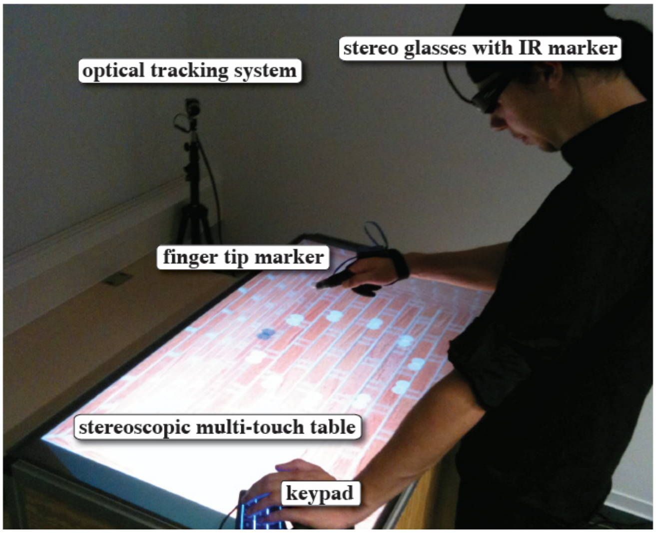


Fig. 3. Experiment setup: photo of a subject during the experiment (with illustrations). As illustrated on the screen, the target objects are arranged in a circle.

All subjects were right-handed. We used the Porta and Dolman tests to determine the sighting dominant eye of subjects [22]. This revealed eight right-eye dominant subjects (7 males, 1 female) and five left-eye dominant subjects (2 males, 3 females). The tests were inconclusive for two subjects (1 male, 1 female), for which the 2 tests indicated conflicting eye dominance. All subjects had normal or corrected to normal vision. One subject wore glasses and four subjects wore contact lenses during the experiment. None of the subjects reported known eye disorders, such as color weaknesses, amblyopia or known stereopsis disruptions. We measured the interpupillary distance (IPD) of each subject before the experiment, which revealed IPDs between 5.8cm and 7.0cm (M=6.4cm). We used each individual's IPD for stereoscopic display in the experiment. 14 subjects reported experience with stereoscopic 3D cinema, 14 reported experience with touch screens, and 8 had previously participated in a study involving touch surfaces. Subjects were naive to the experimental conditions. Subjects were allowed to take a break at any time between experiment trials in order to minimize effects of exhaustion or lack of concentration. The total time per subject including pre-questionnaires, instructions, training, experiment, breaks, post-questionnaires, and debriefing was about 1 hour.

3.2 Materials

For the experiment we used a 62 x 112cm multi-touch enabled active stereoscopic tabletop setup as described in [7]. The system is shown in Figure 3 and uses rear diffuse illumination [24] for multi-touch. For this, six high-power infrared (IR) LEDs illuminate the screen from behind. When an object, such as a finger or palm, comes in contact with the diffuse surface it reflects the IR light, which is then sensed by a camera. We use a 1024x768 PointGrey Dragonfly2 with a wide-angle lens and a matching IR band-pass filter at 30 frames per second. We use a modified version of the NUI Group's CCV software to detect touch input on a Mac Mini server. Our setup uses a matte diffusing screen with a gain of 1.6 for the stereoscopic back projection. We used a 1280x800 Optoma GT720 projector with a wide-angle lens and an active DLP-based shutter at 60Hz per eye. We used an optical WorldViz PPT X4 system with sub-millimeter precision and sub-centimeter accuracy to track the subject's finger and head in 3D, both for 3D "touch" detection as well as view-dependent rendering. For this, we attached wireless markers to the shutter glasses and another diffused IR LED on the tip of the index finger of the subject's dominant hand. We tracked and logged both head and fingertip movements during the experiment.

The visual stimulus consisted of a 30cm deep box that matches the horizontal dimensions of the tabletop setup (see Figure 3). We matched the look of the scene to the visual stimuli used by Teather and Stuerzlinger [35, 36]. The targets in the experiment were represented by spheres, which were arranged in a circle as illustrated in Figure 3. A circle consisted of 11 spheres rendered in white, with the active target sphere highlighted in blue. The targets highlighted in the order specified by ISO 9241-9 [18]. The center of each target sphere indicated the exact position where subjects were instructed to touch with their dominant hand in order to select a sphere. For 3D touch this was the 3D position, and for 2D touch the center of the 2D projection. The size, distance, and height of target spheres were constant within circles, but varied between circles. Target height was measured as positive height from the level screen surface. Subjects indicated target selection using a Razer Nostromo keypad with their non-dominant hand. The virtual scene was rendered on an Intel Core i7 3.40GHz computer with 8GB of main memory, and an Nvidia Quadro 4000 graphics card.

3.3 Methods

The experiment used a 2 x 5 x 2 x 2 within-subjects design with the method of constant stimuli, in which the target positions and sizes are not related from one circle to the next, but presented randomly and uniformly distributed [11]. The independent variables were selection technique (2D touch vs. 3D mid-air touch), target height (between 0cm and 20cm, in steps of 5cm), as well as target distance (16cm and 25cm) and target size (2cm and 3cm). Each circle represented a different index of difficulty (ID), with combinations of 2 distances and 2 sizes. The ID indicates overall task difficulty [12]. It implies that the smaller and farther a target, the more difficult it is to select quickly and accurately. Our design thus uses four uniformly distributed IDs ranging from approximately 2.85bps to 3.75bps, representing an ecologically valuable range of difficulties for such a touch-enabled stereoscopic tabletop setup. As dependent variables we measured the on- as well as off-display touch areas for 3D target objects.

The experiment trials were divided into two blocks: one for the 2D and one for the 3D touch technique. We randomized their order between subjects. At the beginning of each block subjects were positioned standing in an upright posture in front of the tabletop surface as illustrated in Figure 3. To improve comparability, we compensated for the different heights of the subjects by adjusting a floor mat below the subject's feet, resulting in an (approximately) uniform eye height of 1.85cm for each subject during the experiment. The experiment started with task descriptions, which were presented via slides on the tabletop surface to reduce potential experimenter bias. Subjects completed 5 to 15 training trials with both techniques to ensure that they correctly understood the task and to minimize training effects. Training trials were excluded from the analysis.

In the experiment, subjects were instructed to touch the center of the target spheres as accurately as possible (either with 2D or 3D touch), for which they had as much time as needed. For this, subjects had to position the tip of the index finger of their dominant hand inside the 3D sphere for the 3D touch condition, or push their finger through the 3D sphere until it reached the 2D touch surface. Subjects did not receive feedback whether they "hit" their target, i.e., subjects were free to place their index finger in the real world where they perceived the virtual target to be. We did this to evaluate the often-reported systematical over- or underestimation of distances in virtual scenes, which can be observed even for short grasping-range distances [32], as also tested in this experiment. Moreover, we wanted to evaluate the impact of such misperceptions on touch behavior in stereoscopic tabletop setups. We tracked the tip of the index finger in both 2D and 3D touch conditions. When subjects wanted to register the selection, they had to press a button with their non-dominant hand on the keypad. We recorded a distinct 2D and 3D touch position for each target location for each configuration of independent variables, with a total of 20 circles and 220 recorded touch positions per participant.

4 Results

In this section we summarize the results from the 2D and 3D touch experiment. We had to exclude two subjects from the analysis who obviously misunderstood the task. We analyzed these results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level (with Bonferonni correction).

4.1 2D Touch

For the 2D touch technique, we evaluated the judged 2D touch points on the surface relative to the potential projected target points, i.e., the midpoint (M) between the projections for both eyes, as well as the projection for the dominant (D), and the non-dominant (N) eye, as illustrated in Figure 4. Figure 5 shows scatter plots of the distribution of the touch points from all trials in relation to the projected target centers for the dominant and non-dominant eye for the different heights of 0cm, 5cm, 10cm, 15cm and 20cm (bottom to top). We normalized the touch points in such a way that the dominant eye projection D is always shown on the left, and the non-dominant eye projection N is always shown on the right side of the plot. The touch points are displayed relatively to the distance between both projections.

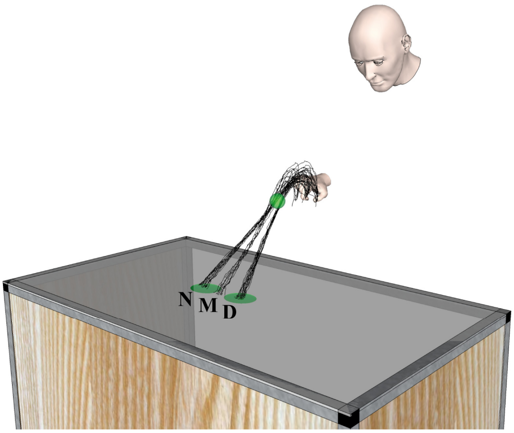


Fig. 4. Illustration of finger movement trails for eye user groups touching towards the dominant eye projection (D), non-dominant eye projection (N), or towards the midpoint. The trails have been normalized and are displayed here for a right-eye dominant user.

As it is illustrated in Figures 4 and 5, we observed three different behaviors when subjects used the 2D touch technique. In particular, eight subjects touched towards the midpoint, i.e., the center between the dominant and non-dominant eye projections. This includes the two subjects for whom eye dominance estimates were inconclusive. We arranged these subjects into the group G_M. Furthermore, three subjects touched towards the dominant eye projection D, which we refer to as group G_D, and three subjects touched towards the non-dominant eye projection N, which we refer to as group G_N. This points towards an approximately 50/50% split in terms of behaviors in the population, i.e., between group G_M and the composite of groups G_D and G_N.

We found a significant main effect of the three groups ($F(2,11)=71.267$, $p<.001$, partial-eta^2=.928) on the on-surface touch areas. Furthermore, we found a significant two-way interaction effect of the three groups and target heights ($F(8,44)=45.251$, $p<.001$, partial-eta^2=.892) on the on-surface touch areas. The post-hoc test revealed that the on-surface target areas, see Figure 5, significantly ($p<.001$) vary for objects that are displayed at heights of 15cm or higher. For objects displayed at 10cm height group G_D and G_N vary significantly ($p<.02$). For objects displayed below 10cm we could not find any significant difference. As illustrated in Figure 5, for these heights the projections for the dominant and non-dominant eye are close together, and subjects touched almost the same on-screen target areas.

Considering the on-surface touch areas, we found that on average the relative touch point for group G_D was $0.97D+0.03N$ for projection points $D\in\mathbb{R}^2$ and $N\in\mathbb{R}^2$, meaning the subjects in this group touched towards the projection for the dominant eye, but slightly inwards to the center. The relative touch point for group G_N was $0.11D+0.89N$, meaning the subjects in this group touched towards the projection for the non-dominant eye, again with a slight offset towards the center. Finally, for group

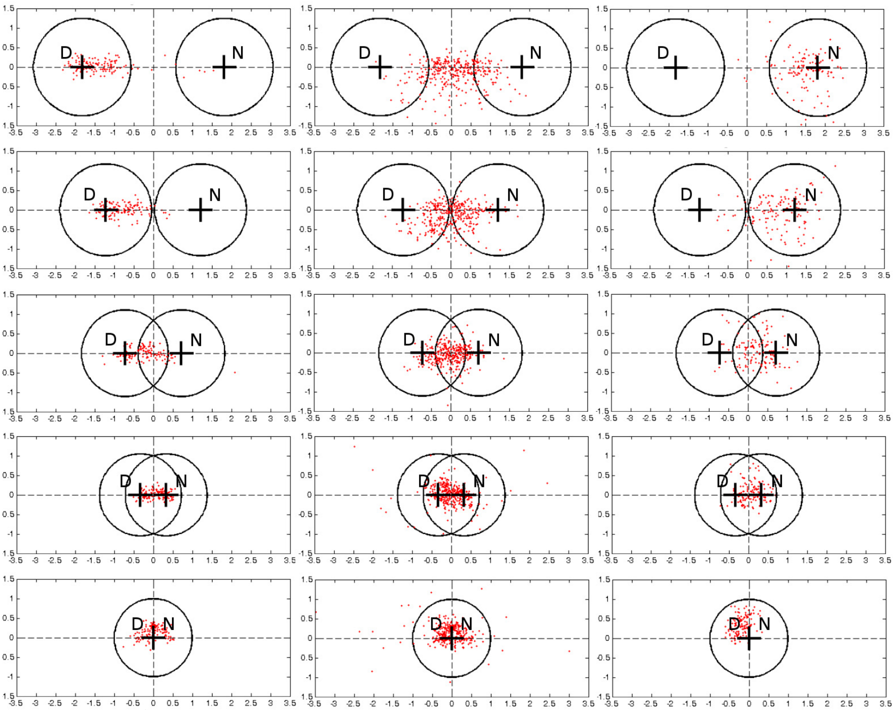


Fig. 5. Scatter plots of relative touch points between the dominant (D) and non-dominant (N) eye projections of the projected target centers on the surface for the 2D touch technique. Black crosses indicate the two projection centers. Black circles indicate the approximate projected target areas for the dominant and non-dominant eye. Top to bottom rows show results for 20cm, 15cm, 10cm, 5cm, and 0cm target heights. The left column shows subject behavior for dominant-eye touches (3 subjects), the middle for center-eye touches (8 subjects), and the right for non-dominant-eye touches (3 subjects). Note that the distance between the projection centers depends on the target height.

G_M we found that on average the relative touch point for this group was *0.504*D+*0.596*N. We could not find any significant difference for the different heights, i.e., the touch behaviors were consistent throughout the tested heights.

However, we observed a trend of target height on the standard deviations of the horizontal distributions (*x*-axis) of touch points for all groups as shown in Figure 5. For 0cm target height we found a mean standard deviation (SD) of 0.29cm, for 5cm SD 0.32cm, for 10cm SD 0.42cm, for 15cm SD 0.52cm, and for 20cm SD 0.61cm. For the vertical distribution (*y*-axis) of touch points and at 0cm target height we found a mean SD of 0.20cm, for 5cm SD 0.20cm, for 10cm SD 0.25cm, for 15cm SD 0.29cm, and for 20cm SD 0.30cm.

In summary, the results for the 2D touch technique show a significant effect for the different user groups on the on-surface touch area over the range of tested heights. These on-surface touch areas vary significantly for objects displayed at heights of 10cm and higher.

4.2 3D Touch

We analyzed the tracked physical 3D “touch” points where subjects judged the perceived center of the mid-air target spheres for the 3D touch technique in terms of their deviation from their actual position in the 3D virtual scene. Figure 6 shows scatter plots of the distribution of judged target positions in relation to the 3D target centers for the different target heights over all trials. The red dots indicate the center positions of the spheres as judged by the subjects. The black wireframe spheres illustrate the actual position and size of the objects. We normalized the judged positions relative to the optical view angle towards the target center. We found no significant difference in the judged positions for the three groups identified in Section 4.1 and pooled the data.
We analyzed the effect of target height on the subjects' judgments. We found a significant main effect of target height on the distances of judged positions from the displayed target centers. Mauchly's test indicated that the assumption of sphericity had been violated for effects of height on the distances of judged positions ($chi^2(9)=62.388$, $p<.001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (epsilon=.302). The results show that the distances of judged positions significantly differs for heights ($F(1.21,15.725)=12.846$, $p<.002$, partial-$eta^2=.497$).

A Tukey post-hoc test with Bonferroni correction revealed that subjects estimated the target centers significantly closer to the actual displayed target centers for the 0cm targets in comparison to targets displayed at 20cm height ($p<.002$). For all other heights the results suggest that the higher the targets are displayed, the larger are the deviations. Pooling over all subjects, we observed mean distances to target centers of M=0.56cm (SD=0.27cm) for 0cm target height, M=0.88cm (SD=0.53cm) for 5cm, M=0.97cm (SD=0.61cm) for 10cm, M=1.32cm (SD=0.93cm) for 15cm, and M=1.90cm (SD=1.48cm) for 20cm. The results suggest that the physical constraints provided by the touch surface at 0cm height reduced judgment errors for objects at zero parallax relative to the other heights. We found no significant difference when comparing to the results for the 2D touch technique at 0cm target height as presented in Section 4.1.

As it can be seen in Figure 6, subjects made larger errors along the view axis than along the orthogonal axes. For the mid-air target positions we found a mean standard deviation of 1.43cm along the optical line-of-sight, a mean SD of 0.36cm parallel to the touch surface, and a mean SD of 0.50cm orthogonal to the other axes. Furthermore, these deviations increased with increasing target heights. For the different target heights above the surface we observed standard deviations of judged positions along the optical line-of-sight of SD=2.20cm (for 20cm target height), SD=1.52cm (15cm), SD=1.05cm (10cm), and SD=0.94cm (5cm). On the other hand, we observed standard deviations of judged positions orthogonal to the view axis parallel to the touch surface of only SD=0.49cm (20cm), SD=0.39cm (15cm), SD=0.30cm (10cm), and SD=0.27cm (5cm). Finally, we found standard deviations of judged positions orthogonal to the other axes of only SD=0.70cm (20cm), SD=0.55cm (15cm), SD=0.41cm (10cm), and SD=0.35cm (5cm). We further analyzed the data to

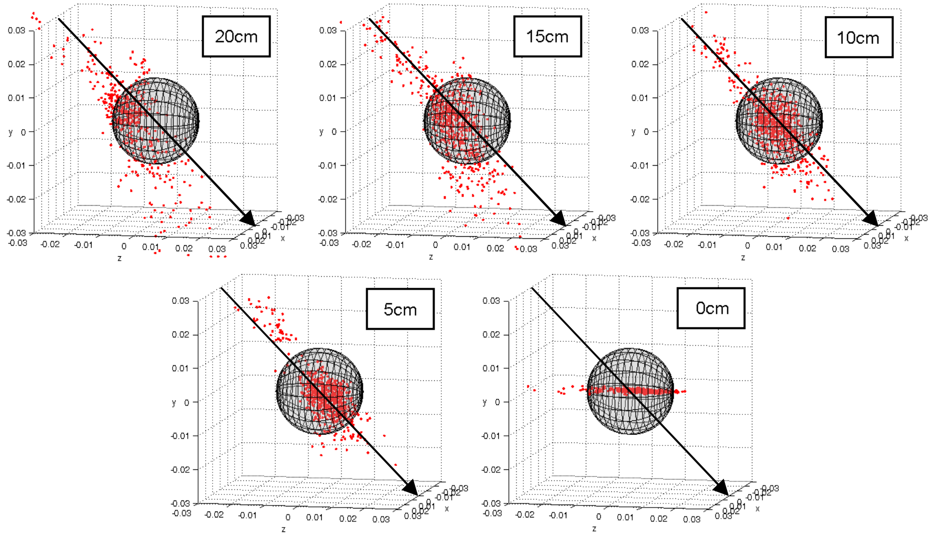


Fig. 6. Scatter plots of judged positions of the 3D target centers for the 3D touch technique over all subjects. Black wireframe spheres indicate the targets. The diagonal arrow illustrates the normalized view angle. The five diagrams show results for 20cm, 15cm, 10cm, 5cm, and 0cm target heights.

determine whether deviations in judged target positions result from under- or overestimation of distances from the observer to the mid-air targets [7,8]. We observed a mean distance underestimation of 0.25% (SD=2.93%). Surprisingly, we found a distance overestimation of M=0.4% (SD=2.00%) and M=1.0% (SD=2.25%) for heights of 5cm and 10cm, respectively. Yet, we found an underestimation of M=-0.54% (SD=2.67%) and M=-0.98% (SD=4.18%) for heights of 15cm and 20cm, respectively.

In summary, the results for the 3D touch technique show a significant effect of stereoscopic parallax over the range of tested heights on the precision and accuracy of judging the position of a target object.

5 Discussion

Our results provide interesting guidelines on how touch interaction in 3D stereoscopic tabletop setups should be realized. First of all and in contrast to previous work [39], our results show evidence for a twofold diversity of 2D touch behaviors of users. As shown in Figure 5, roughly half of the subjects in our study touched through the virtual object towards the center between the projections, and the other half touched towards projections determined by a single eye. The second group roughly splits in half again depending if they touch the projection for the dominant or non-dominant eye. Our results differ from the findings by Valkov et al. [39]. Using a setup with a large vertical projection plane they observed that subjects touched towards the center projection, with a slight offset towards the dominant eye. With 3 subjects touching towards the dominant eye, and 3 subjects towards the non-dominant eye in our study, user behavior in tabletop environments cannot be explained by this model. As a guideline, we suggest that the center between the projection for the left and right eye can be used to detect selections of objects stereoscopically displayed with less than 10cm height, since we did not observe significant differences between subjects at such heights. In order to reliably detect selections for objects higher above the screen, i.e., with larger parallaxes, our results suggest that for each user a calibration would be required. Our results confirm that this approach is highly beneficial, since subjects touched *consistently* for all heights towards the dominant, center, or non-dominant projection.

For practical considerations and to evaluate the ecological validity of using the 2D touch technique for selections of targets at a height between 0cm and 10cm, we computed the minimal on-surface touch area that supports 95% correct detection of all 2D touch points in our experiment. Due to the similar distributions of touch points between the three behavior groups for these heights shown in Figure 5, we determined the average minimal 95% on-surface region over all participants. Our results show that an elliptical area with horizontal and vertical diameter of 1.64cm and 1.07cm with a center in the middle between the two projections is sufficient for 95% correct detection. This rule-of-thumb heuristic for on-surface target areas is easy to implement and ecologically valuable considering the fat finger problem [14, 40]. Due to this problem objects require a relatively large size of between 1.05cm to 2.6cm for reliable acquisition, even in monoscopic touch-enabled tabletop environments.

The results of our second experimental condition reveal that distinct differences exist between the 3D mid-air touch technique and the 2D touch technique. These differences impact the relative performance and applicability for interaction with objects displayed stereoscopically at different heights above the surface. We found no behavior groups or effects of eye dominance on the distribution of judged 3D target positions. Our results show that target height has an effect on precision and accuracy of 3D selections, with large errors mainly along the optical line-of-sight, which we believe to correlate with distance misperception. For 3D objects displayed close to the display surface up to 10cm, touching objects in 2D on the surface by touching "through" the stereoscopic impression is more accurate than 3D mid-air touching. Considering that much research has shown that 3D mid-air touches of virtual objects suffer from low accuracy and precision due to visual conflicts, including vergence-accommodation mismatch, diplopia, and distance misperception [7, 8], it is a promising finding that the reduction of 3D selection tasks to 2D input with the 2D touch technique can improve performance for tabletop surface with stereoscopically displayed objects. However, the results also show that the accuracy for 2D touching of objects displayed above the screen decreases significantly for large negative parallax. The findings are encouraging for stereoscopic visualization on (multi-)touch surfaces. They suggest that virtual objects do not have to be constrained exactly at the zero-parallax level, but may deviate up to 10cm before 2D touch accuracy is significantly degraded [38, 39]. For such distances, the 2D touch technique is a good choice and instrumenting users with gloves or 3D markers can be avoided. Overall, our results show that it is possible to leverage stereoscopic cues in tabletop setups for an improved spatial cognition.

As a guideline for future tabletop setups with direct 2D touch input, the results suggest that touch-enabled 3D objects should not be displayed above an interactive display surface at more than about 10cm height. Above that, the disadvantages outperform the benefits and 3D interaction techniques should be used in that region, as they will provide more accurate interaction possibilities.

6 Example Application: Stereoscopic 3D Widgets

Our experiments have shown that the 2D touch technique has enormous potential as a new interaction paradigm for stereoscopic multi-touch surfaces as long as the objects are displayed with less than 10cm above the surface. In this region our 2D touch technique is the more accurate choice. While this constraint appears to limit the application scenarios in which one could use the 2D touch technique, it also ensures a simple implementation for interaction, in particular, a clear definition of on-display target areas as described in Section 5. Moreover, the size and scale of many virtual objects used in actual tabletop applications suit this constraint. For instance, 3D widgets can be displayed stereoscopically on any multi-touch surface and provide the user with a natural haptic feedback experience when she virtually touches them.

In order to evaluate the quality of the 3D touch technique in a real-world application, we adapted a simple visualization application for virtual caravans (see Figure 7). With this application customers can evaluate various types of caravans with several different features. The 3D widgets on the menu plane allow users to change the visual appearance of the caravan, lighting parameters, turn on signals, headlamps etc. We implemented the on-surface target areas of these 3D widgets as described in Section 5. The highest widgets, i.e., the 3D buttons on the menu panel, are displayed about 10cm above the surface. We used the same physical setup as described in Section 3.2. For this application we used the Unity3D game engine for the generation and rendering of the virtual scene. Unity3D provides a simple development environment for

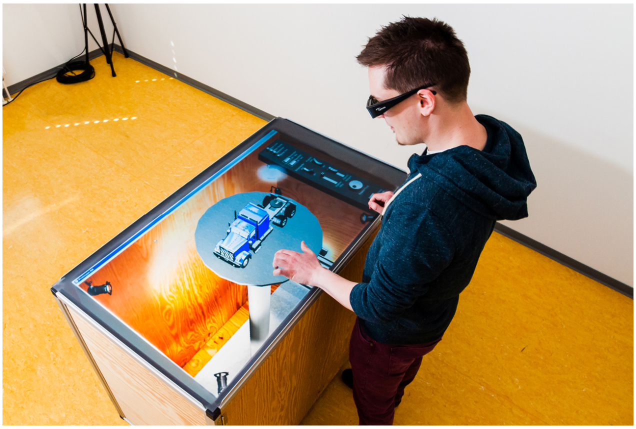


Fig. 7. User interacting with a virtual scene in a stereoscopic multi-touch tabletop setup using touch-enabled 3D widgets. The widgets in the graphical user interface were rendered with negative parallax of up to 10cm height.

virtual scenes, animations and interactions. In order to synchronize virtual camera objects with the movements of a user, we integrated the MiddleVR for Unity software framework. MiddleVR supports streaming of motion data from our tracking system to Unity3D using the Virtual Reality Peripheral Network (VRPN) protocol. With this we stream head poses to Unity3D, resulting in a correct perspective from the user's point of view at all times.

We presented this application to four users, and made several interesting observations. First, all users acknowledged the stereoscopic display when viewing the 3D scene. Second, most users immediately understood that the menu panel with the 3D widgets provides a means to interact with the setup. Surprisingly, when users tried to "touch" the 3D widgets, they adapted their actions to the affordances provides by the widget. For instance, when they pressed the toggle switch, usually they touched its lifted part, although we did not distinguish between touch positions on the surface. We see this as further indication that stereoscopic display in combination with a touch-enabled surface does indeed support the notion of 3D physical interaction elements. Finally, none of the users complained about non-reactive 3D widgets, which might have occurred if they missed the on-surface target areas. This suggests that the shape and size of the on-surface touch areas, as determined by our above study, is sufficient for using stereoscopic 3D widgets in tabletop setups.

7 Conclusion and Future Work

In this paper we evaluated and compared 2D touch and 3D touch interaction techniques for scenes on touch-sensitive tabletop setups with stereoscopic display. We analyzed the differences of 3D mid-air touch input and a technique based on reducing the 3D touch problem to two dimensions by having users touch "through" the stereoscopic impression of 3D objects, resulting in a 2D touch on the display surface. We identified two separate classes of user behavior, with one group that touches the center between the projections, whereas the other touches the projection for the dominant or non-dominant eye. The results of the experiment show a strong interaction effect between input technique and the stereoscopic parallax of virtual objects.

The main contributions of this work are:

- We identified two separate classes of user behavior when touching "through" stereoscopically displayed objects.
- We compared precision and accuracy of 2D/3D direct touch input, which revealed that the 2D touch technique is a viable alternative to 3D touch interaction for object selection up to about 10cm height from the display surface.
- We determined on-surface target regions that support a simple implementation of the 2D touch technique. This enables intuitive touch input for 3D objects and widgets in stereoscopic 3D tabletop applications.

The results are encouraging for stereoscopic visualization in future touch-enabled tabletop setups, since no additional instrumentation and tracking technology is needed for objects with a small stereoscopic parallax. An interesting question for future work is if the results can be applied to portable setups, where the orientation of the touch-sensitive surface varies during interaction. We plan to further pursue these topics to provide compelling user experiences and effective user interfaces for touch-sensitive stereoscopic display surfaces. Moreover, we plan to investigate also how the 2D and 3D touch methods compare in terms of the speed-accuracy tradeoff.

Acknowledgements. This work was partly supported by grants from the Deutsche Forschungsgemeinschaft and the Natural Sciences and Engineering Research Council of Canada.

References

1. iMUTS - Interscopic Multi-Touch Surfaces (2013), `http://imuts.uni-muenster.de`
2. InSTInCT - Touch-based interfaces for Interaction with 3D Content (2013), `http://anr-instinct.cap-sciences.net`
3. Leap Motion (2013), `http://www.leapmotion.com`
4. Benko, H., Feiner, S.: Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection. In: Proc. IEEE 3DUI, pp. 79–86 (2007)
5. Benko, H., Wilson, A.D., Balakrishnan, R.: Sphere: multi-touch interactions on a spherical display. In: Proc. ACM UIST, pp. 77–86 (2008)
6. Benko, H., Wilson, A.D., Baudisch, P.: Precise selection techniques for multi-touch screens. In: Proc. ACM CHI, pp. 1263–1272 (2006)
7. Bruder, G., Steinicke, F., Stuerzlinger, W.: Effects of Visual Conflicts on 3D Selection Task Performance in Stereoscopic Display Environments. In: Proceedings of IEEE Symposium on 3D User Interfaces (3DUI), pp. 115–118. IEEE Press (2013)
8. Chan, L.-W., Kao, H.-S., Chen, M.Y., Lee, M.-S., Hsu, J., Hung, Y.-P.: Touching the void: Direct-touch interaction for intangible displays. In: Proc. ACM CHI, pp. 2625–2634 (2010)
9. de la Riviere, J.-B., Dittlo, N., Orvain, E., Kervegant, C., Courtois, M., Da Luz, T.: ilight 3D touch: a multiview multitouch surface for 3D content visualization and viewpoint sharing. In: Proc. ACM ITS, pp. 312–312 (2010)
10. de la Riviere, J.-B., Kervegant, C., Orvain, E., Dittlo, N.: Cubtile: a multi-touch cubic interface. In: Proc. ACM VRST, pp. 69–72 (2008)
11. Ferwerda, J.: Psychophysics 101: How to run perception experiments in computer graphics. In: SIGGRAPH Core (2008)
12. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psych. 47, 381–391 (1954)
13. Geniva, L., Chua, R., Enns, J.T.: Attention for perception and action: task interference for action planning, but not for online control. Exp. Brain Res. 185(4), 709–717 (2008)
14. Hall, A.D., Cunningham, J.B., Roache, R.P., Cox, J.W.: Factors affecting performance using touchentry systems: Tactual recognition fields and system accuracy. J. Appl. Psychol. 4, 711–720 (1988)
15. Hancock, M., Carpendale, S., Cockburn, A.: Shallow-depth 3D interaction: design and evaluation of one-, two- and three-touch techniques. In: Proc. ACM CHI, pp. 1147–1156 (2007)

16. Hilliges, O., Izadi, S., Wilson, A.D., Hodges, S., Garcia-Mendoza, A., Butz, A.: Interactions in the air: Adding further depth to interactive tabletops. In: Proc. ACM UIST, pp. 139–148 (2009)
17. Holz, C., Baudisch, P.: The generalized perceived input point model and how to double touch: Accuracy by extracting fingerprints. In: Proc. ACM CHI, pp. 1165–1174 (2009)
18. International Organization for Standardization. ISO/DIS 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices (2000)
19. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. 3(3), 203–220 (1993)
20. Loomis, J.M., Knapp, J.M.: Visual perception of egocentric distance in real and virtual environments. In: Hettinger, L.J., Haas, M.W. (eds.) Virtual and Adaptive Environments, volume Virtual and Adaptive Environments, Mahwah (2003)
21. MacKenzie, C.L., Marteniuka, R.G., Dugasa, C., Liskea, D., Eickmeiera, B.: Three-dimensional movement trajectories in Fitts' task: Implications for control. Q.J. Exp. Psychology-A 39(4), 629–647 (1987)
22. Mapp, A.P., Ono, H., Barbeito, R.: What does the dominant eye dominate? A brief and somewhat contentious review. Percept. Psychophys 65(2), 310–317 (2003)
23. Martinet, A., Casiez, G., Grisoni, G.: The design and evaluation of 3D positioning techniques for multi-touch displays. In: Proc. IEEE 3DUI (2010)
24. McIntire, J.P., Havig, P.R., Geiselman, E.E.: What is 3D good for? A review of human performance on stereoscopic 3D displays. In: Proc. of the SPIE, Head- and Helmet-Mounted Displays XVII, vol. 8383, pp. 1–13 (2012)
25. Müller-Tomfelde, C., Schöning, J., Hook, J., Bartindale, T., Schmidt, D., Oliver, P., Echtler, F., Motamedi, N., Brandl, P., Zadow, U.: Building interactive multi-touch surfaces. In: Müller-Tomfelde, C. (ed.) Tabletops - Horizontal Interactive Displays. Human-Computer Interaction Series, pp. 27–49. Springer (2010)
26. Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., Mine, M.: Image plane interaction techniques in 3D immersive environments. In: Proc. ACM I3D, pp. 39–44 (1997)
27. Potter, R.L., Weldon, L.J., Shneiderman, B.: Improving the accuracy of touch screens: An experimental evaluation of three strategies. In: Proc. ACM CHI, pp. 27–32 (1988)
28. Pyryeskin, D., Hancock, M., Hoey, J.: Investigating selection above a multi-touch surface. In: Proc. ACM ITS (2012)
29. Reisman, J.L., Davidson, P.L., Han, J.Y.: A screen-space formulation for 2D and 3D direct manipulation. In: Proc. ACM UIST, pp. 69–78 (2009)
30. Schmalstieg, D., Encarnacao, L.M., Szalavari, Z.: Using transparent props for interaction with the virtual table. In: Proc. ACM I3D, pp. 147–153 (1999)
31. Schöning, J., Steinicke, F., Krüger, A., Hinrichs, K., Valkov, D.: Bimanual interaction with interscopic multi-touch surfaces. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 40–53. Springer, Heidelberg (2009)
32. Sivak, B., Mac Kenzie, C.L.: Integration of visual information and motor output in reaching and grasping: the contributions of peripheral and central vision. Neuropsychologia 28(10), 1095–1116 (1990)
33. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Phil. Trans. R. Soc. B 364, 3549–3557 (2009)
34. Strothoff, S., Valkov, D., Hinrichs, K.H.: Triangle cursor: Interactions with objects above the tabletop. In: Proc. ACM ITS, pp. 111–119 (2011)

35. Teather, R.J., Stürzlinger, W.: Pointing at 3D targets in a stereo head-tracked virtual environment. In: Proc. IEEE 3DUI, pp. 87–94 (2011)
36. Teather, R., Stuerzlinger, W.: Pointing at 3D Target Projections with One-Eyed and Stereo Cursors. In: ACM CHI (2013)
37. Valkov, D., Giesler, A., Hinrichs, K.: Evaluation of depth perception for touch interaction with stereoscopic rendered objects. In: Proc. ACM ITS (2012)
38. Valkov, D., Steinicke, F., Bruder, G., Hinrichs, K., Schöning, J., Daiber, F., Krüger, A.: Touching floating objects in projection-based virtual reality environments. In: Proc. JVRC, pp. 17–24 (2010)
39. Valkov, D., Steinicke, F., Bruder, G., Hinrichs, K.H.: 2D touching of 3D stereoscopic objects. In: Proc. ACM CHI, pp. 1353–1362 (2011)
40. Vogel, D., Baudisch, P.: Shift: A technique for operating pen-based interfaces using touch. In: Proc. ACM CHI, pp. 657–666 (2007)
41. Whitney, D., Westwood, D.A., Goodale, M.A.: The influence of visualmotion on fast reaching movements to a stationary object. Letters to Nature 423, 869–873 (2003)

Understanding Hand Degrees of Freedom and Natural Gestures for 3D Interaction on Tabletop

Rémi Brouet[1,2], Renaud Blanch[1], and Marie-Paule Cani[2]

[1] Grenoble Université LIG, [2] Grenoble Université LJK/INRIA
{remi.brouet,marie-paule.cani}@inria.fr,
renaud.blanch@imag.fr

Abstract. Interactively creating and editing 3D content requires the manipulation of many degrees of freedom (DoF). For instance, docking a virtual object involves 6 DoF (position and orientation). Multi-touch surfaces are good candidates as input devices for those interactions: they provide a direct manipulation where each finger contact on the table controls 2 DoF. This leads to a theoretical upper bound of 10 DoF for a single-handed interaction. With a new hand parameterization, we investigate the number of DoF that one hand can effectively control on a multi-touch surface. A first experiment shows that the dominant hand is able to perform movements that can be parameterized by 4 to 6 DoF, and no more (i.e., at most 3 fingers can be controlled independently). Through another experiment, we analyze how gestures and tasks are associated, which enable us to discover some principles for designing 3D interactions on tabletop.

Keywords: 3D manipulation, multi-touch interaction, tabletop interaction, gesture-based interaction.

1 Introduction

The interactions used to create or edit 3D content need to control simultaneously a large number of degrees of freedom (DoF). For instance, the classical docking task (i.e., defining the position and orientation of an object) requires the control of 6 DoF. The recent rise of tabletop devices seems promising for enabling such 3D interactions. Indeed, those devices have a number of desirable properties: first, despite the mismatch between the 2D nature of the input and the 3D nature of the virtual objects to be manipulated, tabletop interaction is closer to traditional shape design tools (such as pencil and paper, or modeling clay on a support table) than many 3D input devices, requiring to be hold in mid-air. Resting on a horizontal table induces less fatigue, allowing longer periods of activity. It also enables more precise gestures. Lastly, with the advent of multi-touch devices, the number of DoF that can be simultaneously controlled on a tabletop device is high: since each fingertip specifies a 2D position, the use of a single hand theoretically allows the control of 5 fingers × 2D = 10 DoF.

This value of 10 DoF is clearly an upper bound of the actual number of DoF that a user can simultaneously manipulate with a single hand. Several evidences show that the actual number is lower: so far, no multi-touch interaction uses the positions of the

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 297–314, 2013.

five fingers of a hand to control 10 parameters of the object being manipulated. Our common sense tells us that our fingers are not totally independent, since they are linked by the hand, and moreover that even for movements that would be physically doable, we can hardly control each finger independently.

To analyze gestures and DoF, using a new hand parameterization, we successfully decomposed gestures into elementary motion phases, such as translation, rotation and scaling phases. Such phase analysis method permits us to investigate fundamental behaviors of hands and gestures.

A first goal of this paper is to evaluate the upper bound of the number of DoF that can be simultaneously controlled by a hand on a multi-touch device. This is done through an experiment that confirms and refines what our common sense, as well as what the corpus of current multi-touch interaction techniques tell us: the number of DoF of the hand on a surface is between 4 and 6.

A second goal of the paper is to study how those DoF can be mapped to actual 3D manipulations, i.e., which interactions are the most efficient to exploit those DoF. Despite interaction with 3D content on tabletops is not “natural”, in the sense that there is no consensus among participants on how nontrivial 3D manipulations should be performed through 2D gestures, through another experiment, we discover some principles for designing 3D interactions on tabletop, which enable us to disambiguate 3D content manipulations. Possible manipulations correspond to navigation tasks (when the point of view is manipulated), object positioning tasks (i.e., object translation, rotation or scaling) and object deformation tasks (i.e., stretching, compressing or bending some part of an object).

Finally, to compare and validate our research, we investigate how the new phase analysis method fits with the other recent results on multi-touch devices.

2 Related Work

The first manipulation tool humans ever use is their hand, which enables them to touch, grab, pinch, move, or rotate many objects. Thanks to multi-touch devices, these abilities are nearly extended to the digital world.

Before creating a 3D user interface for multi-touch device, understanding the hand gesture is mandatory. Two aspects need to be studied: the hand gestures themselves, and the mapping between these gestures and tasks.

2.1 Hand/Finger Dependencies

Hand gesture analysis is a broad topic connected to many research fields. Every area we have explored notes dependences between fingers while performing a movement or a task. From a mechanical point of view, the hand consists of twenty-seven DoF, although biologically speaking, fingers are linked together by tendons and nerves and so on [1]. Neuroscientists note that a majority of hand movements can be described using two principal components [2]. Martin et al. observe dependence between fingers during voluntary and involuntary finger force change [3].

2.2 Multi-touch Interactions

The manipulation of 3D contents on tabletop is a recent research topic. Hancock et al. compared different techniques to manipulate 3D objects with one, two or three fingers [4]. They extended the RNT (for Rotation 'N' Translation) algorithm [5], and showed that, using spatial modes, one touch input is sufficient to control 5 DoF, while three touch inputs enables the decoupling of interactions and thus becomes more user-friendly. Martinet et al. described techniques to translate 3D objects along the depth axis using a finger of the non-dominant hand together with a unmoving dominant hand [6].

Those works are just two examples of the many 3D user interfaces using tabletop (e.g., [7–9]). A common characteristic of those interactions lies in the limited number of fingers used to manipulate the objects. Indeed, three fingers by hand are used to interact with the virtual environment for the most complex tasks, and the use of the five fingers only occur if the gesture performed is simple (a global translation and/or rotation involving the whole hand).

This rule even holds for interaction techniques designed for tasks more abstract than the manipulation of 3D contents like contextual menus that visualize information [10], or that enable the selection of tools or the switching of modes for manipulating objects [11, 12]. Again, all these interactions, while designed specifically for multi-touch devices, use at most three fingers by hand. Bailly et al.'s works about finger-count menus is a rare exception to this general pattern [13, 14]. Indeed, the number of finger corresponds to the number of the selected field in the menu.

2.3 Hand Gestures Analysis

In the context of multi-touch devices, hand gestures have been analyzed in conjunction with their mapping to particular tasks. Wobbrock et al. studied the "naturalness" of such mapping by letting users define gestures for a given set of tasks [15, 16]. Cohé et al. focused their analysis on object positioning tasks, and demonstrate the importance of finger starting points and of hand forms and trajectories [17].

In contrast, gestures are analyzed by phase analysis techniques. Nacenta et al. studied gestures during object positioning tasks, and discover that an order of manipulation exists [18]. One goal of this paper is to discover principles in order to develop 3D interactions based on phase analysis techniques.

3 Understanding Hand DoF on a Surface

To get a better understanding of possible hand gestures when the fingertips are constrained to remain on a table, we ran a first experiment that does not involve any 3D task. Since our goal was to estimate the number of DoF a user is able to simultaneously control with a single hand, we asked participants to use their dominant hand to perform a number of specific gestures.

3.1 Tasks

The gesture is specified by a starting position and an ending position. Those positions consist of five circles, each circle (resp. labeled with 1, 2, etc.) representing the position of a finger (resp. the thumb, the forefinger, etc.), as depicted on Fig. 1.a. Once a finger is correctly positioned, the corresponding circle turns green. Once all fingers are correctly positioned, the circles vanish, and the ending position appears. Then, the participant has to move his/her fingers to match the ending position, while keeping all fingers in contact with the surface. She/he can take as much time as she wants to perform each gesture.

The experiment was composed of thirty-seven trials. Those thirty-seven gestures are designed to be of various complexities: the simpler ones only involve movements of the whole hand, while the more complex ones involve the combinations of both hand movements and individual uncorrelated finger movements. Our set of gestures set was designed by testing in a preliminary study a comprehensive combination of elementary movements, and by discarding those that were too difficult to perform.

For the first ten trials, an animation between the starting and the ending position was shown to the user prior the trial, whereas for the other trials, no path was suggested. The participants were not asked to follow the suggestion, and its presence had no noticeable effect on the results we report below.

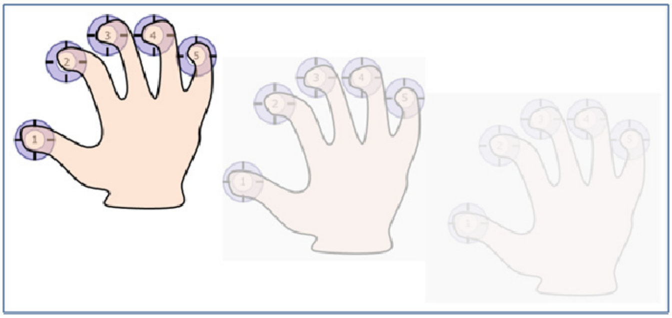


Fig. 1. To analyze hand gestures, we asked users to move their fingers from specific initial positions to specific final positions

3.2 Apparatus and Participants

This experiment was conducted on a 22’’ multi-touch display by 3M (473 × 296 mm, 1680 × 1050 pixels, 60 Hz, 90 DPI). The software was based on the QT and Ogre3D libraries.

31 participants, composed of 8 women and 23 men, were tested. Average age was 30 (min. 22, max 49). All participants had normal or corrected to normal vision. For left-handed participants, the experiment was mirrored. Participant’s background was variable, and not only computer scientist background. Participants’ experience with 3D applications, and tactile devices was variable, but this was not an issue, as the goal of the experiment was to get some understanding of fundamental physical behavior.

3.3 New Parameterization for Hand Analysis

During each trial, the trajectories of the tip of the fingers were recorded. To analyze gestures, we define the following parameterization of the hand: we use the position of the thumb as the origin of a local frame, in order to simplify the decomposition into phases. The first axis of the frame is given by the thumb/forefinger direction of the starting position. Therefore, the hand position is given by the local frame (2 DoF for the position of the origin), and by the position of each finger in this frame (0 DoF for the thumb as it is always at the origin, 2 DoF –distance and angle– for the other fingers).

The position of each finger in the local frame is parameterized by a couple (R_i, S_i) –for rotation and scale– where R_i is the angle defined by the finger of the local frame (i.e., the angle between the thumb/forefinger direction at the starting position and the thumb/finger direction at the current position); and S_i is the ratio between the current distance to the thumb of the finger and its distance to the thumb at the starting position (Fig. 2). With these definitions, a simple translation of the hand keeps the couple (R_i, S_i) constant (only the origin of the local frame changes); a rotation of the hand changes all the R_i by the same amount but does not impact the S_i. In contrast, a pinch gesture will only impact the S_i, making them decrease from 1 (fingers at the same distance from the thumb than while resting in the starting position) to a value smaller than 1 (fingers closer to the thumb).

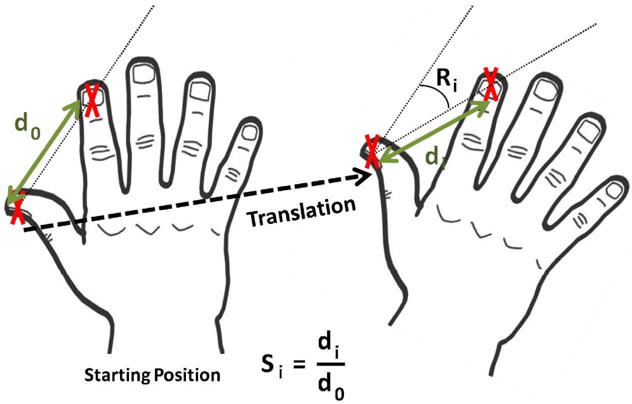


Fig. 2. Hand parameterization: definition of R_i and S_i

3.4 Results

A first look at the traces produced by participants' fingers confirms an intuitive hypothesis: hand gestures on a table can be decomposed into global motion phases (Fig. 3) and some local motion phases.

3.4.1 Global Gestures

The global part consists of the position of the hand (hand translation), its orientation (hand rotation), and how much it is opened (hand scaling). We quantify those parts using the hand translation (T) as the position of the origin of the local frame (i.e., of the thumb); and the hand rotation (R) (resp. scaling (S)) as a weighted barycenter of the R_i (resp. S_i). The weights are chosen to reduce the impact of a finger that is far from the others (i.e., to provide a kind of continuous median value), e.g., for R:

$$R = \frac{\sum_i w_i R_i}{\sum_i w_i} \quad \text{with } w_i = 1 + \sum_{j,k \neq i} (R_j - R_k) \tag{1}$$

We then define phases as periods of time during which a significant variation occurs for those variables, i.e., their first derivative is above a threshold (i.e., threshold are respectively 0.005, 0.5, and 0.001 for translation, rotation and scaling). Fig. 3 shows the variation of T, R and S while performing a gesture (top), and the corresponding phases (bottom). The pattern formed by this example is typical of what can be observed: there is a single phase for the translation, while the rotation and scaling are achieved during several phases (typically less phases are needed for R than for S). The different phases start roughly at the same time but end in this order: first T, then R, and then S. This pattern is similar to the one observed by Nacenta et al. [18], since what they call “period of maximum activity” are the second phase for R and the second or third phase for S.

To further validate this order of manipulations, we can look at the number of phases needed to validate the trials. Fig. 4 summarizes those results: for more than 93% of the cases, users need a single translation phase to correctly position their hand; while a correct rotation is achieved within a single phase for 68% of the trials and a correct scale for 35% of the trials.

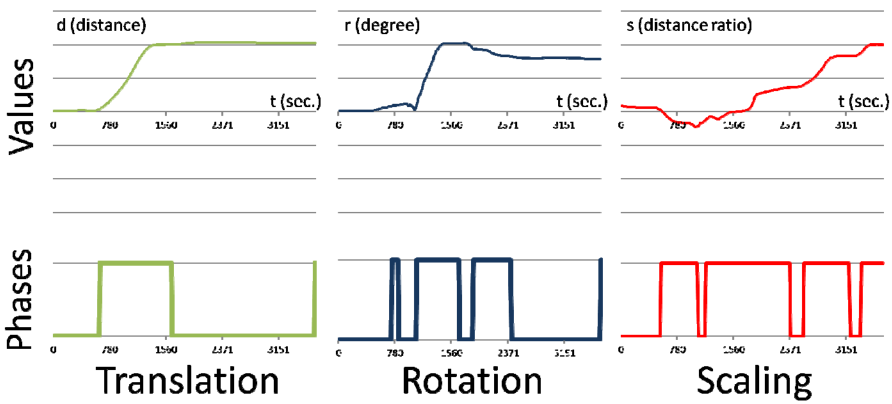


Fig. 3. Global variations (top), and corresponding phases (bottom), during a gesture: variations of translation, rotation and scaling

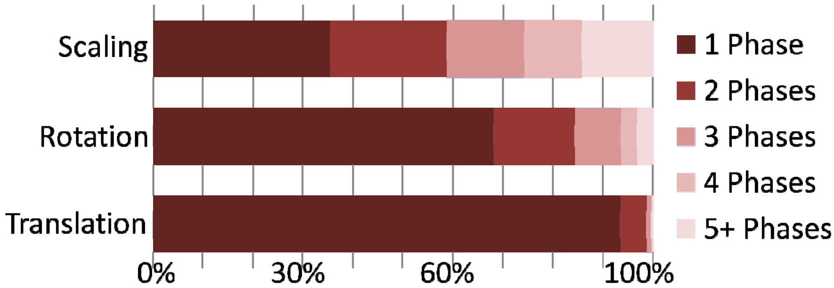


Fig. 4. Percentage of tasks where 1, 2, 3, 4, 5 or more phases are required among all tasks and participants for translation (T), rotation (R) and scaling (S)

Thus we think that hand gestures can be decomposed into sub-parts that have different degrees of stability: from the most stable motion (global translation) to the less stable motion (one finger motion). For instance (Fig. 5), the global translation is the easiest to get right (1 phase only), without any interference afterwards. On the contrary, global translation could induce interferences on rotation (first rotation phase), before that the major rotation motion is performed (second rotation phase). As translation and rotation are simultaneously performed, sometimes rotation motion has to be corrected (third phase).

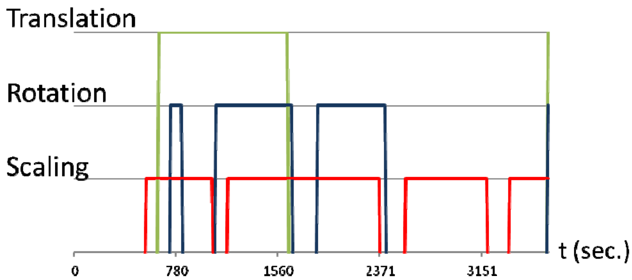


Fig. 5. Phases superposition for the Fig. 3 example

3.4.2 Local Gestures

The local parts of gesture are the components of individual finger movements that are not explained by the global T, R and S described above. A first look at the data shows that those local parts are mainly movements performed by the middle, ring and little fingers. To get a better understanding of those movements, we concentrate our analysis on the trials in which users had to perform movements involving only a subset of those fingers, and in which those movements was the same for the fingers involved.

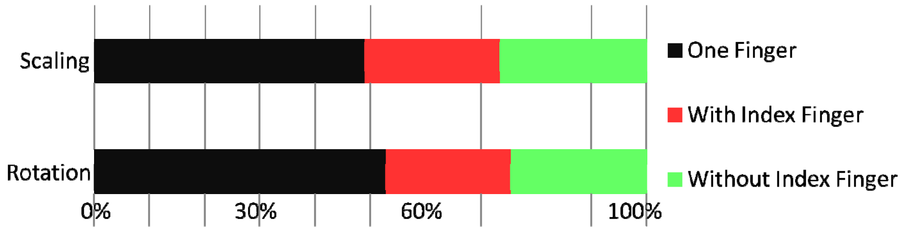


Fig. 6. Average percentage of time spent for moving a single finger, more than other fingers including/not including the index finger for tasks involving the motion of one or more fingers among the last three fingers only

For those tasks, Fig. 6 shows the proportion of time spends moving a single finger ($\approx 50\%$ of the time), the time spends moving more than one finger, including and excluding the index finger ($\approx 25\%$ each). We can note that those proportions are roughly the same whether the participants are asked to perform a scaling task, i.e., to control the S_i (top) or a rotation task, i.e., to control the R_i (bottom). It is also interesting that the movement of one (or more) of the last three fingers involves the motion of the index finger despite that in those tasks the index finger was not supposed to move. This shows how it is difficult for users to control the three last fingers simultaneously and independently. The interdependence between those fingers is consistent with the study conducted by Martin et al. [3].

To further investigate the interdependencies among the last three fingers, we split the trials into three groups, depending on the number of fingers the users have to move among the middle, ring and little fingers. Fig. 7 shows for each group (vertically: 1F, 2F, 3F), the relative time spent moving 1, 2 or 3 of those fingers. It is interesting to note that even when asked to move a single finger (1F), the participants spend more than 30% of their time moving two or more fingers. On the other hand, when participants have to perform the same motion for the last three fingers (3F), only one third of the time is used to move the fingers together, while $\approx 40\%$ of the time the fingers are moved individually.

This confirms that the three last fingers cannot be used to control something independently of the index finger, even if they are used together as a whole. Such dependencies induce difficulties for users to efficiently control the hand 10 DoF, and decrease this upper bound around 4 or 6 DoF (two or three independent fingers).

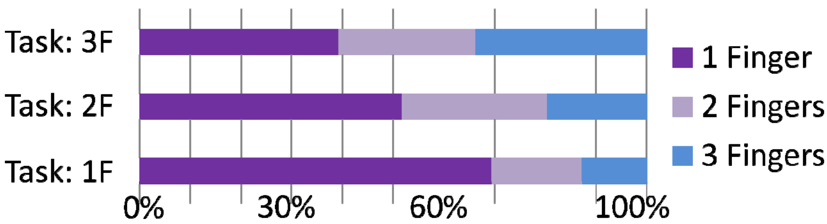


Fig. 7. Average percentage of time spent for moving 1, 2 or 3 fingers among the last three fingers, when the user is asked to move 1, 2 or 3 of them (1F, 2F, 3F)

4 Mapping Gestures and 3D manipulation

We ran a second experiment to understand the most natural mapping between 2D gestures and 3D tasks. Recent researches have focused either on navigation tasks (e.g., [4, 6]) or object positioning tasks (e.g., [7, 15]). Mixing both kinds of task increases the number of possible mapping. Therefore, one of our goals was to discover if the implicit information included in an interaction could be used to automatically switch between interaction modes, rather than having to provide explicit widgets for mode selection

4.1 Tasks

The participant observes an animation of the desired task on the first part of the screen (Fig. 8a, b) (left)), and then he/she performs a gesture of their choice to perform this task (right). The experiment was composed of thirty-six trials, divided into three classes: eleven navigation tasks, nine object positioning tasks and sixteen object deformations tasks. For navigation or object positioning tasks, the scene was composed of two cubes, a grid, and a background picture (Fig. 8a). For object deformations, only the grid and 3D object were shown (Fig. 8b).

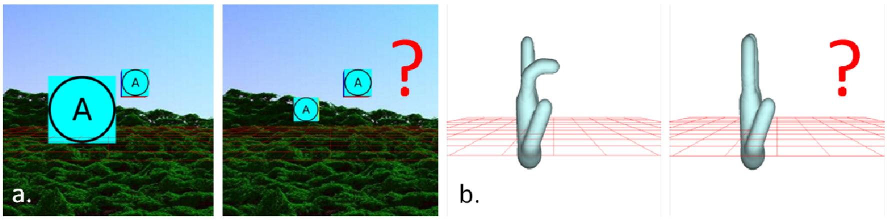


Fig. 8. a) Example of setting for discovering fundamental behavior for navigation / object positioning tasks. An animation is shown on one screen (left), while users perform gesture on second screen (right). b) Similar setting for object deformation tasks.

4.2 Hand Phase Analysis

The analysis process performed for the first experiment was reproduced with little differences. However, we had to adapt the hand parameterization to the number of fingers in contact with the table. Contrary to the first experiment, where each finger could be identified by the starting position, all interactions did not always involve the five fingers (e.g., the thumb was not always used).

The first experiment demonstrated that the thumb is usually the most stable finger (this was our reason for using it as origin of the local frame). Therefore, we assumed the thumb to be the one that was moving the less (this assumption can be wrong when the gesture is a translation, but this is non-issue, since all the fingers are being moved the same way in this case). The other fingers do not need to be distinguished. We also had to perform two distinct phase analyses, one for each hand, to interpret the gestures.

4.3 Results

4.3.1 Hands/Fingers Uses

To deeper investigate the efficient DoF a hand can control, we first observe that only three participants used more than 3 fingers by hand. Those cases mostly involved navigation tasks. In more details, when participants involved more than 3 fingers to manipulate the 3D contents, the principal phase of their interaction corresponds to translation phase (i.e., the most global motion). On average, fewer fingers by hand are used to handle objects than to navigate (Table 1). The difference between numbers can be explained by the use of the second hand. Further explanations are developed in the next section.

Table 1. Second Experiment results

	Generality		2nd Hand Data			Phase Analysis			Best
Tasks	Dist. to Obj.	Avg. #Fingers	% Tasks	2nd Hand #Fingers	Type	#T. Phase	#R. Phase	#S. Phase	Gesture 1st/2nd
Navigation									
Translation /xy	**1.5**	**2.5**	**03**	**1.0**	**Sym.**	**1.2**	**0.1**	**2.0**	**Tr./-**
Translation /z	**1.9**	**2.9**	**46**	**1.7**	**Sym.**	**1.5**	**0.1**	**1.0**	**Tr./Sym.**
Rotation /xy	**2.3**	**2.8**	**52**	**1.4**	**Sup.**	**1.1**	**0.3**	**1.5**	**Tr./Sup.**
Zoom	**1.6**	**3.5**	**54**	**1.8**	**Sym.**	**1.5**	**0.1**	**1.2**	**Tr./Sym.**
Zoom to Object	**1.2**	**3.0**	**40**	**1.6**	**Sym.**	**1.5**	**0.2**	**1.1**	**Tr./Sym.**
Object Positioning									
Translation /xy	**0.2**	**1.2**	**00**	**-**	**-**	**1.0**	**0.1**	**0.3**	**Tr./-**
Translation /z	**0.5**	**2.1**	**24**	**1.3**	**Sup.**	**1.2**	**0.1**	**0.6**	**Tr./-**
Rotation /z	**0.5**	**2.3**	**00**	**-**	**-**	**0.7**	**1.1**	**1.9**	**Rot2./-**
Rotation /xy	**0.7**	**2.3**	**57**	**1.1**	**Sup.**	**0.9**	**0.1**	**0.9**	**Tr./Sup.**
Scaling	**0.4**	**3.0**	**19**	**2.5**	**Sym.**	**1.4**	**0.4**	**1.0**	**Sca2./-**
Obj. Deformation									
Extrusion	**-**	**1.6**	**42**	**1.4**	**Sup.**	**1.1**	**0.1**	**0.2**	**Tr./Sup.**
Bending /z	**-**	**2.2**	**19**	**1.3**	**Sup.**	**0.2**	**0.9**	**1.2**	**Rot1./-**
Bending /xy	**-**	**2.3**	**77**	**1.2**	**Sup.**	**0.7**	**0.2**	**0.2**	**Tr./Sup.**
Local Scaling	**-**	**2.4**	**21**	**1.8**	**Sup.**	**0.9**	**0.2**	**1.3**	**Sca2./-**
Deleting	**-**	**1.2**	**11**	**1.5**	**Sup.**	**1.1**	**0.0**	**0.2**	**Tr./-**
New Object	**-**	**2.6**	**17**	**1.3**	**Sup.**	**1.4**	**0.2**	**1.4**	**Sca2./-**
Object Selection									
Selection	**0.5**	**1.0**	**-**	**-**	**-**	**-**	**-**	**-**	***see 4.3.2**

However, many users interacted using both hands. From our observations, the non-dominant hand had two main functions: a support function (Sup.) (e.g., frequently indicating the parts of the scene that should not move by keeping a still hand on them); or a symmetric function (Sym.) (e.g., doing symmetric gestures with both hands for scaling). The support function is most frequently used, specifically on object manipulation tasks where it is used to maintain some objects or some part of the object of interest in place.

4.3.2 Modes Disambiguation

The vast majority of users (87%) performed ambiguous gestures, i.e., used similar gestures for two different tasks. This leads us to look for ways to disambiguate those gestures.

A first clue for disambiguation is the location of the fingers at the start of the gesture: the first finger is hardly put on or around an object when a navigation task is involved (distance > 1, Table 1), directly manipulating on the background image. Furthermore, the grid is sometimes manipulated to perform indirectly navigation tasks such as panning along the depth axis. On the other hand, object manipulations typically start in or nearby the object (distance < 1). Although this criterion enables us to distinguish navigation tasks from object manipulation tasks, further investigation has to be done to disambiguate object positioning from object deformations.

A second clue for disambiguation is the number of fingers used. The average number of fingers involved to navigate is about 3 while this number decreased to 2 for object positioning. Though, the non-dominant hand gives the most relevant number of fingers: 1 finger used for navigation, no finger for object positioning and 1 or more for object deformation. In a large proportion, the non-dominant hand fingers reached the border of the screen for navigation tasks when it has a support function.

Therefore, the different modes could be automatically distinguished during user interaction by mixing these two criteria: a finger-count method [13] would give the selected interface mode, while finger locations could tell to which object the interaction is to be applied, if not to the whole scene.

4.3.2 Group Selection

Another issue investigated is how a transformation could be applied to a couple of objects. The same gesture was usually performed for both objects (≈ 75% of users), each object involving one hand. But this does not scale to more than two objects, and cannot be applied to gestures requiring both hands.

Instead of simultaneously/sequentially manipulating the different 3D contents, fewer participants (≈ 20%) preferred to first select the object by clicking (or double clicking) before manipulation. Only two users performed a "lasso" gesture to select object before performing the transformation. After the object selection, the gesture was performed either on one of the object, or near the barycenter of the group. This leads us to conclude that a specific widget should be created to represent the selected group.

4.3.3 Scaling Interferences

During most tasks, the participants performed scaling phases while performing their interaction. In many cases, the DoF involved by scaling was meaningless.

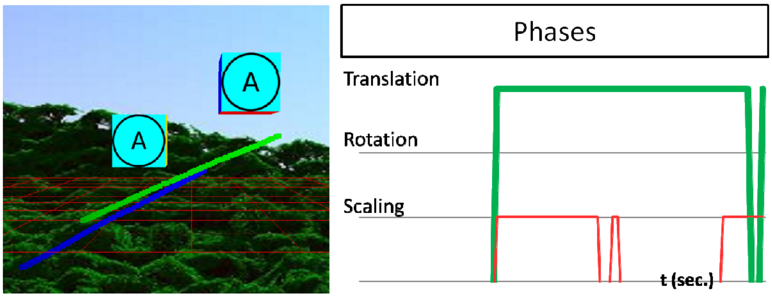


Fig. 9. Gesture for a translation task and its phases

For instance, Fig. 9 illustrates the gesture of a participant during a navigation task: a translation in the (x, y) plane. In this illustration, more than 90% of the motion was analyzed as translation phase, while short scaling phases occurred in parallel. As stressed when analyzing the first experiment, the stable and useful part of scaling motions usually takes place when the translation and rotation phases of a motion have ended. Therefore scaling phases should not be taken account when they occur concurrently to other phases, and the gesture in Fig. 9 should be interpreted as a bare translation.

4.3.4 Navigation Tasks: Zoom vs. Depth Axis Translation

Two consecutive trials were depth axis translation and zooming tasks. To distinguish the different kinds of trials, a background image was added to the 3D scene. Though, every user but two asked for the differences. Once answered, they mainly succeed to understand the shown transformation.

Moreover, we can also notice that, although they did know the difference (as they asked for it), half of the participants still performed the same gestures for both tasks.

4.3.5 Combining Different Manipulations

Some tasks of the experiment consisted in combining elementary motions – for instance, object translation and rotation. In order not to influence the participants, and leave them free to invent their own interaction mode, only a before/after animation was shown in this case. Analyzing data by phase analysis techniques enables us to easily distinguish whether users prefer to perform each “elementary” motion sequentially, or simultaneously. The results are gathered on Table 2.

Table 2. Table illustrating whether users prefer to separate the different motion (left), or not

Tasks	% Sequential motions	% Concurrent motions
Translation + z/Rotation	**58%**	**42%**
Translation + xy/Rotation	**79%**	**21%**
Translation + Scaling	**61%**	**39%**

In two third of the cases, participants preferred to decompose gestures into “elementary” ones. This is consistent with Martinet et al. work [19]. In details, performing a translation and a depth axis rotation are mainly decomposed into a translation and a rotation phases. The higher number of participants performing simultaneously these two phases (42%) are consistent to Wang and Nacenta works [18, 20], as these two phases slightly interferes each other. On the other hand, when translation is coupled with a rotation along the other axis, the phase analysis mainly identified two translation phases; the second phase corresponding to the second hand gesture: a trackball like rotation (see further details in the next section) [21].

4.3.6 Starting Finger Positions for Deformation Tasks

We already observe that starting positions of fingers is relevant in order to disambiguate navigation from object manipulation tasks. Further investigations about fingers starting positions have been performed for object deformation tasks.

When participants use their non-dominant hand, their fingers typically remain far away from part of the object that is deformed (even sometimes at the opposite side). By performing such gesture, participants keep in place the object, while she/he works on a region of interest of the object –such as designers keep in place their paper while drawing [22].

However, the dominant hand gestures are typically performed around the deformed object. For instance, on bending tasks, the thumb position corresponds to the center of rotation, and remains static, while a rotation gesture is detected by phase analysis (Table 1). Local scaling (such as stretching or compression tasks) is typically performed by a shrink gesture, where the gesture barycenter is located nearby the center of the part of the object that is being deformated.

4.3.7 Noticeable Gestural Design Pattern

As we already observed, a majority of users performed ambiguous gestures, and therefore the interface need some disambiguation between modes. On the opposite, we note that some manipulations can be linked together, enabling us to identify typical gestures and a gestural pattern for each mode.

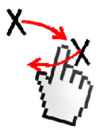
Gestures	Translation	Rotation 1	Rotation2	Scaling 1	Scaling 2*
Phase:	Translation	Rotation	Translation + Rotation	Scaling	Translation. + Scaling
Type:					

Fig. 10. Five typical gestures for one hand interaction, identified through our experiment. Scaling 2 is difficult to identify due to scaling interferences (see section 4.3.3)

Once phases are analyzed, hand gestures on a surface can be easily classified into 5 main classes (Fig. 10). Due to scaling interferences, no gesture is identified when all three phases are detected. On the contrary, the detected gesture is Rotation 2. In Table 1 (last column), we associate each task of the user study to the corresponding typical gestures for the first hand.

The gestural pattern is summarized in Table 3. On the first hand, manipulations that transform the scene/object on the 2D screen plane mainly used one-handed gestures (e.g., translation/extrusion along x, y axis tasks are performed by one hand translation gestures). Scaling manipulations can be gathered into two possible gestures, which both represent a shrink gesture, either performed by one or two hands.

Table 3. Table grouping a set of action – either usable on navigation tasks, or object positioning/deformation tasks – and the users associated gestures

Action	**Translation / xy**	**Translation / z**	**Rotation / z**	**Rotation / xy**	**Scaling / Zoom**
Gestures (Phases)	Translation	?	Rotation	Translation + Support	1 or 2 handed Shrink gesture.

On the other hand, manipulations that required depth axis motions need more attention. For instance, rotation tasks are usually performed with two hands: one hand is keeping the object in place, while the second hand is "pushing" the object, like in the trackball technique [21]. Though, manipulations that correspond to a translation along depth axis are outsiders: no gesture was consistently used to perform these tasks.

Using such a gestural design pattern for all 3D multi-touch interfaces would be a real advantage, since users would need to learn the pattern only once, and would immediately be efficient with new tools.

5 Comparison with, and Application to Previous Work

5.1 Other Multi-touch Gestures Analysis: Cohé and Hachet Work

Cohé and Hachet recent research lead them to another approach of understanding gestures for manipulating 3D contents [17]. Their paper focused on object positioning tasks.

Their approach was to classify gestures using three parameters: form, initial point locations, and trajectory. They identified gestures by the number of moving/unmoving fingers (the form), their starting locations (initial point), and the kind of motion (trajectory), while exploring object translation, rotation or one axis scaling tasks.

For those tasks, while we used a different methodology, our results are largely consistent with their findings: in our case, form and trajectory parameters are considered by the phase analysis. Nearly all their classifications are coherent with the gestural pattern that we defined above. For instance, their rotation gestures (except for R3 and R8) are identical to our rotation phase. Moreover, both papers observe that a majority of users prefer to start on or nearby the object.

The main difference is the parameters used to define the starting locations. While we only defined the neighborhood of object to distinguish between modes, they divided this parameter according to cube elements (faces, edges, corners and external). Both classifications bring their own advantages. Using cube elements to directly manipulate complex 3D content such as large triangular meshes would be meaningless. On the opposite, manipulating 3D content with 3D transformation widgets could always make use of cube like widgets, and therefore use the proposed decomposition.

5.2 Direct Interaction Techniques: 1-, 2- and 3-Touch Techniques

A first kind of 3D manipulation is interactions that are directly performed onto the objects. Hancock and Cockburn researches identify 3 techniques, based on the number of fingers used (which are extended by Martinet's works for depth axis translation) [4, 6] .

Their paper is focused on the comparison between three techniques that enable users to perform translations and rotations. The first technique, involving only one-touch interactions, corresponds to an extension of the RNT algorithm [5]. By doing so, the interface can manipulate 5 DoF with a single finger. The second technique, involving two touch interactions, the first finger correspond to the RNT algorithm for translations and yaw motions, while the second finger is used to specify the remaining rotations. The last technique maps each group of motion to a specific finger – translation to the thumb, yaw rotation to the second finger and the remaining motions to the last finger.

It is noticeable that they stop their comparison up to three-finger techniques that corresponds to our effective upper bound number of fingers. They compared the three techniques in two experiments. For both tasks, they concluded that the three-touch technique was the fastest to use, while the one touch techniques was the less efficient one.

We will further focus on the differences between these methods, compared with our phase analysis method. Even though the one touch technique is the most stable gesture (as it can only provide translation phases), the technique suffers of DoF distinctions: all interactions are mapped to the same gesture. On the other hand, the three-touch technique easily decomposes translation and z-axis rotation to translation and rotation phases into the two first finger motions. Translation and rotation phases can be mainly performed at the same time, with little interference between them, so users are more efficient while performing such techniques.

Though, the last finger suffers from the same issue on two and three touch techniques. Indeed, as the roll and pitch rotation are mapped in the Cartesian frame, rotation and scaling local phase are mixed during the last finger gestures. Therefore, performing pure roll or a pitch rotation are interfering each other.

5.3 Indirect Interaction/Widget Technique: tBOX Analysis

Another kind of 3D manipulation involves a widget that acts as a proxy to the real object. 3D transformation widgets are commonly used in 3D applications. A recent example of 3D transformation widgets for multi-touch devices is the tBOX [7]. To easily manipulate 3D objects, they are enclosed in their bounding box that is made interactive. This is an extension of the standard manipulation widget (represented by 3 arrows). The existing manipulations on objects are translation, rotation and scaling. All user gestures have to involve the cube widget – specifically the vertices, edges or faces of the cube. For instance, pushing a single edge performs rotations, while translating along edge performs translations. A shrink gesture on both sides of the tBOX widget represents a single axis scaling.

The first observation about tBOX, once analyzed into phases, is that all object manipulations are translation phases only (scaling corresponding to one hand translation and a symmetric second hand role). In terms of stability, such gestures are the most efficient, as no interferences can occur. Moreover, such widget leaves a lot of possible interactions for other manipulations (such as deformations).

To the tBOX authors mind, one goal of their interactions was to discriminate between rotation and translation. Therefore, users cannot efficiently switch between these two manipulations: they have to stop their first gesture and reach again the required edge. On the other hand, phase analysis based interface would permit to easily switch between these manipulations, maybe at the cost of stability.

6 Discussion

Theoretically, multi-touch devices offer the possibility of manipulating 3D scenes while simultaneously controlling many DoF: up to 20 actually, if the two hands were used. However, this upper bound is never reached. Because of the interferences between fingers and to their restricted motion when moved in contact with a plane, complex gestures involving all fingers are often unstable, and the time it takes to perform them would be prohibitive for an interactive use.

As shown by the second experiment, users easily invent gestures to interact with 3D content. Quite interestingly, they tend to use all fingers for global hand gestures such as translation, rotation, and scaling, although two or three fingers would be sufficient (in this case, using all fingers is easy, since there is no local hand motion to control). For more complex interaction gestures, users naturally limit themselves to one to three fingers per hand. This leads us to the following methodological rules when designing 3D interaction on a multi-touch table:

- Firstly, the number of DoF effectively controlled by the user (never more than 8 for the two hands in our experiments) is actually much smaller than the number of DoF required for navigating, plus moving and deforming objects in a 3D scene. Therefore, using an interaction system based on several interaction modes is mandatory.
- Secondly, the number of fingers actually on the device during the interaction gesture could be easily used to distinguish between simple navigation tasks, and more complex object positioning/deformation tasks: full hand interaction could be used to select and control navigation, since simple global gestures, which the user preferably performs with all fingers, are sufficient in this case. For object

manipulation/editing tasks, the interface could disambiguate the required mode by counting the number of finger on non-dominant hand.

- As noted in our experiments, the location where the gesture starts is often meaningful: users typically use it to select the object to which the action is applied. In addition to controlling object selection, the hand location at the start of the gesture could be another way of automatically selecting between navigation (if the gesture starts on the background) and object positioning (with some limitation for crowded scene, where some free background space would need to be artificially preserved for navigation).
- Global phase analysis is quite coherent for mapping gestures and tasks: gestures are easily classified. Even more, a design gestural pattern for 3D contents manipulations emerged from the experiments, which are reproduced inside each tested mode, and could be extended to any other 3D content transformation mode. Though, scaling phases should be analyzed independently, when the other gesture phases have stopped, as they can be produced as side effect of other phases.
- Lastly, using the full hand to grab groups of objects on which to apply a gesture (such as all the objects covered by finger tips, or by the convex envelop of finger tip positions) would be a further extension of this technique. However, extra gestures such as double-clicking with a finger, or circling the object to select it (as done by some of our users), would be needed to add distant objects to the group.

7 Future Work

The first goal of this paper was to understand hand gestures on a surface. The phase analysis technique we proposed provides a simple, yet consistent way to analyze and classify gestures, especially regarding global hand motion. Therefore, an interesting direction for future research would be to develop new interaction methods directly relying on such phase analysis to drive task control.

Acknowledgements. This research was partially funded by ERC advanced grant EXPRESSIVE and by the G-INP BQR Intuactive.

References

1. Wilhelmi, B.J.: Hand anatomy (2011)
2. Santello, M., Flanders, M., Soechting, J.F.: Postural hand synergies for tool use. J. of Neuroscience 18, 10105–10115 (1998)
3. Martin, J.R., Zatsiorsky, V.M., Latash, M.L.: Multi-finger interaction during involuntary and voluntary single finger force changes. Exp. Brain Research 208, 423–435 (2011)
4. Hancock, M., Carpendale, S., Cockburn, A.: Shallow-depth 3d interaction: design and evaluation of one-, two-and three-touch techniques. In: Proc. CHI 2007, pp. 1147–1156 (2007)
5. Kruger, R., Carpendale, S., Scott, S.D., Tang, A.: Fluid integration of rotation and translation. In: Proc. CHI 2005, pp. 601–610 (2005)

6. Martinet, A., Casiez, G., Grisoni, L.: The design and evaluation of 3d positioning techniques for multi-touch displays. In: Proc. 3DUI 2010, pp. 115–118 (2010)
7. Cohé, A., Decle, F., Hachet, M.: tBox: A 3D Transformation Widget designed for Touchscreens. In: Proc. CHI 2011, pp. 3005–3008 (2011)
8. Grossman, T., Wigdor, D., Balakrishnan, R.: Multi-finger gestural interaction with 3d volumetric displays. In: Proc UIST 2004, pp. 61–70 (2004)
9. De Araùjo, B.R., Casiez, G., Jorge, J.A.: Mockup builder: direct 3D modeling on and above the surface in a continuous interaction space. In: Proc. GI 2012, pp. 173–180 (2012)
10. Francone, J., Bailly, G., Lecolinet, E., Mandran, N., Nigay, L.: Wavelet Menus on Handheld Devices: Stacking Metaphor for Novice Mode and Eyes-Free Selection for Expert Mode. In: Proc. AVI 2010, pp. 173–180 (2010)
11. Hancock, M., Hilliges, O., Collins, C., Baur, D., Carpendale, S.: Exploring tangible and direct touch interfaces for manipulating 2D and 3D information on a digital table. In: Proc. ITS 2009, pp. 77–84 (2009)
12. Scoditti, A., Vincent, T., Coutaz, J., Blanch, R., Mandran, N.: TouchOver: decoupling positioning from selection on touch-based handheld devices. In: Proc. IHM 2011, pp. 37–40 (2011)
13. Bailly, G., Müller, J., Lecolinet, E.: Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. IJHCS 70, 673–689 (2012)
14. Bailly, G., Demeure, A., Lecolinet, E., Nigay, L.: MultiTouch menu (MTM). In: Proc. IHM 2008, pp. 165–168 (2008)
15. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092 (2009)
16. Morris, M.R., Wobbrock, J.O., Wilson, A.D.: Understanding users' preferences for surface gestures. In: Proc. GI 2010, pp. 261–268 (2010)
17. Cohé, A., Hachet, M.: Understanding user gestures for manipulating 3D objects from touchscreen inputs. In: Proceedings of the 2012 Graphics Interace Conference, pp. 157–164 (2012)
18. Nacenta, M.A., Baudisch, P., Benko, H., Wilson, A.: Separability of spatial manipulations in multi-touch interfaces. In: Proc. GI 2009, pp. 175–182 (2009)
19. Martinet, A., Casiez, G., Grisoni, L.: Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. IEEE Transactions on Visualization and Computer Graphics 18, 369–380 (2012)
20. Wang, Y., MacKenzie, C.L., Summers, V.A., Booth, K.S., et al.: The structure of object transportation and orientation in human-computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 312–319 (1998)
21. Chen, M., Mountford, S.J., Sellen, A.: A study in interactive 3-D rotation using 2-D control devices. In: ACM SIGGRAPH Computer Graphics, pp. 121–129 (1988)
22. Guiard, Y.: Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. Journal of Motor Behavior 19, 486–517 (1987)

Considering Communities, Diversity and the Production of Locality in the Design of Networked Urban Screens

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

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

Abstract. Highly diverse settings such as London (with people from ~179 countries speaking ~300 languages) are unique in that ethnic or socio-cultural backgrounds are no longer sufficient to generate a sense of place, belonging and community. Instead, residents actively perform place building activities on an ongoing basis, which we believe is of great importance when deploying interactive situated technologies in public spaces.

This paper investigates community and place building within a complex multicultural context. We approached this using ethnography, complemented with workshops *in the wild*. By studying the relationships arising between different segments of the community and two networked screen nodes, we examine the place building activities of residents, and how screen nodes are incorporated into them. Our research suggests that urban screens will be framed (and eventually used) as part of this continuing process of social, spatial and cultural construction. This highlights the importance of enabling socially meaningful relations between the people mediated by these technologies.

Keywords: Diversity, communities, ethnography, workshops, in the wild, urban screens.

1 Introduction

When studying the deployment of interactive screens in urban public settings [14-17] it is important to consider not only their immediate use and spatial location [2, 13, 17], but also their assimilation within the wider socio-cultural context of the locale. This is of great importance because the interplay between these spatial and cultural elements can strongly influence the long-term success of these technologies. One way in which this interplay can be addressed is by investigating the *production of locality* [1, 5, 18]. This refers to the processes that communities use to build a sense of place; in other words, how the communities create cultural conventions to control both space and people's behavior within them in ways that allow individuals to develop their own identities in relation to both the locale and other people [1]. By defining who has access to what spaces, at what times, under what conditions and based upon which

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 315–322, 2013.

narratives or behaviors, a community can define itself in terms of its conventions, practices, values and aspirations, hence providing a shared social identity.

Place building activities are varied, and they are as much about appropriation as they are about the negotiation and control of space when interacting with other people. For instance, Dant and Deacon observed that homeless people living in hostels emphasized control over what they were allowed to do, such as being able to make a cup of tea whenever they wished, as a way to transform an impersonal space into a place they felt they belonged to [4]. Similarly, some residents in the London neighborhoods where we worked emphasize a degree of control over certain public spaces (e.g. Library halls) to install art exhibitions for public display. In these two cases, certain practical actions (such as making tea or periodically renewing art exhibits) enable social exchanges and conviviality, which translate into a feeling of commonality that underpins the wider social structures that support the community. The analysis of place building is of particular interest in locales where there is a large heterogeneity in ethnicity, nationality, language, religion, educational level and income. In these highly diverse contexts [19] a sense of place does not arise implicitly as a simple product of long held traditions; instead, the local inhabitants must explicitly create it through active efforts of place building [1].

The literature on urban screens has explored the crossroads between communities, public space, technical factors and social interaction around interfaces [6, 10, 12, 16, 18]. Although some works address the social aspects of screen use, the study of how wider socio-cultural variables influence long-term community support of urban screens remains an open question. In general, there has been an emphasis on social interactions mediated through screen content on shorter timescales [13, 15, 17], without the consideration of longer term, wider scope socio-cultural variables. Only very few longitudinal case studies have taken a wider contextual approach such as [14, 18].

The present paper argues that this wider social context matters if one wants to understand the different perspectives and behaviors surrounding long-term community support for public interactive screens. To study the ways in which such wider cultural context affects screen node use, we explore how urban, networked screens are integrated within the cultural practices of a highly diverse urban locale in London. We examine the social aspects of screen use through an analysis of place building practices, and inquire into the reasons why certain people emphasize certain uses for the screens as opposed to others. Our research takes a targeted, longitudinal approach to understand a complex range of social, technical and interactional issues [2, 6, 12].

Our data gathering methodology includes ethnography and workshops *in the wild*. The main contribution of this work is to show how action, space and sociality mutually impinge upon one another forming a wider cultural logic, which in turn has a great influence on how a network of public interactive screens is experienced, treated, understood and embraced in practice.

The structure of the paper is as follows. First, we describe our methodology in Section 2. In Section 3 we discuss the integration of screen nodes within pre-existing place building activities. We provide further analysis and conclusions in Section 4.

2 Methodology

Our project involves a network of four interactive touch screens designed to encourage public participation and to explore how networked urban screens can augment urban experience to support communities and culture [6]. There were three main screen node applications discussed with our participants during workshops and ethnography: *Slideshow*, *SoundShape* and *ScreenGram*. Slideshow simply loops through a set of images. *SoundShape* enables people connected remotely to create a collaborative musical pattern. ScreenGram leverages common technologies (e.g. Twitter) to enable users to upload images to the screen nodes. For further description see [12].

Diversity is a fundamental consideration of this project: two of these screens are located in London, which is home to people from ~179 countries and speaking ~300 languages [19]. Furthermore, the borough where these are located (Waltham Forest) is one of the most ethnically diverse in Britain [22]. One of our screen sites is *The Mill* in Walthamstow, a non-profit organization that runs a building that has become a hub for local communities; the second one, *Leytonstone Public Library*, fulfills a similar social role. Amongst many other uses, locals use these two sites to hold business meetings, playgroups and art exhibitions, and also to give and/or take free courses. Both The Mill and Leytonstone Library are popular public places where people organize and enact initiatives that actively build a sense of local identity (a *sense of place*) for both their neighborhoods and the borough itself. Both sites bring together a diverse cross-section of local residents to pursue varied social activities, and hence represent a natural sample of people present in the locality.

In order to engage with a wide range of social groups within these diverse communities, we carried out ethnography and conducted two workshops in Leytonstone and four in Walthamstow. Rather than defining our groups in terms of statistical or demographic properties, we chose to target "grassroots" groups that were already being sustained through various degrees of involvement with the locality. This enabled us to interact with people in their usual social configurations and pre-established networks of relationships. A total of 70 people from the locality participated in the workshops. These people came from a wide variety of cultural backgrounds and age brackets; with roughly equal numbers of toddlers, children, teenagers, adults and seniors. In addition to the workshops, we also carried out ethnography, consisting of observations during the various stages of implementation for over a year, and conducted 64 semi-structured interviews with locals belonging to different social groups. In all our research encounters we discussed a broad selection of issues that included the locale, the screen, and issues relating to a sense of community and belonging.

Each one of the workshops was built around the activities of one or more local groups with which the researchers established contact after engaging ethnographically with the locale for ~8 weeks. These groups emerged from social meeting grounds such as churches, local schools, activism societies and hobbyist communities. Each group was chosen to be not only as homogeneous as possible, but also to fit with other groups so that, in aggregate, it would provide as wide a view into the dynamic of the locale as possible. Although we did not specifically create our groups on this basis, as the research progressed we found that age provided a convenient means to identify sets of groups which manifested consistent behaviors. Hence, regarding the analysis of our data, we chose to divide the whole population into three age-related groups: *children/teenagers*, *adults/young adults*, and *seniors*. This classification, although mainly intended to facilitate data analysis, is supported by the social dynamics of our sites. For instance, a group of children in the locale that participated in our workshops also participate in playgroups with their parents. Another example of a local group that participated in our workshops is *The Recycled Teens,* a group of seniors that gathers every Tuesday in a senior residence to sing old tunes, watch old films and discuss current affairs in relation to past events. Of course, some groups are not as homogeneous as these. An example of this kind is a knitting group, that includes participants ranging from children as young as 6 to others over 65 years old. In between creating scarves and hats, they discuss local, national and international politics. They use their knitting to support and draw attention to causes. Also this group participated in one of our workshops. In the following section we discuss how our informants integrated screen nodes within these pre-existing place-building activities. We will follow the three age-centered categories we defined earlier in this section.

3 Embedding Screen Nodes within Place Building Activities

3.1 Screens Approached through *Competitive Physical Action*

One of the ways in which children and teens generate a sense of belonging, community and place is by playing in public spaces after school hours. These include playgrounds, parks, markets, and sites like The Mill and Leytonstone Library. It is here where they meet friends that can become an integral part of their life. Discovering and mastering the spatial landscape and its community helps children and teenagers build their personal and social identity, while at the same time, helping them feel their neighborhoods as actually *theirs*. In a way, children and teens seek to assert certain degree of control over their spaces and their social experience within, eventually claiming these places as an extension of their own selves. To a great extent, this is done through a highly embodied, physical engagement. This means that a park or public space only becomes theirs *in action* [20]: by wandering around, playing in it, and interacting with friends and family both *in* the space and *with* the space.

From both ethnographic observations and workshops we noticed that, when interacting with the screen node, children and teens use the same action-centered approach that they use when they make public spaces theirs. They like to explore the opportunities provided by the technology: the sensitivity of the touch foil, the fluidity in application response, and their ability to exert control over the system. Whether it is by simply moving the pointer or by systematically exploring the possibilities in touch response, their approach is dominated by the actions that they can perform rather than by the explicit interactions that the developer envisaged for the application running in the node. Children and teens hence use the node in novel ways, inventing *competitive* games that rely on physical interaction (e.g. free pointer motion) and play them with others. In this way they create their own uses for the screen, which are quite independent from those intended by the application designers (who in our project imagined mainly *collaborative,* rather than competitive applications). In this case, users invent competitive games because it is easier for them to embed competition within their playground politics, as compared with the abstract collaboration ideal presented by the application itself. In both of our sites children and young teens attend playgroups where they learn to share the space with others; they enjoy books, toys and resources that are not theirs, but which they can borrow and play with. These objects and spaces are not seen as alien, but as *personal*, in that "public property may become viewed as private possessions and thereby potentially contribute to a sense of self" [3]. Children frame the screen node using a similar collective ownership approach, and they treat it as personally enjoyable, yet community-owned. For these participants, the screen node is experienced in the same way as a park or a playgroup, and social behavior is enacted accordingly.

3.2 Screens Used as Springboards for Local Culture

When constructing a sense of belonging, adults emphasize less the physicality of the environment and take a more instrumental view. They are interested in using the screen node to create what they consider a positive environment by improving education, increasing historical awareness or fostering better communication flows between different groups within their neighborhoods. This place building perspective has resulted in the formation of many of initiatives, from the creation of grass-roots business improvement districts, to art trails and community centers. In this context, generating a "Walthamstow/Leytonstone culture" is relevant to adults as an intellectual, symbolic and material practice that not only enables the production of a sense of place, but also promotes "the human capacity to expand worlds towards other potentially distant horizons and more complex outcomes of life" [11]. For our adult participants, place building is an opportunity for the purposeful pursuit of change in the locality; it is a *scaffold* [8], an instrument to mediate between their imagined view of their locality and the reality in which it stands. Hence, our adult participants see their place building activities as a springboard to generate local culture and bridge the gap between what their locale *is* and what it *could become*; a stage for the display of the unrealized potential development of the locale. For instance, those explicitly engaged in transformative activities (such as local activism) wanted to use it to extend their efforts, informing and educating the local population about relevant projects. Members of the Knitting Group, expressed this view, urging researchers to upload to ScreenGram images of knitted characters they were selling to raise funds for their local community center. Local artists thought that the node could extend the available physical exhibition spaces by imagining it as a window to display digital art. The managers of The Mill imagined the node as a virtual notice board, extending their own functions and enabling users to connect with each other even if the building itself was closed. Even those adults not involved in specific community-building projects identified the node as an opportunity for dissemination of community ideas, particularly as outlets of local useful information: recommendations on what to do in the area, news on what is happening and mechanisms to connect with other locals.

As shown above, most adults in our workshops and interviews reported that they expected the screen node to embody a higher purpose. They approached the screen as an extension of their pre-existing place building endeavors, uploading photos or short messages that were relevant to their community. For many, the mere physical presence of the node itself legitimized previous efforts for place building, showing them to have been efficacious, as new resources for the realization of the community had been produced. These locals presented the node in local events with pride, as a symbol of their achievements to bring the community closer to its unrealized potential.

3.3 Screens Used as Collective Memory Fixing Mechanisms

One important way in which seniors generate a sense of place is through remembrance narratives involving the past of the neighborhood and their lives within it. This activity is one of *anchoring* [7], in which subjects bring the past to inform the present, and ensure that valuable historical lessons remain with us. Although this is particularly visible when they interact with each other during community group meetings, these narratives are also important in other contexts. Whether sharing these stories with children in local schools or with local historical societies, seniors enjoy talking about the past in public spaces.

Given this predilection for reminiscence, our senior participants enjoyed the node applications that enabled them to preserve and share memories. Hence, they quickly identified the ScreenGram and Slideshow applications as a potential collective photo-album, which could help preserve and rework memories, acting as a vehicle for their remembrance narratives. In addition, since the screen node is located within one of the social hubs in which our senior participants congregate, it can also serve as a focus point to talk about these memories. Hence, the visual elements of the screen facilitate the verbal storytelling with which they usually accompany photographs, and which they traditionally employ to build a sense of belonging.

Although seniors were not able to interact easily with the screen node (researchers helped them upload images and display them in the node), they enjoyed the social opportunities provided by using the node to share images in their community group. Rather than a superficial feeling derived from watching a large number of generic images, these participants preferred to dwell for longer in specific images that held importance for lives of others in the neighborhood; they preferred the images that better allowed them to retain collective memories, rather than those that simply had aesthetic appeal. They stressed that they would have liked to have a pause button to dwell for longer on specific images and discuss them, rather than having an automatic timed loop. Capturing an image in a photograph, and re-capturing the moment by pausing its display on the screen node, enhances the process of building intimate connections with the place and with other people. Hence, for these users remembrance is less about simple archiving and more about reclaiming the experiences of others as own, and enabling their own experiences to be reclaimed in the same way by others. Hence, our senior participants expressed a need to discuss specific images that participants of other ages did not require, and which directly connected with their place building efforts. Screen nodes helped seniors share their place building perceptions with each other, reinforcing their commonalities and leading to a more rewarding sense of place identity based on collective participation and mutual reinforcement.

4 Discussion and Conclusions

When interacting with the screen nodes, most participants (irrespective of their age) enact *performances* that connect them to their existing place building practices. These performances consist of patterns of behavior "whose repetitions situate actors in time and space, structuring both individual and group identities" [9]. Although these performances are based on repetition, mimicry and reproduction of social interaction in the neighborhood, they are also varied: to an extent, each performance is unique, allowing locals the expression of individuality even when following normative patterns of interaction. On the one hand, the repetition associated with each performance serves to deeply embed patterns of interaction and produce a shared sense of place. On the other, the freedom that each performance affords the individual can become a vehicle for the exploration of the relationships between the individual and the community. Hence, these performances are both agents of social change and aids to reproduce current social structures of interaction.

Although anchoring, scaffolding and action based approaches are universal human experiences regardless of age, we found that the behavior of our participants towards the screen could be understood (albeit in a simplified manner) by considering one of these approaches as dominant. We posit that this simplification can provide useful guidance when designing applications for public interactive screens, particularly regarding their longer-term integration with the social dynamics of a locale.

In this paper, we emphasized the commonalities between regular place building practices and the performances that locals enacted when interacting with screen nodes. To understand this performance, simply conducting interviews and recording the spoken word is not enough, as it misses the sociological and cultural milieu in which the performance acquires its meaning and cultural significance. Hence, we argue that understanding the wider context through ethnography and workshops *in the wild* provide an invaluable tool to make sense of the ways in which technologies become embedded within the place building practices of a locale.

Place building is directed towards the generation of a *structure of feeling* [21]; a deep subconscious familiarity with our everyday living spaces. Hence, it is not a purely emotional construction: it is also physical, leading to a deep-seated bodily experience in which the residents *belong* in the space and are able to navigate it effortlessly. It depends not only on the values, beliefs and customs of locals; it also depends on the unique ways in which their pre-existing place building practices create a bridge between their culture and the locale in which they live. In this sense, place building takes the spatial analysis of interaction with users and extends it with cultural dimensions, making it a useful addition to the tools that can be applied when considering the deployment of a technology in a given public setting.

Acknowledgements. This work is funded by the UK DE Program grant EP/I031839/1 and EP/I031413/1. We would like to specially thank our host venues and research participants.

References

1. Appadurai, A.: Modernity at Large: Cultural Dimensions of Globalization. Minnesota Press, USA (1996)
2. Behrens, M., et al.: Exploring the Effect of Spatial Layout on Mediated Urban Interactions. In: Proc. of the ACM Int. Symp. On Pervasive Displays, PerDis (2013)
3. Belk, R.W.: Possessions and the extended self. Jrnl. of Consumer Res., 139–168 (1988)
4. Dant, T., Deacon, A.: Hostels to homes: The rehousing of single homeless people. Avebury/Gower, Aldershot (1989)
5. Dourish, P.: Re-space-ing place: 'place' and 'space' ten years on. In: Proc. of CSCW, November 4-8 (2006)
6. Gen, F., Schieck, A., et al.: Research in the Wild: Exploring the potential of Networked Urban Screens for Communities and Culture. In: DIS Workshop: Research in the Wild, Newcastle, UK (2012)
7. Giddens, A.: The Construction of Society. Polity, Cambridge (1984)
8. Greenfield, P.: A Theory of the Teacher in the Learning Activities of Everyday Life. In Everyday Cognition, pp. 117–138. Harvard University Press (1984)
9. Kapchan, D.: Performance. Journal of American Folklore 108, 479 (1995)
10. Memarovic, N., et al.: Using public displays to stimulate passive engagement, active engagement, and discovery in public spaces. In: Proc. of MAB (2012)
11. Miller, D.: Tales from Facebook. Polity (2011)
12. North, S., et al.: Tension Space Analysis: Exploring Community Requirements for Networked Urban Screens. In: Proceedings of IFIP INTERACT. LNCS. Springer (2013)
13. O'Hara, K., et al.: Understanding collective play in an urban screen game. . In: Proc. CSCW 2008, San Diego, CA, USA (2008)
14. Redhead, F., et al.: NNUB: the neighbourhood nub digital noticeboard system. In: Proc. of ACM OzCHI, pp. 418–419 (2010)
15. Peltonen, P., et al.: It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In: Proc. of ACM SIGGHI (2008)
16. Satchell, C., et al.: Suburban nostalgia: the community building potential of urban screens. In: Proc. of ACM OzCHI (2008)
17. Taylor, K.: Programming video art for urban screens in public space. First Monday, Special Issue #4: Urban Screens (2006)
18. Taylor, N., et al.: Social interaction around a rural community photo display. International Journal of Human-Computer Studies 67(12), 1037–1047 (2009)
19. Vertovec, S.: Towards post-multiculturalism? Changing communities, conditions and contexts of diversity. International Social Science Journal 61(199), 83–95 (2010)
20. Warnier, J.P.: A praxeological approach to subjectivation in a material world. Journal of Material Culture 6(1), 5–24 (2001)
21. Williams, R.: The Long Revolution. Pelican, Harmondsworth (1961, 1984)
22. (2011), http://www.ons.gov.uk/ons/guide-method/census/2011/index.html

Growing Existing Aboriginal Designs to Guide a Cross-Cultural Design Project

Margot Brereton[1], Paul Roe[1], Thomas Amagula[2], Serena Bara[2], Judy Lalara[2], and Anita Lee Hong[1]

[1] Queensland University of Technology, [2]Anindilyakwa Land Council
{m.brereton,p.roe,anita.leehong}@qut.edu.au

Abstract. Designing across cultures requires considerable attention to inter-relational design methods that facilitate mutual exploration, learning and trust. Many Western design practices have been borne of a different model, utilizing approaches for the design team to rapidly gain insight into "users" in order to deliver concepts and prototypes, with little attention paid to different cultural understandings about being, knowledge, participation and life beyond the design project. This paper describes a project that intends to create and grow a sustainable set of technology assisted communication practices for the Warnindilyakwa people of Groote Eylandt in the form of digital noticeboards. Rather than academic practices of workshops, interviews, probes or theoretical discourses that emphasize an outside-in perspective, we emphasize building upon the local designs and practices. Our team combines bilingual members from the local Land Council in collaboration with academics from a remote urban university two thousand kilometers away. We contribute an approach of growing existing local practices and materials digitally in order to explore viable, innovative and sustainable technical solutions from this perspective.

Keywords: Cross-cultural, Aboriginal, slow design, sustainable design, digital noticeboards, urban screens, interface design, Human-computer interaction.

1 Introduction

Aboriginal communities in Australia see great potential for technologies to engage their youth, preserve their language, strengthen their culture and assist their communication within and outside their communities. The Anindilyakwa Land council of Groote Eylandt together with academic researchers has embarked upon a project to design "digital noticeboards" suited to and in service of the Warnindilyakwa people to support bilingual communication about news, culture, health, education etc. However, effective introduction of technology in a remote community requires consideration of how it will fit and thrive amidst the everyday practices of the Warnindilyakwa.

In this paper we examine design practice across cultures in general and then introduce the Warnindilyakwa people of Groote Eylandt. We then explain our research approach for the context of Groote Eylandt. This provides the background for our approach of extending existing Indigenous designs digitally in order to (a) best

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 323–330, 2013.

engage the expertise, cultural knowledge and aspirations of the local people and (b) demonstrate technological possibilities in a concrete and culturally relevant way.

2 Design Practice across Cultures

There is recent recognition of the problem of design practices not translating across different cultures and settings [6,14]. Winschiers demonstrated that common Western Participatory Design (PD) methods such as workshops and brainstorming were incompatible with the socio-cultural habits of Namibian participants [14]. Irani et al [10] argued that methods are transnationally produced and dynamic, and we can expect that problems will arise if methods are assumed to move easily and stably from one setting to another. Irani et al call for embracing heterogeneity in design, thinking about design in terms of engagements between different groups, the complexities of articulating perspectives and implications of translation between sites.

Brereton and Buur [5], in moving beyond the idea of users and defined participant groups, found that *“new formats of participation can be characterized by their sensitivity towards new types of networked relations among people, the diverse motivations of people to participate, the subtle balance of values and benefits involved in collaborative endeavours and the inherent power relations between participants”.*

Different cultures have understandings about being, knowledge, participation and life. Winschiers et al [15] in the African context also found it is more useful to emphasize communities than individual users. In Sub-saharan Africa, “the way of life is deeply rooted in a paradigm of ‘connectedness of all’ expressed in the aphorism, ‘a person is a person through other people’ ”[15]. When participation is already a core value in a community, the role for researchers engaging in design with a community is to read and respond to community practices and way of life such that they themselves become “participated”.[15] There is thus recognition among engaged design practitioners in developing contexts that hybrid practices in technology design must emerge that are sensitive to the contexts and networks of relations at hand. Moreover understanding of the evolving context and networks of relations takes time to develop and hence design is worth growing over time [3].

3 Groote Eylandt and the Warnindilyakwa People

The Groote Eylandt Archipelago is a unique and diverse environment in the Gulf of Carpentaria off the northern coast of Australia; it is the traditional homeland of the Warnindilyakwa people. The Warnindilyakwa people were brought to Groote Eylandt on a series of songlines [16] which created the land, rivers, animals and people and which named everything pertaining to the region. The language, “Anindilyakwa” is spoken by the 14 clan groups, which make up the two Moieties on Groote Eylandt. There are approximately 1400 Warnindilyakwa people living in three communities, and they are all formally related. Traditional collective culture governs much of the Warnindilyakwa people’s lives, but people in this region endeavour to “combine a traditional lifestyle with the comforts of the 21st century" [7] building upon the opportunities presented largely by manganese mining and tourism.

The history of Groote Eylandt has changed dramatically over the past century [8], with a nomadic lifestyle and oral culture replaced by a more Western one bringing many problems as well as benefits. The cumulative effects of long-term disengagement between governments and the three Anindilyakwa communities on Groote Eylandt has led to poor comparative socio-economic outcomes [7,8].

The Anindilyakwa Land Council (ALC) is a progressive land council that has initiated approaches across the spectrum to reconnect and re-engage the people of Groote. The executive membership consists of elected members, Elders from each of the 14 clans on the island. Contemporary challenges include: poor school attendance rates and declining engagement of youth in Warnindilyakwa culture, ceremony and history.

Fundamental to the ALC's initiatives is to seek ways of using technology within an Aboriginal context to support communication and connection on Groote. New digital technologies have the potential to address isolation problems leading to improved education, health and employment outcomes and improved cultural exchange and understanding [9]. Schools have adopted iPads in the classroom and many Aboriginal residents own iPads (over 100 owned by the 1400 residents).

With limited internet access, low literacy, an oral language tradition and a collective Aboriginal culture that is fighting to remain vibrant and sustainable, there is considerable work to do to understand how to best to support communication and connection through technologies. In summary, the Warnindilyakwa people seek to embrace the opportunities and challenges of the modern world by maintaining a firm footing in both Warnindilyakwa traditional culture and Western culture and seek to exploit technologies in order to do this. [1]

4 A Research Approach for the Groote Context

Groote Eylandt is a cultural and environmental jewel. As such the Warnindilyakwa have been extensively researched, often by fly in fly out researchers, with little perceived benefit to the community. It is important that new knowledge and technologies are owned by the people and that the project is sustainable: not an expensive toy. So in this project the team has tried to take a different approach to design "of the noticeboards" by growing existing design practises and thereby ensuring sustainability, familiarly and above all ownership of the project and its design by the community, building upon the ethos of Participatory Design as construed by Beck [1].

Previously we have we emphasized the importance of building relationships across cultures through the Aboriginal cultural practice of "yarning" and time spent together in hands on practical activities [4]. As academic researchers one of us was first

[1] For the purpose of definition, we take "culture" to simply mean the way of life of a group of people. People by virtue of their family history, environment, relations, hobbies and work may belong to many different cultures and sub-cultures. Particular design questions and participating groups tend to bring into focus different cultural perspectives.

engaged in work on Groote to address the environmental threat of cane toad invasion, a non-native species that is a pest in Australia. This led to the deployment of environmental monitoring projects in collaboration with the local Indigenous Ranger groups. Such practical projects involve time spent together engaged in activities on country and in the lab and time out in tea breaks. This provides time for talking, noticing each other's different ways, and helping each other out. The digital noticeboard project was first conceived in this context: on country, yarning and with local designs.

With consent from the Elders who make up the Land Council Executive membership, the project was put to the Australian Research Council and approved. Having received funding to work together to develop a series of digital noticeboards, one might expect the workshops and design activities to begin. However, differences in cultures are such that even contemplating holding an event or workshop is revealing of a large number of unknowns, so much so that it becomes intuitively clear that a workshop is not the place to start. Moreover in a collective culture it is important to make sure that everyone feels included and that Elders guide the project.

As a first step the non-Anindilyakwa members of the project undertook a cross cultural course and spent much time speaking with different members of the community, other stakeholders and other Aboriginal Australians. Bi-cultural people (Anindilyakwa who have spent time off island in Western communities) proved immensely important in understanding issues and being able to relate them across cultures.

The project has proceeded through dialogue with the Elders to explain our current thinking and to seek suggestions and advice. Dialogue and activities have also taken place with the Linguistics Centre staff, the Indigenous Land and Sea Rangers, the School principals and other groups on the Island. Activities have included ecological monitoring, fishing with the Rangers, and much fixing of computers and transferring photos between iPads. This has given insights into issues using everyday technology.

The goal of our project is to research, design, build and evaluate novel public communal technologies harmonised to the Warnindilyakwa. It is important to understand who will use the noticeboards, where they will be placed, what form they will take (noticeboards, tabletops or other forms that emphasize orality [2]), what kinds of interaction modalities they will support, protocols for publishing content, how they will be maintained etc.

The Warnindilyakwa culture is so different from Western culture and there seems to be so much potential for mistakes that it is difficult to know where to start. So we start from an existing point, working together, researchers with bilingual Warnindilyakwa members of the Land Council, beginning with Warnindilyakwa designs.

5 Warnindilyakwa Existing Design Practise

The Warnindilyakwa have invested considerable time and resources into designing and producing several artefacts whose principal purpose is for communication, both within the community and to the outside world. These include:

1. A poster presenting the Anindilyakwa Indigenous Protected Area (IPA) [11];
2. The land council newsletter "Ayakwa" [12]; and
3. Photos, videos and maps.

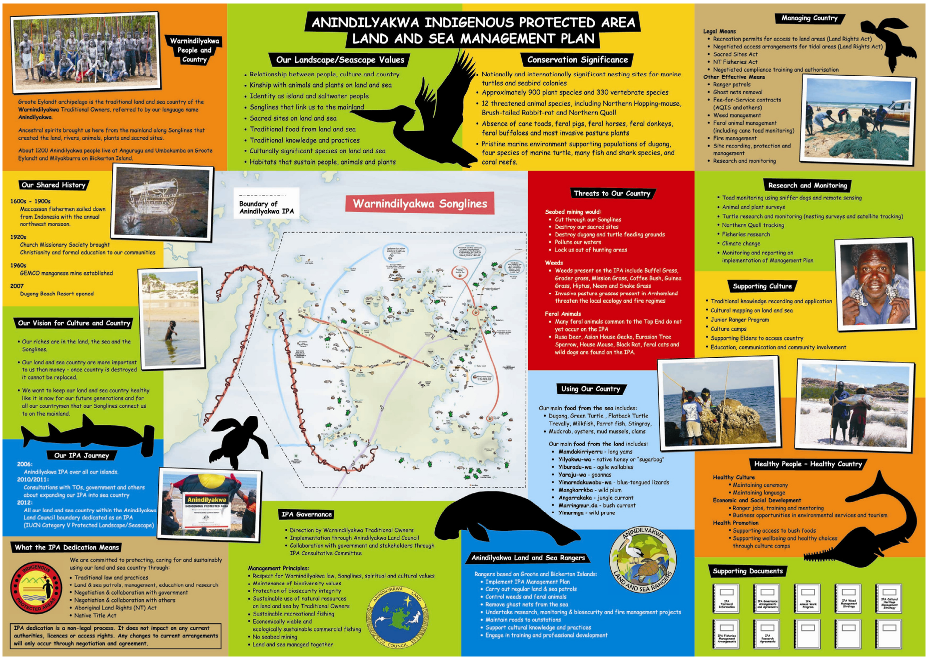


Fig. 1. Poster explaining the Groote Eylandt Indigenous Protected Area

The IPA poster (Fig. 1) was designed for communicating the importance of the Anindilyakwa Indigenous Protected Area agreement with the community and outside world. The IPA poster went through many iterations incorporating feedback from each of the different Warnindilyakwa clans. In particular the poster and its components required permission from Elders of each clan in order to be used and displayed.

The Ayakwa newsletter is a bi-monthly publication of the Anindilyakwa Land Council, see Fig. 2. The newsletter informs the community of news, event and important social messages. It is produced by the ALC and vetted by other members of the community before being published. Whilst being English text based the newsletter also contains stories "in language" and is highly visual.

Photos, videos and maps are used by the community, anthropologists, educators, health workers and rangers to discuss and present ideas. For example videos of ranger activities are shown to old people at aged care facilities both for their enjoyment but also to discuss issues, garner feedback and obtain permission. Videos have also been published for people remote from Groote to view on island activities [13].

6 First Steps Towards Cross Cultural Design of Digital Artefacts

Growing existing designs of the IPA poster, Ayakwa newsletter and maps and videos into digital and slightly interactive forms proved to be effective starting points for the design. We constructed digital versions of these which could be rendered on iPads (fairly common in the community), PCs and a large touch screen noticeboard which was shipped to the island. We stayed faithful to the original graphics and designs, with modifications to accommodate aspects such as screen size, interactive navigation and audio and video elements. The familiarity of the designs (IPA post, Ayakwa and ranger videos) enabled the community to immediately understand the potential of the interactive digital form, to begin generating further ideas and to discuss sensitivities.

Fig. 2. Ayakwa newsletter

Working with the familiar respects community, what the community has done and puts them in control because they have all of the knowledge that surrounds those things. It also helps non-community members to tap into community design ideas.

Initially the University researchers had hypothesized that sophisticated permissions systems might be necessary to control access to digital material and to obtain permission: systems which understood clans and community relationships. In fact existing straightforward protocols already exist: the Elders of all the clans, or those concerned, decide. This became immediately apparent through understanding how existing artefacts had been designed. Furthermore by starting with and growing existing artefacts the team was able to immediately trial novel technologies, e.g. digital noticeboards, seeded with designs including images and videos for which permission had already been obtained. (Were the designs to be used outside a trial e.g. for publication on the web permission would need to be requested again.)

It became clear for example from the initial demonstration of the digital noticeboard that the Elders were very pleased to be able to present positive stories from the Ayakwa magazine spoken "in language" on the digital noticeboard. At present these stories may not reach many in the community with low literacy. There was much discussion of subtitles and speaking in both Anindilyakwa and English languages, so as to foster literacy in English, and to further the correct spoken form of Anindilyakwa, reinforced by subtitles. The discussion revealed further concern about the way in which the design of some social media and texting practices encourages the use of creole, since they do not support oral forms, and because the written form of Anindilyakwa is quite long.

Showing a Google map of Groote Eylandt on the large touchscreen noticeboard led to proposals from the Elders for the possibility to show songlines, ecological names, clan estates and their boundaries and stories of the people, so that they can better impart and share their culture with the youth. It also revealed the quite different qualities between the Google map and a map created by an anthropologist, charting the names of the country according to the local people. The Elders were keen to enable these sorts of representations through technology.

From the conception of the project there have been different expectations and understandings of the project and different perspectives on what might or might not be acceptable regarding interfaces and publication. In a collective culture such as the Warnindilyakwa there are no leaders, the only consensus is on what Elders from all the clans might agree too. Some members of the community are keen to publish and promote culture to other parts of the world; others are more conservative.

The reaction of the non-Anindilyakwa was to avoid any cultural sensitivities and to initially target 'safe' aspects of the community life like the ranger program (environment) and sport. This proved both naïve and unnecessary: naïve in that culture pervades aspects of community life and unnecessary in that existing designs had already been undertaken and were ideal seeds from which to grow designs.

7 In Conclusion

Familiarity and expertise are key starting points that enable people to make design contributions. Building upon existing Indigenous designs and growing them into digital forms was an effective method of beginning a cross-cultural design project. It enabled the design team to engage the expertise and cultural knowledge of the local people and demonstrate technological possibilities in a concrete and culturally relevant way. In cross cultural design projects, approaches that respect cultural practices and learn from earlier local initiatives form an important stepping stone towards developing shared understandings and making greater innovations together.

Acknowledgements. First and foremost we wish to acknowledge the ALC, linguistics, rangers and Warnindilyakwa community for supporting this project. Elayne Fernandez, Vincent "Billy" Van Uitregt, Jason Wimmer, Joel Anderson and Ben Hall

all provided valuable input, as did QUT students Monique Alvis and Bethany Ward. This work was supported by an Australian Research Council Linkage Research project number LP120200329.

References

1. Beck, E.E.: P for Political: Participation is not enough. Scandinavian Journal of Information Systems 14(1), 1 (2002)
2. Bidwell, N.J., Reitmaier, T., Marsden, G., Hansen, S.: Designing with mobile digital storytelling in rural Africa. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1593–1602. ACM (April 2010)
3. Redhead, F., Brereton, M.: Growing local participation through long term design. In: Proceedings of the Participatory Innovation Conference 2012, Melbourne, pp. 1–4 (2012)
4. Brereton, M., Roe, P., Hong, A.: Evolving a relationship for cross-cultural participatory innovation. In: 2012 Participatory Innovation Conference Digital Proceedings, pp. 1–5. Swinburne University (2012)
5. Brereton, M., Buur, J.: New challenges for design participation in the era of ubiquitous computing. Co-Design 4(2), 101–113 (2008)
6. Puri, S.K., Byrne, E., Nhampossa, J.L., Quraishi, Z.B.: Contextuality of participation in IS design: a developing country perspective. In: Proceedings of the Eighth Conference on Participatory Design, vol. 1 (July 2004)
7. Ramsey, G., et al.: New Ownership New Responsibilities: Review of Education in the Groote Eylandt and Bickerton Island Region. Terteducon Pty Ltd. (July 2009)
8. Brasche, I.: Cultural resilience and social wellbeing: A case for research on Groote Eylandt. Australian Aboriginal Studies Journal (2), 93–98 (2008)
9. FAHCSIA, Department of Families, Housing, Community Services and Indigenous Affairs, Closing the Gap: The Indigenous Reform Agenda, Australian Government (2009), http://www.fahcsia.gov.au/sa/indigenous/progserv/ctg/Pages/default.aspx (access March 23, 2013)
10. Irani, L., Vertesi, J., Dourish, P., Philip, K., Grinter, R.E.: Postcolonial computing: a lens on design and development. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1311–1320. ACM (April 2010)
11. Indigenous Protected Areas, Australian Government, http://www.environment.gov.au/indigenous/ipa/index.html (accessed on March 23, 2013)
12. Ayakwa newsletter, Anindilyakwa Land Council, http://www.anindilyakwa.com.au/periodicals-and-annual-reports (accessed on March 23, 2013)
13. Ranger video, http://www.environment.gov.au/indigenous/publications/video-anindilyakwa.html (accessed on March 23, 2013)
14. Winschiers, H.: The Challenges of Participatory Design in a Intercultural Context: Designing for Usability in Namibia. In: PDC, pp. 73–76 (January 2006)
15. Winschiers-Theophilus, H., Chivuno-Kuria, S., Kapuire, G.K., Bidwell, N.J., Blake, E.: Being participated: a community approach. In: Proceedings of the 11th Biennial Participatory Design Conference, pp. 1–10. ACM (November 2010)
16. http://en.wikipedia.org/wiki/Songline

Web Accessibility in Africa: A Study of Three African Domains

Daniel Costa[1], Nádia Fernandes[1], Sofia Neves[1], Carlos Duarte[1], Raquel Hijón-Neira[2], and Luís Carriço[1]

[1] LaSIGE, University of Lisbon, Portugal
[2] Universidad Rey Juan Carlos, Spain
{dancosta,nadiaf,abatista,lmc,cad}@di.fc.ul.pt, raquelhijon@gmail.com

Abstract. Being the most used method for dissemination of information, especially for public services, it is of paramount importance that the Web is made accessible as to allow all its users to access the content of its pages.

In this paper, we evaluated 2250 Governmental Web pages from each one of three different African countries (i.e., Angola, Mozambique and South Africa). This report compares the accessibility quality and the level of structural complexity of these African countries government's Web pages. We found that hand coded pages tend to have larger number of HTML elements and also to present higher number of accessibility problems. Finally, it suggests some recommendations to repair the most common problems in these pages.

Keywords: Web Science, Web accessibility, automated evaluation.

1 Introduction

In many countries, the Web is the main vehicle used by governments to spread information, education, allow civic participation and other public services. It also is an important medium for receiving and providing information and interacting with society. Therefore, it is essential that the Web is accessible in order to provide equal access and equal opportunity to people with or without disabilities. Besides, an accessible Web has the potential to help people with disabilities and the elderly to participate more actively in society.

The United Nations (UN) estimates that approximately 10% of the world's population are persons with disabilities [2]. It is difficult to estimate how many people are affected by Web accessibility problems, nevertheless, if we move forward to an ideal situation, where only a reduced percentage of the population faces accessibility barriers, then technology is serving society in the right way.

The importance of Web accessibility is increasing in the international context, and especially in the European Union [1]. In Europe, more and more countries have legislation requiring that government Web sites be accessible. In contrast, developing countries in Africa have less stringent laws, if any [2]. Governments worldwide have

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 331–338, 2013.

several stimuli to adopt accessibility. Demonstration of social responsibility by provisioning information and services to all citizens is one of them.

In this paper, we present a report of the state of Web Accessibility in three countries located in the African continent. The evaluation of accessibility we describe is based on the Web Accessibility Guidelines (WCAG) 2.0 [3].

1.1 Web Content Accessibility Guidelines 2.0

To help creating accessible Web pages, WCAG 2.0 defines guidelines that encourage designers/developers to craft Web pages according to a set of best practices. These guidelines are also used for accessibility evaluation.

WCAG 2.0 contains several guidelines written as testable sentences and chosen to address specific problems related with accessibility. Each guideline has a testable success criterion, which is supported by techniques that can be true or false when testing Web content against them.

Although, it is possible to use the guidelines to manually evaluate Web pages, due to the nature of this study (i.e., the large number of Web pages evaluated) we used an automated evaluation tool: QualWeb [4].

1.2 QualWeb

QualWeb is a Web automatic accessibility evaluation tool. The main advantage of this tool is the in browser context evaluation [6], i.e., after the Web browser processes the Web page and all resources are loaded. To this end, the Webkit-based Phantom[1] headless browser is used, allowing us to assess the page's code after browser processing. In terms of techniques, QualWeb covers 51% of the HTML and 73% of the CSS techniques.

An additional distinguishing feature of this tool is the ability to find different states of the Web page [4]. This means QualWeb is capable of interacting with DOM elements and detecting changes to the DOM of a page. QualWeb stores a new state if more than content is replaced after interaction (e.g., introduction of new HTML elements on the DOM tree). We consider the total number of states found, the level of complexity of a Web page as this reflects the dynamism we can find on the current state of the Web.

2 Experimental Study

For this study, the first step was to obtain a list of governmental Web pages for each of the three countries: Angola, Mozambique and South Africa. Starting from each of the main government's pages, we used a Crawler to look for clickable elements in it. Every time a clickable element redirected to another URL on the same domain name (gov.ao for Angola; gov.mz for Mozambique; or gov.za for South Africa), this new URL was kept as an object to be evaluated and the algorithm continued to execute. Using this method, we collected a sample of 2250 government Web pages, from each country.

[1] PhantomJS: `http://phantomjs.org/`

Afterwards we performed the evaluation itself, on each one of these 2250 Web pages per country. Every Web page was assessed with the QualWeb evaluator to check for conformance with WCAG 2.0 HTML and CSS techniques. The evaluation produced a list of Warnings, Passes and Fails that are analysed in the results section. In the interest of classifying the complexity of the evaluated Web pages, the QualWeb feature allowing the identification of different page states was used to determine the total number of states in the pages evaluated.

2.1 Results

Our evaluation yielded differences in the HTML documents in terms of number of HTML elements, between domains of different countries (Figure 1). The pages of South Africa (za) present a higher number of elements with an average of 846.37 elements per page, followed by pages of Angola (ao) with an average of 360.17 elements per page and finally by the pages of Mozambique (mz) with an average of 344.60 elements per page.

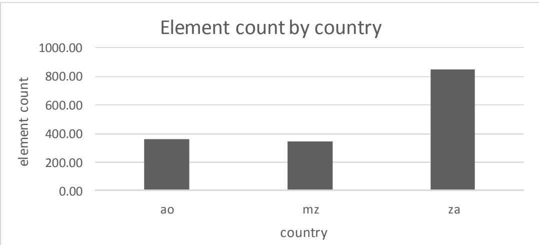


Fig. 1. Average number of elements per page for each country's governmental pages

Figure 2 presents how the evaluation outcomes (fail, pass and warning) differ between the African countries' Web pages. A *failure* occurs in the cases where the evaluator can detect automatically and unambiguously if a given HTML element has an accessibility problem. A *pass* ensues from elements that, unambiguously, are classified as having no accessibility problems. *Warnings* are raised when the evaluator can partially detect accessibility problems, but which might require additional inspection (often by experts). Table 1 presents the percentage of outcomes (pass, fail and warning) by country. Inspecting these results with additional detail, the Web pages have the following evaluation outcomes:

- **Fail**: Even though the compliance with accessibility techniques is quite different in all three countries, the common factor between the Web pages of Mozambique and South Africa is that fails are slightly above 50%. In addition, the Angolan Web pages are just above 40% for fails.
- **Pass**: Angola's governmental Web pages register the highest percentage of passing elements, reaching over 40%. Mozambique ratio decreases to around 37% and South Africa registers the lowest value, around 19%.

- **Warning**: Mozambique's Web pages elements register the lower percentage of warnings, around 10%. Followed by Angola's Web pages with 13% and South Africa with 27%. The three countries have total of fail and warning above 50%: Mozambique just above 60%; Angola around 55%; and South Africa approximately 80%. South Africa registers the highest total of potential accessibility problems.

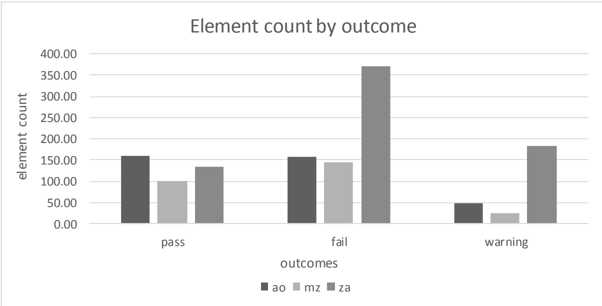


Fig. 2. Average number of HTML elements by evaluation outcome by country

Table 1. Distribution of evaluation outcomes (absolute values and percentages) by country

Country	Pass	% Pass	Fail	% Fail	Warning	% Warning
Angola	159.66	**43.61%**	157.92	**43.14%**	48.51	**13.25%**
Mozambique	101.45	**37.42%**	143.77	**53.03%**	25.88	**9.55%**
South Africa	133.47	**19.46%**	370.05	**53.94%**	182.45	**26.60%**

Evaluation by Technique

In the following analysis, we will focus on the accessibility results by technique, identifying the more compliant and the more infringed techniques for each country. Figure 3 shows the techniques where occurred passes and their average. All three countries present higher pass values for techniques C23 and C19. The third higher pass value is C8 for South Africa, C9 for Mozambique and C21 for Angola. These techniques evaluate the following conditions:

- C23 – if div elements in main content have background colour;
- C19 – whether text is incorrectly altered to "look" as if it has an align right or centre;
- C8 – for paragraphs and headings, looks for a wrong usage of extra spaces between letters to simulate the letter spacing property;
- C9 – whereas decorative images are specified in CSS rules and therefore removable when disabling CSS;
- C21 – checks if the line-height property is used with relative values and if these values range between the ones recommended.

The first observation that can be made is that HTML techniques present lower values of pass comparatively with CSS techniques. This can be explained by the fact that CSS techniques are more specific than HTML ones, which means that an automated evaluation can more easily determine pass for these, while HTML return higher number of warnings.

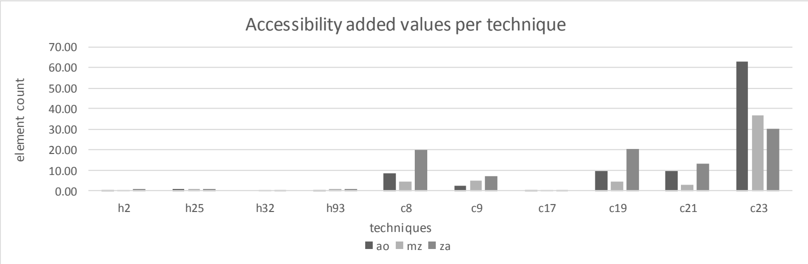


Fig. 3. Average number of passes by technique per country

The average number of possible problems and problems (fails and warnings) per technique is presented in Figure 4. All three countries present higher values in techniques C15 and C7. For South Africa and Angola, the subsequent high value technique is H30, while for Mozambique is H73. Respectively these techniques evaluate the following conditions:

- C15 – if anchor and input form components present a visual alteration when interacted with;
- C7 – whether anchor elements are followed by a span tag with a textual description of the link hidden by a CSS rule;
- H30 – if the link text describes the purpose of the anchor;
- H73 – checks the correct usage of the summary attribute in tables.

From these results, we can deduce the most common elements with potential accessibility problems. In South Africa and Angola these are anchors or input form components, and in Mozambique tables are added to these.

Incompliance with certain techniques is more pronounced in some countries. For instance:

- H33 – if a title attribute supplements a link, is a more common problem in South Africa (average of 24.61), comparing with the other countries (average of 2.95 for Angola and 1.50 for Mozambique);
- C23 – which presents an average of 9 elements with problems for Angola, being negligible in the other two countries;
- H39 – verifies the usage of caption elements to associate data tables captions with data tables, shows the same behaviour as H33, with an average of 39.58 for South Africa (average of 1.10 for Angola and 11.42 for Mozambique).

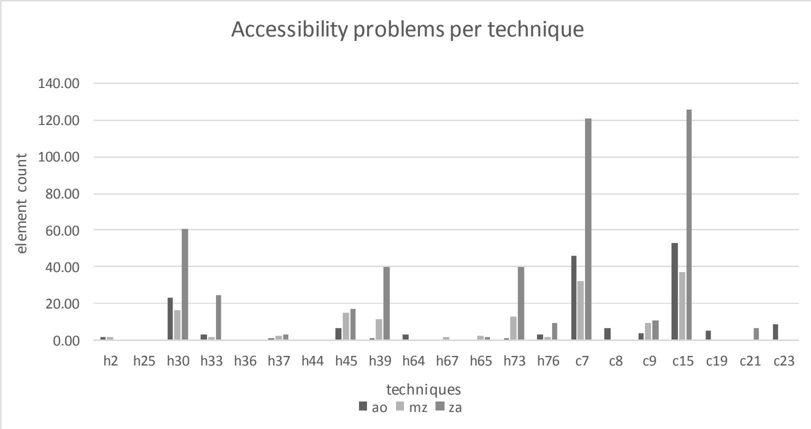


Fig. 4. Average number of fails and warnings (possible problems) by affected technique

Level of Complexity

We found that the average complexity for the three different domains is approximately 1. The results gathered for Mozambique and South Africa show that the highest level of complexity is 2 (found in 2 Web pages). For the Angolan Government pages, the highest level is 3 (found in 3 Web pages), while 17 pages had level 2.

From these results we can conclude that, for these countries, dynamic changes to the governmental pages layout or interaction elements (thus excluding changes to their content) is not common. When these changes are required, a new page will be loaded, instead of changing the DOM.

3 Discussion

We found there are differences between the three African domains government's Web pages accessibility quality. The South African Government's Web pages have a larger number of HTML elements, but also present a larger percentage of elements raising fails and warnings, comparatively with the Web pages from Angola and Mozambique. This goes towards the conclusions of *Lopes et al* [5], where it was found that the size of the pages influences its quality (i.e., smaller Web pages have less accessibility problems than bigger ones).

Concerning the techniques, it was observed that CSS techniques have a greater influence on the positive accessibility values for all the countries domains than HTML techniques. Techniques C23, C21, C19, and C8 were found to be the ones with highest compliance levels.

When considering potential accessibility problems (fails and warnings), we perceived that they also have higher values in CSS techniques but the difference to HTML techniques is not as pronounced as we found when analysing passes.

The techniques most often violated were C7, C15, H30 and H73. It is interesting to note that what we observed for one of the techniques with more problems, H30 (which verifies if the link text describes the purpose of the anchor), is consistent with what was already seen in a previous accessibility study of two hundred of the most used Web pages in the entire world [6].

The majority of the HTML problems found are related with the accessibility quality of tables, specifically when they do not have captions and summary elements, and if links do not have text descriptions. If those were carefully reviewed and redone the accessibility quality of the pages would considerably improve.

The results show that government Web pages would greatly benefit from reviewing their CSS, since the majority of their problems are located in techniques C7 and C15, especially for the South African government's Web pages. Problems with these techniques can be solved by adding a description of the link given in the anchor element, inside a span tag and hidden by a span CSS, as recommended by the WCAG 2.0 description. For technique C15, the solution would be to ensure that every anchor link and input box changes its colour whenever it is interacted with. People would greatly benefit from this visual aid and contrary to technique C7, it is much easier to enforce. Correcting these situations would help separate the normal paragraph's text and the interactive text in the anchor element, as well as help signalling which form input element is selected at a specific instant when it is being interacted with.

After finishing the automated evaluation, we performed a manual inspection of some of the government's pages from each country. This inspection was performed following the indications of the WCAG 2.0. For the South Africa's Web pages we observed that: the limitations of the several divisions of the pages was not always clear; link elements were confused with parts of the text; the general structure was quite similar to a newspaper and did not denote a lot of accessibility concerns. For Mozambique's Web pages, decorative images do not have either alt or title attributes when they should have them with empty values; some colours are also perceived as being too bright; table captions were also almost inexistent; there are also some flash objects directly embedded without any textual descriptions. For Angolan Web pages, since they generally follow the same structure, they all could benefit from adding captions to tables and textual descriptions to images and anchor elements. We can see that some of the issues found manually confirm the findings of the automated evaluation.

It was also possible to detect that Angola and Mozambique's Web pages benefited from tools that help code generation (such as Flyout and Plone, respectively). On the other hand, South African pages, taking into account the quantity of comments in the code and its specificity, were probably manually coded. This probably contributed to the bigger number of CSS problems, because code generators avoid several CSS problems, such as the use of relative font-sizes.

Regarding the level of complexity of the Web pages, we found that dynamic changes to the pages' DOM are mainly used to change the content of the pages and not to add new elements to the page (i.e., less structural complexity). In what concerns the accessibility quality, the slightly higher complexity found in Angolan Web pages does not reflect any significant change in the overall accessibility score.

4 Conclusion

The Web is the main vehicle used by many governments to spread information, education, allow civic participation and other public services. If these pages are not accessible they fail to reach their target population.

In this paper we evaluated 2250 Governmental Web pages from each one of three different African countries: Angola, Mozambique and South Africa. This report shows that the South Africa Government Web pages have more elements than the other countries but have less quality in terms of accessibility. The Angolan Government Web pages scores the best ratio of passes when comparing with the other countries. Mozambique's pages have the lower rating of fails and warnings combined. Regarding the level of structural complexity, we did not find major differences between the different countries' Web pages.

A manual inspection of a sample of the pages suggested that Angolan and Mozambican Web pages might have benefited from the support of code generation tools during their development, while this is not so clear in South African Web pages. The accessibility evaluation, concomitantly, has shown more accessibility problems in South African pages, with some of these problems being in some cases more easily addressed and prevented with the use of code generation tools.

This overall view of the current state of accessibility in these African governments Web pages by WCAG 2.0 techniques facilitates establishing a set of recommendations to repair the most common problems.

References

1. eAccessibility – Opening up the Information Society. European Commission (December 2010), http://ec.europa.eu/information_society/activities/einclusion/policy/accessibility/index_en.htm/ (last accessed on March 20, 2013)
2. Kuzma, J., Yen, D., Oestreicher, K.: Global e-government Web Accessibility: An Empirical Examination of EU, Asian and African Sites. In: Second International Conference on Information and Communication Technologies and Accessibility, Hammamet, Tunisia, pp. 83–90 (2009)
3. Cooper, M., Reid, L.G., Vanderheiden, G., Caldwell, B.: Techniques for WCAG 2.0 – Techniques and Failures for Web Content Accessibility Guidelines 2.0. W3C Note, World Wide Web Consortium (W3C) (October 2010), http://www.w3.org/TR/WCAG-TECHS/
4. Fernandes, N., Costa, D., Neves, S., Duarte, C., Carriço, L.: Evaluating the accessibility of rich internet applications. In: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A 2012), Article 13, 4 pages. ACM, New York (2012), doi:10.1145/2207016.2207019
5. Lopes, R., Carriço, L.: Macroscopic characterisations of Web accessibility. New Rev. Hypermedia Multimedia 16(3), 221–243 (2010)
6. Fernandes, N., Lopes, R., Carriço, L.: Evaluating Web Accessibility at different processing phases. New Review of Hypermedia and Multimedia (2012)

Webpage Designs for Diverse Cultures: An Exploratory Study of User Preferences in China

Yin Su[1], David Liu[2], Xiaomeng Yuan[2], Justin Ting[2], Jingguo Jiang[3], Li Wang[2], and Lin Gao[2]

[1] Lenovo (Beijing) Co., Ltd, 6 Shangdi West Road, Haidian District, Beijing, China
suyin86@126.com
[2] Microsoft Co., Ltd, 5 Danleng Road, Haidian District, Beijing, China
{davli,xiyuan,justtin,applewang,lingao}@microsoft.com
[3] Zhejiang University Hangzhou, China
firjjg@163.com

Abstract. A wealth of studies has revealed a cross-cultural difference in the user preference on webpage designs. Users from other cultures often criticize a widely accepted webpage design in one culture. Designs for diverse cultures are thus expected to be specific to address diverse user preferences. This study investigated the preferences of Chinese users on four essential design elements related to the readability of texts of the result pages of search engines. The results suggested that the search result pages of the Bing search engine designed for typical 'US users' did not satisfy Chinese users. Chinese users, in general, preferred huge-sized texts for titles, a more compact layout of the search result pages, and keywords to be highlighted in red. The findings of the study contributed to webpage design guidelines for Chinese users, and may serve as a catalyst in exploring user preferences in designing for diverse cultures.

Keywords: webpage design, cross-culture, diversity, Chinese users.

1 Introduction

As the number of Internet users in China increases (up from 11,100 (8.5% of the population of China) to 56,400 (42.1%) in the last 7 years), some international companies that offer online services have realized the potentially huge business opportunities and rushed into the Chinese market. Designs of websites of these companies usually follow the ones that have been accepted in the respective homeland of the company, with language the only difference. Chinese users are sometimes not satisfied with a mere translation of the content. The question may be asked as to why a websites with good "pedigree" may not be successful in China.

1.1 Cross-Cultural Difference in Preferences of Webpage Design

A wealth of studies has revealed a cross-cultural difference in user preferences and judgments on the webpage design [1-6]. For instance, Simon [6] found that Asians

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 339–346, 2013.

disliked triangles and squares on webpages, whereas North Americans and Europeans preferred combinations of those shapes. Faiola, Ho, Tarrant, and MacDorman [7] suggested that the U.S. and South Korean people perceived the aesthetics of home pages differently. These findings suggest that designs that are blindly copied or slightly modified from other cultures may actually not be satisfying.

Attempts have been made to provide explanations for the cross-cultural difference in preferences on the webpage design. Many studies suggested that the cognition and communication styles in diverse cultures influence how people learn and interact with online information [6, 8-12]. This learning process, in turn, influences how users interpret a website's aesthetics [13-15]. For instance, in low-context cultures (e.g., Germans and Swiss), communication occurs predominantly through explicit statements in text and speech. In high-context cultures (e.g., Japanese and Chinese), messages include other communicative cues such as body language and the use of silence [16]. Previous studies showed that the high-context people browse information faster and prefer fewer links to find information than the low-context users [15]. The high-context users appreciate the webpages with a compact layout more, while the low-context ones prefer the pages that looked more open. These studies provide theoretical evidence for the cross-cultural differences in the judgments and preferences of webpage designs.

1.2 Motivations for the Present Study

Considering the cross-cultural differences in the judgment and preference of webpage designs, it may be advantages for the international companies to provide specific designs to cater to the preferences and requirements of the users from diverse cultures. However, there are only a few theoretical or empirical studies that provide detailed guidelines for designers. Accordingly, some webpages, which were specifically designed for Chinese users, were found not to be satisfying due to the limitation of the degree to which the designers understood the preferences of Chinese users.

Our study was performed to investigate the specific preferences of Chinese users, aiming to make attempts to enrich the guidelines of the webpage design for Chinese users and to act as a catalyst for future studies that explore the specific preferences of users from diverse cultures on the designs (i.e., not limited to the webpage design alone).

1.3 Design Elements Tested in the Present Study

Numerous studies have been conducted to explore impact factors of readability of texts on computer screens, indicating that many essential design elements do have an effect on the readability of computer-displayed texts [21-26]. For instance, the font-size and font-color were indicated as having effects on both of the readability of English and Chinese texts, which reflected in accuracy and reading speed [21-22]. Chan and Lee [24] suggested that the line-spacing also influence an individual's reading speed, as well as the comprehension of offered information. A similar finding regarding to the impact of line-spacing on readability indicated that texts with wider line spacing lead to better accuracy and faster reaction times [26]. Researchers also found the line-length of text on websites to influence its clarity and comprehension [25].

Based on these observations, our study examined whether and how four design elements that have been suggested to relate to the readability of texts (i.e., the font size, style of keyword-highlighting, line height, and search result margin (SR-margin)), influenced Chinese user preferences on the webpage design. Specifically, a search result was composed of a title, an attribution, and a snippet. The SR-margin referred to the space between search results, while the line-height was a combination of the space between the title and attribution, the space between the attribution and the snippet, and the space between lines within the snippet. The terms of the line height, the SR-margin, and the title, attribution, and the snippet of search results are illustrated in Fig. 1.

Fig. 1. An illustration of the terms of the line height, the SR-margin, and the title, attribution, and the snippet of search results

2 Methods

A total of 1009 participants (roughly 50% female; 90% white collared workers and 10% students; 59% aged from 25 to 30, 12% aged from 18 to 24, and 29% aged from 31 to 35) were recruited by iResearch, a professional consulting company (www.iresearch.com), from cities of Beijing, Shanghai, and Guangzhou. All participants randomly completed no more than 3 prepared tasks. Participants in each task ranged in number from 318 to 361.

Four kinds of font-size, six kinds of keyword-highlight, six kinds of line-height, and six kinds of SR-margin were separately tested using four tasks. Detailed parameters of all designs of these elements are summarized in Table 1, including the parameters of the designs of the Bing search engine (www.bing.com) for typical 'US users' (the so-called US designs). A set of search result pages (SERPs) that contained exactly the same contents with the only difference in only one of the four design elements was prepared in each task.

Table 1. The detailed parameters of all designs of the four elements tested in this study

	Font-size (T/A&S)	Keyword-highlight (T/S)	Line-height (T/A/S)	SR-margin
A	13px/13px	Red/Red	1px/1px/16px	17px
B	16px/12px	**Red/Red**	3px/1px/18px	19px
C	16px/13px	Blue/Blue	4px/1px/19px	21px
D	20px/13px	**Blue/Blue**	5px/3px/21px	23px
E	-	**Orange/Orange**	7px/5px/23px	25px
F	-	Red/N/A	9px/7px/25px	27px

Note. T, A, S denoted the title, attribution, and snippet. Underlined parameters indicate the design of the Bing search engine for typical 'US users'.

In each task, participants were side-by-side presented with two random SERPs from the SERP set and were required to indicate their preferences. After that, the "loser" SERP disappeared and a new "competitor" from the SERP set showed up. The participants were required to indicate their preferences again. Then another new "competitor" replaced the "loser", and so on. After all SERPs in the set have shown up, the final winner was recorded as the page with the most preferred design.

3 Results

We separately calculated the percentages of total participants that preferred each kind of design for each of the four design elements. We aimed to find out which kind(s) of design(s) were most preferred by participants. Further, with respect to each design element, we separately analyzed whether and how participant preferences were influenced by the individual gender, age, and occupation. We did not analyze the interactions between gender, age, and occupation on participant preference due to the limitation of the sample size.

3.1 Font-Size

Four kinds of font-size combinations, as listed in Table 1, were tested in this study. A chi-square test revealed a significant font-size effect on the percentage of participants that preferred each kind of combination most ($\chi^2(3) = 29.70$, $p < .01$). The popularity of these designs increased with the increasing size of the font of titles. That is, overall, a significant majority of participants (34.77%) preferred the design with the largest font-size (20px) of titles most (ps < .001), while the design with the smallest font-size (13px) of titles was the least preferred one (13.54%). The results indicated that the percentage of Chinese users that most preferred the font-size combination (16px/13px) in the so-called US design (25.51%) was significantly less than the percentage of users that preferred the 'largest one' (20px/13px) ($p < .01$). Further analyses revealed no significant effects of gender ($\chi^2(3) = .52$, n.s.), age ($\chi^2(6) = 7.04$, n.s.) and occupation ($\chi^2(3) = 2.34$, n.s.) on the distribution of participants that favored each kind of font-size.

3.2 Keyword-Highlight

Six kinds of keyword-highlight design were prepared for testing in this study. A chi-square test revealed a significant difference between individual preferences on SERPs with these designs ($\chi^2(5) = 52.23$, $p < .01$). Most participants preferred either the red-normal design (25.79%) or the red-bold design (22.64%) of keyword-highlight most. No significant difference was observed in the individual preference between these two kinds of design, n.s. Participants most dislike the blue-normal and the blue-bold designs. Specifically, only 5.66% of participants indicated that they preferred the blue-bold design most, which serves as the so-called US design of the Bing search engine. The percentages of participants that favored the orange-normal design (19.50%) and that favored the red-N/A design (19.18%) were in the middle of these two extremes. Further, neither individual gender ($\chi^2(5) = 3.61$, n.s.), age ($\chi^2(5) = 14.44$, n.s.), nor occupation ($\chi^2(5) = 3.24$, n.s.) was observed to have a significant impact on the distribution of participant preferences on kinds of keyword-height design.

3.3 Line-Height and SR-Margin

We compared the participant preferences on six kinds of line-height combinations. A chi-square test revealed a significant difference between individual preferences on SERPs with these combinations of line-height ($\chi^2(5) = 14.68$, $p < .01$). The combination preferred by the most participants (21.88%) was as follows: 3px wide between the title and the attribution, 1px wide between the attribution and snippet, and 18px wide between lines in snippet. The popularity of the combination roughly increases as the layout of search results was designed to be more compact. In particular, only 16.41% of participants preferred the line-height combination of 5px wide between the title and the attribution, 3px wide between the attribution and snippet, and 21px wide between lines in snippet, which serves as the so-called US design of the Bing search engine ($p < .05$). Further chi-square tests revealed that there were no gender ($\chi^2(5) = 7.52$, n.s.), age ($\chi^2(10) = 8.22$, n.s.), and occupation ($\chi^2(5) = 7.82$, n.s.) effects on participant preferences on each kind of the line-height combination.

We prepared six kinds of SR-margin, which ranged from 17px to 27px with intervals of 2px for testing in this study. A chi-square test revealed that there was a significant difference in participant preferences on the six kinds of the SR-margin design ($\chi^2(5) = 15.57$, $p < .01$). Take a closer look at the preference data, we found that a significant majority of participants (22.56%) favored SERPs with 19px SR-margin, ps < .01. Specifically, among the four SERPs with the SR-margin that was wider than 19px, the popularity roughly decreased with the increasing of the width of the SR-margin. A total of 19.51% of participants preferred the design for typical 'US users' with 21px SR-margin, which was significantly less than the percentage of participants that preferred the design with 19px ($p < .01$). No significant effects of individual gender ($\chi^2(5) = 7.99$, n.s.), age ($\chi^2(10) = 7.09$, n.s.), and occupation ($\chi^2(5) = 4.68$, n.s.) were observed on the distribution of the participant preferences on SR-margin designs. These findings were generally consistent with those of the line-height design, suggesting that participants more appreciate SERPs with a relatively compact layout.

4 General Discussion

Our study tested Chinese user preferences on four essential design elements of the search result pages. Overall, the results revealed that most of the search result pages that were designed for 'US users' actually did not satisfy Chinese users. The four main findings are summarized below.

First, Chinese users consistently preferred the huge-sized texts for titles to the medium-sized ones used in the designs for typical 'US users'. Previous studies have suggested that enlarging the font-size can improve the readability of texts [17, 23-24]. When viewing search results, Chinese users were used to fixating on the titles of search results and almost ignored the attributions and snippets. Chinese users mainly focus on the good readability of search result titles, but pay less attention to the aesthetics of the whole pages.

Second, Chinese users consistently preferred keywords to be highlighted in red. We reasoned that Chinese users might be used to first fixating the texts around keywords when judging the value of a search result. They desired one way of highlighting that could help them distinguish keywords among texts immediately. Therefore, the reason why Chinese users preferred keywords to be highlighted in red is probably because the keywords in a more distinguishing color were more legible than those in bold.

Third, Chinese users consistently favored a relatively compact layout of the search result pages. For the pages that were designed on the basis of the favorite combination of SR-margin and line-height of Chinese users, the amount of information displayed on one screen was approximately twice as much as that offered by the pages that were designed for 'US users". We proposed that Chinese users were sometime prone also to view some of other information that seemed to be irrelevant to their original target search. For example, consumers with the aim of purchasing clothes would also like to view promotion information concerning other goods. Therefore, they would not reject pages containing a great deal of information.

These findings indicate that when designing webpages for Chinese users, the following general guidelines should be followed: (1) try to use a relatively large font; (2) highlight the keywords in red; and (3) make the layout of the webpage compact with a relatively great deal of information.

This study added to the wealth of evidence that users from different cultures can perceive webpages differently and often prefer different designs [1-6, 25-26]. We suggested that designers should understand the preferences of users from diverse cultures and provide specific designs to address their preferences. The findings of this study may assist to enrich the guidelines of the webpage design for Chinese users, and are expected to act as a catalyst for future studies that explore the specific preferences of users from diverse cultures on the designs.

Two limitations of this study should be acknowledged. Firstly, the experimental design should be more rigorous. We examined the main effects of the four design elements on user preferences, but we did not take the interactions between these elements into account due to the limitation of the sample size. Secondly, using the paradigm of the current study, we can measure the user preferences on the design of webpages, but we cannot determine the exact reasons behind their judgments and the actual influences of the webpage design on user information search and processing. Considering both aspects, we suggest that a larger participant pool, a more rigorous experimental design and various methodologies to measure individuals' unconscious information search and processing (e.g., think aloud and eye-tracking) to be used in future studies.

References

1. Badre, A.N.: The effects of cross cultural interface design orientation on World Wide Web user performance (2000), `http://www.cc.gatech.edu/gvu/reports/2001/abstracts/01-03.html` (retrieved)
2. Chau, P.Y.K., Cole, M., Massey, A.P., Montoya-Weiss, M., O'Keefe, R.M.: Cultural differences in the online behavior of consumers. Communications of the ACM 45, 138–143 (2002)
3. Faiola, A., Matei, S.A.: Cultural cognitive style and web design: Beyond a behavioral inquiry into computer-mediated communication. J. Comput. Med. Commun. 11, 375–394 (2005)
4. Kamppuri, M., Bednarik, R., Tukiainen, M.: The expanding focus of HCI: Case culture. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 405–408. ACM Press, New York (2006)
5. Shen, S., Woolley, M., Prior, S.: Towards culture-centred design. Interact. Comput. 18, 820–852 (2006)
6. Simon, S.J.: A cross cultural analysis of Web site design: An empirical study of global Web users (1999), `http://marketing.byu.edu/htmlpages/ccrs/proceedings99/simon.htm` (retrieved)
7. Faiola, A., Ho, C.-C., Tarrant, M., MacDorman, K.: The Aesthetic Dimensions of U.S. and South Korean Responses to Web Home Pages: A Cross-Cultural Comparison. Int. J. Hum. Comput. Interact. 27, 131–150 (2011)
8. Marcus, A., Gould, E.W.: Crosscurrents cultural dimensions and global web user interface design. Interactions 2, 32–46 (2000)
9. Nisbett, R.E.: The geography of thought. The Free, NewYork (2003)
10. Norman, D.: Emotional design: Why we love (or hate) everyday things. Basic Books, New York (2004)

11. Picard, R.W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Strohecker, C.: Affective learning: A manifesto. BT Technology Journal 22, 253–269 (2004)
12. Tractinsky, N.: Towards the study of aesthetics in information technology. In: Proceedings of the 25th Annual International Conference on Information Systems, pp. 771–780. ACM Press, New York (2004)
13. Karvonen, K.: The beauty of simplicity. In: Proceedings of the ACM Conference on Univeral Usability, pp. 85–90. ACM Press, New York (2000)
14. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of Web sites. Int. J. Hum. Comput. Interact. St. 60, 269–298 (2004)
15. Plocher, T., Patrick Rau, P.-L., Choong, Y.-Y.: Cross-cultural design. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 4th edn. John Wiley & Sons, Inc., Hoboken (2012)
16. Würtz, E.: Intercultural communication on web sites: A cross-cultural analysis of web sites from high-context cultures and low-context cultures. J. Comput Med. Commun. 11, 274–299 (2005)
17. Huang, D.L., Rau, P.L., Liu, Y.: Effects of font size, display resolution and task type on reading Chinese fonts from mobile devices. Int. J. Ind. 39, 81–89 (2009)
18. Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G.: Examining children's reading performance and preference for different computer-displayed text. Behav. Inform. Technol. 21, 87–96 (2002)
19. Boyarski, D., Neuwirth, C., Forlizzi, J., Regli, S.H.: A study of fonts designed for screen display. In: CHI 1998 Conference Proceedings, pp. 87–94 (1998)
20. Tullis, T.S., Boynton, J.L., Hersh, H.: Readability of fonts in the windows environment. In: Conference Companion on Human Factors in Computing Systems, Denver, Colorado, United States, pp. 127–128 (1995)
21. Garcia, M.L., Caldera, C.I.: The effect of color and typeface on the readability of on-line text. Comput. Ind. Eng. 31, 519–524 (1996)
22. Dyson, M.C., Kipping, G.J.: The effects of line length and method of movement on patterns of reading from screens. Vis. Lang. 32, 150–181 (1998)
23. Bernard, M.L., Chaparro, B.S., Mills, M.M., Halcomb, C.G.: Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. Int. J. Hum. Comput. St. 59, 823–835 (2003)
24. Chan, A., Lee, P.: Effect of display factors on Chinese reading times, comprehension scores and preferences. Behav. Inform. Technol. 24, 81–91 (2005)
25. Zahedi, F., Van Pelt, W., Song, J.: A conceptual framework for international web design. IEEE Transactions on Professional Communication 44, 83–103 (2001)
26. Ling, J., van Schaik, P.: The influence of line spacing and text alignment on visual search of web pages. Displays 28, 60–67 (2007)

Your Phone Has Internet - Why Are You at a Library PC? Re-imagining Public Access in the Mobile Internet Era

Jonathan Donner[1] and Marion Walton[2]

[1] Microsoft Research India, Lavelle Rd, Bangalore, India 500080
[2] The Centre for Film and Media Studies, University of Cape Town, Upper Campus, Arts Block, Cape Town
jdonner@microsoft.com, marion.walton@uct.ac.za

Abstract. This study focuses on teenage users of public internet access venues (PAVs) in low-income neighborhoods of Cape Town. It documents their cultivation of detailed ICT repertoires to make the most of available ICTs. It highlights the continuing importance of PAVs as supplements for poorly equipped schools and reveals the incompleteness of any supposed transition to mobile-only internet use. While the mobile internet is opening up opportunities for young people, its current form still conflicts with the easy (global) rhetoric of a closing digital divide and the end of the PAV. We recommend policy and design actions (effecting rules, training, messaging, functionality, and Wi-Fi) to reconfigure PAVs to be more useful "in the age of the mobile internet". Though some actions require support from policymakers, this is fruitful ground for designers and technologists. We identify steps that can be undertaken immediately, rather than waiting for future device convergence or lower tariffs.

Keywords: Libraries, ICT4D, Shared Access, South Africa, Developing Regions, Human Factors, Mobile Phones.

1 Introduction

This paper revolves around a question: “If you have the internet in your pocket, why do you still visit a public access venue?” Mobile Internet access is growing worldwide, propelled by lower cost smart phones and data enabled “feature phones” costing as little as USD$50. Though mobile handsets may not provide the same functionality as a PC, the mobile internet promises to bring a billion or more people online [1].

The arrival of this private, accessible, but perhaps not optimal mobile internet has implications for ‘traditional’ public access venues (PAVs), such as libraries, telecenters and cybercafés serving low-resource communities. We suggest that recent assertions in the practitioner literature (e.g., [2, 3]) about the irrelevance of public access in the age of the mobile deserve further scrutiny. Only a handful of studies have systematically addressed the interplay amongst these forms of access [4–6], and no one has examined the effect of mobile internet use on particular groups engaging in specific activities or the strategic choices made by users confronted with a potential repertoire [7, 8] of access choices. This paper describes a study in Cape Town, South Africa,

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 347–364, 2013.

emphasizing an ongoing role for public access venues even among some mobile internet users, and suggesting ways to redesign and reconfigure public access venues to be more useful in the era of the mobile internet.

This study focuses on older teenage PAV users in low-income neighborhoods of Cape Town, identifying the roles of PAV and mobile phones in their educational, cultural, civic, and health-related involvement. The consequences of PAV use are important for young people who confront various information-related challenges associated with transitions between school, tertiary studies, and a forbidding job market where only one in eight adults under 25 years of age find formal employment.

The mixed-design deployed four activities: (1) Semi-structured initial interviews with operators of 36 PAVs in the Western Cape Province; (2) Detailed interviews, activity/drawing probes, and task analyses with 53 teenage PAV users in 6 sites, including neighborhood libraries, larger 'central' libraries, and some cybercafés; (3) a closed-end questionnaire administered to 280 PAV users in Cape Town, and (4) an extensive debriefing session with leaders of three PAV organizations in Cape Town.

The results of the three main activities are detailed in [9], prepared as part of a broader multi-country study on public access to computing. This paper condenses the report [9], and adds the results of the consultation to find:

1. *Teenage users have developed practices which help them negotiate the respective strengths and weaknesses of public access and private mobiles.*
2. *The PAV ecology supports a valued and non-substitutable repertoire of practices for resource-constrained users, even those with 'the internet in their pocket'.*
3. *Teens can use a combination of mobile and public access internet resources to participate in media production* (though not all of them do so).
4. *PAV operator policies influence the chances for simultaneous, complementary use of the mobile internet in the venue.*

We discuss specific implications for design and policy, implementable in the short term, to increase the utility of PAVs in the mobile age. In essence, the PAV can 'welcome the mobile' into the venue through rule changes (to allow file transfers), staff training, Wi-Fi, and spatial reconfiguration to support sociality and play.

2 The South African Context

South Africa remains a society with extreme differences between rich and poor [10]. Its distinctive patterns of ICT use [11, 12] make it an important case in policy initiatives around digital inclusion [13]. Hardware is expensive due to import duties and lack of domestic manufacturing. Cell coverage is good but data, whether via terrestrial DSL or wireless GPRS/3G connections, remains relatively expensive. DSL lines are frequently capped with monthly limits as low as 1GB. Wireless data is purchased by bundles (similar to prepay airtime minutes), encouraging careful attention to a "running meter" [14]. Public access [15] to the internet in South Africa is available in libraries, NGO-run telecenters, schools, and cybercafés [16, 17].

By 2010, South Africa had over 100% mobile penetration (50 million subscriptions) but only 743,000 fixed broadband subscribers [18]. Market research suggests that 9 million unique users subscribe to the GPRS data channel [19]. Meanwhile MXit, a Java-based GPRS “internet-lite” chat application, has become popular enough, as a first-time mobile internet experience, to spawn moral panics, new phrases, and political sagas [20–24]. Thus "mobile centric" [25] internet use is increasingly prevalent, particularly in urban areas and among youth [26, 27]. Germane to our study population, most grade-11 teens in low-income township schools in Cape Town used their mobiles to access the internet as early as 2008 [28].

Books, libraries, and computers are scarce in most South African schools [29]. Unequal access to good teaching and facilities testify to contemporary class inequalities and to racial discrimination in past provisioned under apartheid [30]. Since 2002/3, public schools in the Western Cape Province, where Cape Town is located, have benefited from a rollout of computer labs and at around the same time, the SmartCape project began providing internet access to libraries in disadvantaged areas of Cape Town. With simply not enough PCs or internet access points at school or at home, teens turn to libraries, cybercafés, and other PAVs to augment their access. We had no difficulty finding teens in Cape Town libraries every afternoon and on weekends, waiting for their allotted periods of free internet access on the SmartCape computers.

The participants in this study have been growing up in post-democracy in an era of dramatic social change. Opportunities increased for Black households, just as growing unemployment plunged many into poverty. Unemployment is particularly serious for young people who leave school facing national unemployment rates of 25% [31].

If young people are unsuccessful in their exams or cannot afford to study further, contacts in their social networks may be their major source of opportunities for employment or income. Ideally, they need to find a way to develop interests or networks that connect them with the worlds of work and higher education. Many are skilled users of the ICT and their practices suggest a continuing role of PAVs in supplementing for poorly equipped schools, highlighting shortcomings of the mobile internet as currently experienced.

3 Framing/Perspective

The literature on the roles of mobiles in the theory and practice of “Information and Communication Technologies for Development” (ICT4D) is too voluminous [32] to cover here. Our interest is not voice calls or even text messaging but mobile internet use in relation to shared PC and internet access. We do not presume that mobile internet use is a substitute for public access or that it is irrelevant to those who use PAVs. Indeed, it may often not be the same “internet” on the mobile, as access modes range from accessing WAP sites to running enclosed applications, such as a chat clients, to downloading premium content or games from operator websites [33].

We join other recent HCI papers to explore communicative ecologies and repertoires in resource constrained settings [34, 35]. Communicative ecologies involve interactions among technologies, sociality, and discourse [36, 37]. Information or cultural ecologies are characterized by ‘hybrid’ interactions of physical and digital [8, 38]. We adopt ecological metaphors to explain the roles of mobile internet, other ICTs, social networks, and PAV spaces in the repertoires of young people. This approach allows us to acknowledge young people’s agency and supplement it with an awareness of power, first, the long, powerful shadows cast by institutionalized schooling centered around adult institutional authority and second, the overall contours of participation, as sculpted and truncated by economic forces.

4 Methods

Our project was intended to explore the leading edge of a non-equally-distributed phenomenon (complementary use of mobile internet and PAVs). The theoretical target population was urban teens from historically disadvantaged populations in South Africa who lack access to resources. Due to cost constraints, we pursued non-representative, purposive sampling during all phases of data gathering. Our sample is urban, restricted to Cape Town. Our informants were all PAV users. Thus our results are exploratory in nature, and should be interpreted to identify patterns and issues relevant to design and to policy, rather than to fix specific estimates of behavioral frequencies in the broader population. All questionnaires and recruiting procedures were prepared in accordance with the University of Cape Town regulations for research with human subjects, including written parental consent for participants under the age of 18.

In **phase one**, we conducted semi-structured qualitative phone or in-person interviews with 36 PAV operators in the Western Cape (11 cybercafés, 5 telecenters/NGO facilities, and 21 libraries). We selected sites at random from a list of Cape Town libraries located in low-resource neighborhoods and included the city’s two central libraries. Telecenters and Cybercafés were selected using a convenience sampling method, focused on the same low-resource neighborhoods in Cape Town.

The interviews were the same in the phone and face-to-face conditions. The questions were formulated to consider the use mobile internet. Interviews were conducted in the language of the respondents (English, isiXhosa, or Afrikaans). These brief 15-20 minute interviews were recorded, transcribed, translated, and then entered into Excel for analysis. Face-to-face interviewees were provided a prepay airtime voucher worth 40 Rand ($4.48) as an honorarium.

In **phase two**, we recruited 53 teenage participants from public access venues around Cape Town, primarily the public libraries in Delft (14%) and Langa (14%), Cape Town’s Central Library (13%), and the African Axess Internet Cafe in Langa (11%). Selected venues provide insight into a mix of government and commercial operators in diverse neighborhoods. We interviewed both Afrikaans and isiXhosa speaking participants (who were mostly multi-lingual). The group included 24 young women (45%) and 29 young men (55%) who ranged in age from 13 to 19 years old, with an average age of 17 years.

The interaction with participants varied depending on availability of PCs, safety of bringing a mobile to the venue, time, and the evolving needs of the project.

An interaction consisted of some or all of the following components. (a) A semi-structured interview on PAV and mobile internet behaviors inquired about the task that had led the participant to come to the venue. (b) Assisted drawing of a project network diagram detailed the socio-technical resources (including PCs and mobiles) recruited for the participant's most recent school project [39]. Finally, (c) a video-taped facilitated task analysis on a PC or mobile using the "think aloud" method [40]. Participants received a R15 ($1.68) airtime honorarium.

In **phase three**, we distributed a questionnaire that focused on validating and quantifying issues and patterns found in earlier phases. We used leave-behind or self-administered questionnaires, with a R10 ($1.12) incentive for completing the 15 minute task. The questionnaire was available in English-only, Afrikaans + English, and isiXhosa + English versions.

In June and July 2011, we distributed questionnaires at seven PAVs. Overall, 294 users, 171 users at 4 libraries and 141 users at 3 cybercafés (missing value on venue=3), responded to the survey. Overall, 67% of respondents were male and 33% were female. Respondents ranged in age from 12 to 55 years of age. The highest completed education among the adults ranged from high school and "matric" (passed final high school exams) up through some university training.

The questionnaire captured a diverse set of Cape Town residents, not strictly the poorest of the urban poor. Overall, 56% of respondents spoke isiXhosa as their first language, followed by English (27%), Afrikaans (24%), and Zulu (1%). English was listed as a second language by 52% of respondents. Furthermore, 62% of respondents said they were unemployed, 17% were employed, 10% were part-time employees, and 8% were self-employed. Concerning their living arrangements, 92% had electricity in their homes and 81% of respondents could get to the PAV in 20 min. or less.

Finally, in **phase four**, we held a three-hour workshop to discuss findings with representatives of African Access (a cybercafé chain), The Cape Town Libraries Smart-Cape initiative, and the manager of the Cape Access project of the Provincial Government of the Western Cape. During the workshop, we tested and refined initial findings and developed recommendations with these key stakeholders.

5 Findings

5.1 Teenage Users Have Developed Practices Which Help Them to Negotiate the Respective Strengths and Weaknesses of Public Access and Private Mobiles

In this section, we describe how teens navigate the interrelated affordances and constraints made available via mobiles and PAVs. Free use (as in a library) supports more resource-intensive goals (requiring storage space, time, and bandwidth) and stable media production. Paid use (such as a phone) supports time-sensitive goals, interpersonal communication, and low bandwidth media use.

In phase three, we asked the respondents about locations, beyond the shared access venue, where they could get Internet access. Figure 1 shows that nearly half of respondents had access at home. In our target group of teens, 42% of survey respondents had home access to a PC while 30% of in-depth interviewees owned a PC. These estimates are higher than *national* averages but are consistent with independent estimates for *urban* areas[1]. The South African census reported that 38% of Cape Town residents have PCs at home [41]. Another study found PCs in 34% of homes in Cape Town and 38% of homes in Johannesburg [42]. Additionally, 86% reported access via phone, at least once in the past. As expected, teens were more likely to report accessing the Internet via school PCs and less likely to report access via work PCs. More adults than teens reported visiting cybercafés. In terms of **activities**, browsing/searching was the number one activity, with over 80% of both teens and adults planning to do so that day. About 40% of teens and adults planned to seek help on the Internet while at the PAV.

Additional analysis revealed that 90% of teens and 95% of adults reported owning a cell phone. More interesting is that 82% of adults reported bringing their phone to the PAV while only 56% of teens did so. The significant difference (x2 p<.001) is probably due to school regulations, which prohibit students from having and using cell phones during the day because of concerns about safety and theft.

The phase two interviews provided additional insight into the **interplay** of platforms, venues, and uses. Google search (98%) and word processors (92%) were used by most teens: the mobile chat program MXit was popular (88%) while the more recently adopted social network Facebook was also used by many teens in this group (69%).

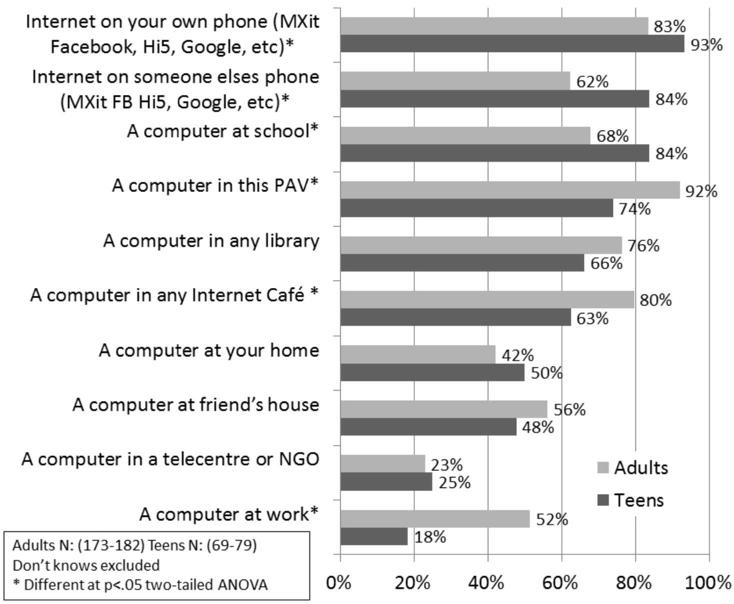


Fig. 1. Phase III Survey "Before today, where else have you accessed the internet?"

[1] PC ownership does not imply home internet access; according to the 2011 census [41] around 18-19% of urban PC owners report no home internet access.

Some tasks were clustered to take advantage of the functions and affordances of a specific platform (PC or mobile). No one used word processors on their phones. Similarly, no one preferred to use MXit on a computer. The inability to multitask on, print from, download, or display large images on a mobile phone were seen as disadvantages, and writing down information from a mobile screen was viewed as a highly inconvenient way of transferring information.

Hybrid or cross-platform practices characterized the use of Google Search and Facebook. PCs were the preferred platform for Google search, with users able to transfer their skills to mobile. Cross-platform use seemed to be the preferred way of using Facebook, with participants electing to use the accessibility and privacy of mobiles together with the convenience and economy of computers. This allowed them a less expensive mode of access, better ergonomics as well as an expanded set of features, particularly instant messaging (not available on mobile Facebook at the time).

Costs were a persistent factor. L explains: *"Sometimes I don't have money to go to the Internet cafe and don't have money to use my sister's phone, so I do come here",* and A. concurs*: "I use a computer when I do not have airtime and I use cell phone when I have airtime".* Phones were a way of accessing the mobile web ‘ka ncinci’, just a tiny bit - costly but cheaper than hourly fees in an Internet cafés.

Many web pages download painfully slowly on phones. Some networks bill for data by time rather than by megabyte, and our observations of mobile internet use showed that young people avoided certain actions on the mobile web to avoid costly downloads.

Time constraints were also a factor in free PAVs, where restrictions on PC use times as short as fifteen minutes prompted quick-turnaround ‘match and grab’ search-copy-paste-print routines adapted to the pressurized library environment. In other cases, longer periods at the computer (as in the internet cafés) or regular visits to the library allowed more sustained involvement in editing and visual design as well as integration with handwritten and photocopied material.

Those with airtime and data resources took delight in contacting peers on Facebook via mobiles. MXit allowed a more economical, less interrupted mode of interaction:

– N: *If I don't have airtime, I have like one Rand, I go on MXit, chat with my friend the whole night. (f 15)*

Since MXit costs were so low, interactions proliferated, shifting freely into popular games. The low costs of MXit also allowed evasion of parental surveillance and regulation. This freedom and abundance of communication allowed for a multitude of interactions via MXit, which we did not observe with the more expensive channels.

Some interviewees framed their interactions on MXit as a separate space, cut off from institutional ecologies. They were adamant that their use of MXit, games, or other mobile applications was entirely for pleasure, and bore absolutely no relation to schoolwork or other institutionally approved ‘instrumental’ uses. While the mobile internet was well established in the informational ecosystem, both project network diagrams in Phase II interviews and the Phase III survey indicated that it was not often utilized for schoolwork.

5.2 The PAV Ecology Supports a Valued and Non-substitutable Repertoire of Practices for Resource-Constrained Users, Even Those with 'the Internet in Their Pocket'

Clearly, phones support elements like social networking, media sharing (SD cards, photo sharing, and music playback) [43] and chat through teens' friendship-based ecologies and peer practices. Do teens still require PAVs? We highlight three assertions in the affirmative below.

First, we asked PAV operators: "In your community, has it made a big difference to have public access to the internet in your venue?" Overall, 33 of 36 respondents answered yes. The most common theme, mentioned by 15 respondents, was the way the PAV helped people search for employment. Other responses, less common, included confidence (11 mentions) and the provisioning of a safe space (7 mentions). As predicted [44], the biggest difference between free and paid venues was in the issue of cost: 10 of 25 of the free venues mentioned their primary effect in terms of the subsidy of free access. Contrast these quotes from venue operators:

- Cybercafé: *We have very good parking facilities and are located in a mall, so people feel very safe to come here, or to drop their children off here*
- Library: *It's free of charge. And it's a central safe place. Sometimes if we're offline, people will go look for other libraries. It's helped so much for development. People can keep an eye on tenders, or look for jobs online, create CVs – at no cost.*
- Library: *For school children it helps to get the most recent information we cannot provide with books. That makes their life more easy. And also from our side, information is not on books yet. So we go onto the internet…*

PAV operators were asked if there were other options for getting internet in their communities. Generally, the libraries mentioned the cybercafés and vice-versa. The differentiator is price. Consider these responses:

- Cybercafé: *The library is around the corner, where people have free internet for 45 minutes. We often get the people who do not have patience to wait at the library. Even though they have to pay, the rates are good and they would rather pay than have to wait…in a queue.*
- Library: *People come to the library cause the internet is free, even phones are too expensive. We've even had clients saying that they do have internet at home, but they prefer to work at the library cause they don't have to pay for the data*

Overall, cybercafés framed their effect as transactional based on convenience and access. Libraries stressed transformational elements of confidence and training, but also the financial subsidy.

Second, in 39 of the 53 Phase II interviews, we asked teens why they were in the PAV that day. The majority (69%) of respondents in this group had come to the venue for reasons, which they presented as primarily serious, such as working on a school project or studying. Two had come to assist a younger sibling with schoolwork, and three had come for other reasons (sending faxes for a parent's business, attending a training course, or creating publicity posters for a non-profit concert to raise funds). Many had come to the venue with more than one purpose, one related to schoolwork and another related to socializing, either online (18 %) or at the venue (5%), or pursuing an interest or hobby.

The South African school curriculum envisages an ecology, which simply does not exist, requiring project-based learning despite the limited availability of resources, such as computers, books, and libraries in public schools. As a result, both schools and young people rely on public libraries. As A. (age 19, library) explained, she would be 'prevented from learning' if she did not have access to the public library: 'all the schoolwork requires you to get information here at the library'. Public access venues also extended to home ecologies, which often did not provide computers or appropriate spaces for homework.

Several participants claimed that the availability of the libraries and internet cafes had helped them achieve higher grades at school. For example, M (m, 17) attributed success of passing the school year to web searches: *'one can say [Google] helped me pass!'*

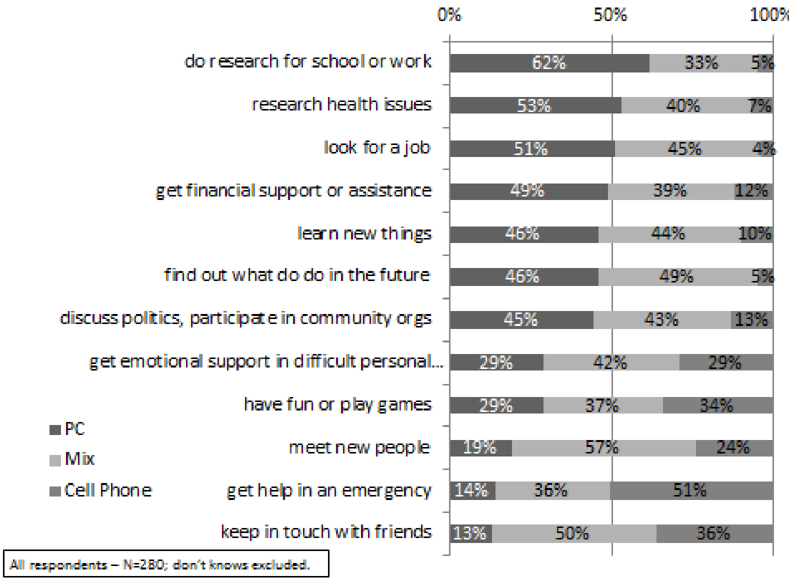


Fig. 2. Phase III Survey Stated Preferences for Tasks

Interviewees had a number of theories about what had helped increase their grades, including whether the venue had helped them to find and edit information, improved their visual presentation and neatness, and assisted them in meeting tight deadlines. They mentioned both the convenience of digital writing and the dividends paid when a 'neat' and 'professional' appearance of their projects translated into higher grades.

Third, we asked survey respondents whether they preferred to complete a series of tasks exclusively on the PC in the PAV, exclusively on their mobile, or using a mix of the two devices (for parsimony, we were unable to separate out mobile internet from the mobile in general). There were no significant differences between the preferences of teen vs. adults, although the results showed plenty of variance amongst the tasks. There are two key patterns here. First, participants preferred completing instrumental items (research, health, job search, learning new things) on PC only than on mobile-only. Second, the most common answers were often “a mix” of PC and phone – respondents did not want to choose between the devices. This is evidence for emerging complementary rather than competitive roles of PCs and phones.

Taking away the “mix” option, we asked: “if you had to choose between using only your phone for a month or only the computer in the center, which would you pick?” Teens were slightly more likely than adults to say the phone (61% vs. 52%), although the difference was not statistically significant. We also asked “if Internet access was free on your cell phone network, would you still use the PAV?” Overall, 54% of teens and 63% of adults said they still would, although again the difference was not significant.

5.3 Teens Can Use a Combination of Mobile and Public Access Internet Resources to Participate in Media Production

Some examples from studies on social support and media production illustrate different ways in which fixed and mobile internet are complementary, and hint at ways PAVs could be more welcoming of the mobile internet. Almost all of the interviewees used MXit as an academic backchannel to update one another about schoolwork or to assist one another with homework and projects. For example, A (f, 18) was able to use MXit to catch up on missed homework (‘*I would ask my friend through MXit*’), or in the case of an accounting project, to co-ordinate. Mobile phones are powerful enablers of photography and audio and video recording, but publishing and downloading audio-visual media via mobile networks is expensive and can be slow, thus phones were used for media production and (to a lesser extent) editing media, while computers and PAVs were used to produce CDs and DVDs or browse visually intensive sites. One member of a singing group used a computer to produce a CD of their songs, while another participant used the library to administer an amateur drama group and Facebook to sell a DVD of their production. These examples reveal how bandwidth and airtime constraints limit networked sharing of audio and video.

Most of the networked media production we encountered involved simple photo editing on feature phones, L. (f, 16) explained that she had edited her Facebook profile image using software on her feature phone, annotating it with the following message using a green typeface *[ee….]* Hahahaah …. preettyy …* * . The distinctive orthography, imitation of speech, and use of symbols suggest that such images are an extension of her mobile messaging practices. For her, annotated and tinted shots held a certain distinction: *A picture is just a boring picture. Everything must be edited, it looks so stylish…..*

5.4 PAV Operator Policies Strongly Affect the Chances for Simultaneous, Complementary Use of the Mobile Internet in the Venue

Phase one discovered differences between cybercafés and free venues in mobile internet use. Among the 11 cybercafés, 8 said that users printed files from their phones, and 8 had seen other examples of ‘simultaneous use’ of mobiles and PCs. Among the free venues, these observations were less frequent. Ten out of 25 reported that phone users printed files from their phones. Twelve out of 25 had seen instances of simultaneous use.

Some of this gap was likely due to differences in clientele, with older, more affluent users visiting the cybercafés. Beyond this, different venues offered different services and enforced different rules. Contrast these operator responses:

- Cybercafé: *We help with everything. Whatever question. We don’t mind. If it has anything to do with internet. On the phone or on the computer. We help. That is our service.*
- Cybercafé: *Sometimes people ask how to set up email on the phone. We help them, but charge for it. 15 min help = 10 Rand.*
- Library: *The system always gives us problems with USB connection, when they use the phone on the computers. No one was allowed to log in. We had to shut down the server and then switch on again.*
- Library: *When we see them on MXit or Facebook we are lenient. But we tell them that it is very dangerous to have the phone with them.*

Thus a venue's rules and policies are a major influence on complementary, simultaneous use of mobile internet and PAVs. Nine of the eleven cybercafés had no rules about cell phone use. By contrast, six of the libraries had total bans, eight said phones should be silent, and another four specified phones should not be connected to PCs. Five cybercafé operators reported having helped users register for phone-based services or configure their phone; four specifically allowed users to upload photos. Four free venues blocked USB ports for phone uploads/downloads. Only three helped configure phones, and another three said their users knew more than they did about phones. Cybercafés charged the equivalent of a dollar or two to help a user set up a phone, and they were more than happy to allow uploads and downloads from the phone.

6 Discussion

The discussion builds on these findings by questioning the declaration of a closed "digital divide" and identifying steps designers, policymakers, and venues can make to allow PAVs to be more effective in the coming age of mobile internet.

Though we are not the first to critique oversimplifications of how mobile internet promises to 'close the digital divide' [45], our work shows the continuing importance of safe, well-equipped venues such as these. Computers in particular play a key role in ecologies of resource-constrained feature phone users. Young people who could, in theory, be "mobile-only" internet users have instead constructed a "mobile-centric" repertoire, relying on the PAV to complete certain tasks required by school and work as well as to save money. Mobiles played a central role in participation in network-based peer interactions but the PAV was central to other kinds of participation. Public access continues to offer (at this time and in this population) critical value in certain activities. Mobile internet is making great strides, but does not yet substitute for public access, considering hardware, network, cost, space, printing, and guidance. Following [46], public access via the PC and private mobile access may be different enough to complement rather than substitute each other. Our findings echo those of [47] in the context of higher education, suggesting that while phones are used extensively in educational contexts, computer users enjoy a wider range of choices and greater convenience.

6.1 Policy and Design Changes

Phase III respondents rated their interest in PAVs' mobile phone allowances (Figure 3). There was relatively more interest in allowing Wi-Fi connections, printing directly from phones, and booking places at PCs by SMS or phone than in rule changes or booths to allow voice. Our results suggest changes in policies and design in five domains, detailed below. Our Phase IV consultations with the PAV stakeholders influenced discussion and interpretation of these results. The design and policy recommendations are intended to inform not only in the South African context, but also for public access points in other resource-constrained settings where users are increasingly likely to have data-enabled phones.

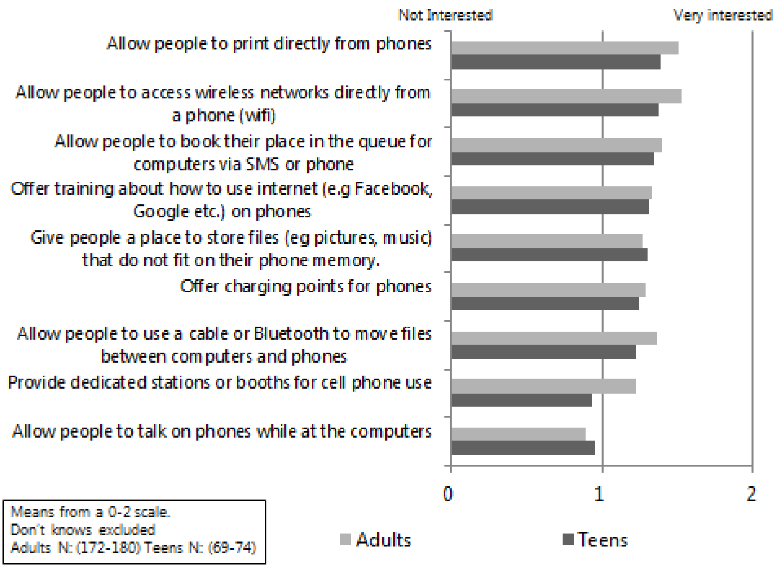


Fig. 3. Actions PAVs Could Take to Improve Phone Use

Rules. At a basic level, the “rules” prohibiting cell phone use in libraries are simply about noise management, but blanket bans could be replaced with rules to limit noise. All of the phase IV consultees were amenable to this rule change.

Affordances. If connected in some way, phones could be better storage devices, increasing the coherency of short PAV sessions and cutting down on costly inter-session printing. Since viruses are a major concern and a drag on scarce support resources [48], phase-IV consultees preferred web-based and wireless solutions to physical connections. With the assistance of design, PAVs could support the more cost-effective methods of transferring text or images to mobile-friendly cloud-based storage, and free transfer to phones via Bluetooth or USB cable.

Equally important, designs allowing phones to interface better with paper in school ecologies might include mobile interfaces for PC-less printing or mini design apps for creating printable display text or diagrams on feature phones. Phone cameras or scanning apps could help shorten queues for copiers.

Skills. Librarians and other PAV operators outside of those in for-profit cybercafés may benefit from specific training and encouragement oriented towards the opportunities presented by the mobile internet. With the proper skills, PAV staff could help users save time waiting for shared resources and encourage them to get more out of the internet in their pocket. PAV operators need new skills to help mobile-centric users with things like configuring email on phones to searching, cloud storage, and local caching, and less technical (but critical) skills, such as managing time, contacts, online reputation, and use of mobile-accessible resources for leisure and school. In Phase IV, we discussed the development of posters, covering topics such as cloud-based storage and photo processing, which might augment and backstop training provided to PAV staff. Individual tips could also be saved as small graphics and distributed via Bluetooth.

Framing. These shifts in rules and skills are not simply a matter of technical changes. Our interviews highlighted the suspicion with which PAV operators (particularly librarians) viewed phones, emphasizing negative associations with time wasting, dangerous social networks, destructive viruses, and the like. Many missed the remarkable potential that some teens had identified – the value of linking their personal digital devices to a shared access resource. Phase IV participants suggested that libraries could curate and promote electronic sources designed specifically for mobile platforms to encourage people to explore to identify and take advantage of what their phones can do as a complement to PC internet access. Free public access venues can (continue to) support young people’s mobile-centric activities beyond schoolwork, such as games, media production and distribution, and social networking, thus reframing PAVs to support participation in youth culture beyond the dominant ecologies of schooling.

Some librarians would be opposed to teens using media in the libraries, particularly noisy gaming; thus, they may oppose groups of young gamers were likely to disturb other patrons. *"We do allow games but the librarian has the right to say that if they are too rowdy they must quiet down or cut the sessions"*. Another problem relates to the fact that parents might be using the libraries as a 'babysitting' facility. Although librarians object to becoming afterschool supervisors, there is recognition that kids have few other places to go, and at least the library is a safe place. Hence, it is important to make the PAV an interesting and enjoyable place for kids pursuing various goals. Other PAVs in settings where security is less of an issue or where there are more alternatives may select different policies in this case.

Bits. Part of the appeal of the public PAVs remains free (subsidized) bits. In a country where internet data remains expensive, the long lines for PC access are understandable. On the one hand, this is a key observation regarding mobile design for resource-constrained settings. The free community hotspots common in more prosperous contexts [49] remain rare in places where bits are relatively expensive. The Wi-Fi feature of a handset is of little use to those who cannot afford to turn it on.

A more complementary policy view of Wi-Fi-enabled handsets could reduce pressure on library PCs to provide enough time for everyone. Phase IV consultees described how two of the central libraries had instituted Wi-Fi options. However, they were adamant that provisions needed to be in place to refund PAV's for the cost of their bandwidth.

Nevertheless, disaggregating the access to the machine from subsidy of the bits and increasing the flexibility, utility, and accessibility of both could be an effective move. One study suggested that worldwide, 50% of mobile phone users access the internet via Wi-Fi rather than the cell network [50]. With the spread of lower-cost data-enabled handsets, resource constrained people (without access to Wi-Fi at home, school, or work) would like free hotspots, too.

In summary, there is an opportunity for public PAVs to not simply follow the lead of cybercafés, but rather to go further in supporting their users' increasingly mobile-centric internet behaviors. Through a combination of staff training, updated rules, a Wi-Fi connection (cost permitting), and perhaps some cabling and charging stations, PAVs could provide valuable mobile-related services to users with a relatively modest investment in materials and time. Design could support specific shared-PC and own-mobile scenarios around printing, file management, and storage. Even PAV owners and frontline staff with low-resources could begin to treat the mobile internet less as an affliction (social networking) or threat (substitution) and more like the supplement to PC Internet access, as suggested by our analysis and interviews suggest it can be.

6.2 Limitations and Next Steps

The methodological limitations of a non-representative sample were discussed in the methods section. Here, we reflect on broader issues of generalizability. On the one hand, the purposive sample was appropriate for this research topic at this time. Young people in urban townships in this area are a leading case vis-à-vis mobile internet use, and the patterns we identify here will become prominent as more adopt data-enabled phones emerge in South Africa and elsewhere on the continent. Relatively good public facilities in Cape Town suggest possibilities of public access for state, schools, and businesses alike. Lower levels of PC use or of mobile internet adoption would not have allowed us to explore demand for coexistent, complementary, or competitive uses with any certainty. On the other hand, the uses observed in 2011 are a “snapshot” of an ongoing process. Conditions are changing as more mobiles become more internet-enabled, as smart phones and tablets become more affordable, as schools invest in ICTs, and as bandwidth tariffs come down. The progression towards lower cost and ‘converged’ devices may over time erode the current stark differences in affordances between public PCs and private mobiles. Thus, we are reluctant to make forecasts regarding how long the conditions (and repertoires) we observed will persist or to how long it might be before teens, as the ones we met will be able to do without public access internet. This is likely to happen before they will stop needing safe public spaces, which are conducive to learning, social networking, and cultural participation. Instead, we hew closer to current conditions to suggest some immediate steps that policymakers and PAV operators can take.

Space constraints in the paper have prevented us from addressing how broader patters of social and economic constraints shape participation and available ecologies, but this will be addressed elsewhere in a later publication. In brief, South African inequalities show how starkly power and economic stratification shape ecologies and user repertoires. Simultaneously, school and work shape demand, requiring and rewarding the genres that index middle-class ecologies.

7 Conclusion

Ecological metaphors encourage us to see PAVs as places of agency, which support learning, growth, and adaptation. Nonetheless, PAVs also interface with broader markets of participation and access. Those who can pay enjoy relative freedom to communicate, while those who cannot, must conform to rules designed to serve institutional ecologies and the adult ‘owners’ of PAV spaces, who often, particularly in the case of free public services, stigmatize, misrecognize, and fail to support young people’s mobile repertoires.

The snapshot we offer is one of stark practical and conceptual splits between (public) PCs and (private) mobiles. We see little evidence in this context that demand for PAVs among resource constrained, mobile internet-using teens will decline in the near-term, but also identify opportunities associated with the increasing prevalence of data-enabled handsets among PAV users. Thus, we suggest improvements in rules, changing affordances, upgraded skills, revised framing, and a different provision of bits (Wi-Fi) in order to help "welcome the mobile internet into the public venue".

These steps require support and funding from policymakers, and they are also a fruitful ground for interaction designers and technologists. All can be undertaken immediately to help this generation of teens, rather than waiting for convergence or lower mobile tariffs to help the next one.

Acknowledgements. This research was part of the Global Impact Study of Public Access to Information & Communication Technologies (globalimpactstudy.org) implemented by the University of Washington's Technology & Social Change Group with support from the International Development Research Centre (IDRC) and a grant from the Bill & Melinda Gates Foundation. We are also grateful for the contributions of the project team members, Silke Hassreiter, Anja Venter, Minah Radebe, Muya Koloko, Nombuyiselo Dziba, Juliette Manitshana, Baxolele Zono and Thabisa Xhalisa. Furthermore, we thank Amaal Davids and the support of Library and Information Services, City of Cape Town, Cape Access (Provincial Gov. of the Western Cape), African Axess and Silulo Internet Cafés, and the participants, without whom the research would not have been possible.

References

1. Morgan Stanley Research: The Mobile Internet Report: Ramping Faster Than Desktop Internet, the Mobile Internet Will Be Bigger Than Most Think. Morgan Stanley & Co., New York (2009)
2. Ajao, O.: Are Mobile Phones Pushing Cyber Cafes Out of Business? (2010), `http://www.ictworks.org/news/2010/05/10/are-mobile-phones-pushing-cybercafes-out-business`
3. Samii, R.: Mobile Phones: The Silver Bullet to Bridge the Digital Divide? Participatory Learning and Action 59, 44–50 (2009)
4. Chigona, W., Lekwane, O., Westcott, K., Chigona, A.: Uses, Benefits and Challenges of Public Access Points in the Face of Growth of Mobile Technology. The Electronic Journal of Information Systems in Developing Countries 49, 1–14 (2011)
5. McEwen, R., Scheaffer, K.: Orality in the Library: How Mobile Phones Challenge Our Understandings of Collaboration in Hybridized Information Centers. Library & Information Science Research 34, 92–98 (2012)
6. Prasad, R.: Rupamanjari Sinha Ray: Telecentres Go Where Mobile Phones Fear to Tread: Evidence from India. In: Regional International Telecommunications Society India Conference, New Delhi, pp. 1–26 (2012)
7. Licoppe, C.: "Connected" Presence: The Emergence of a New Repertoire for Managing Social Relationships in a Changing Communication Technoscape. Environment and Planning D 22, 135–156 (2004)
8. Nardi, B., O'Day, V.: Information Ecologies: Using Technology with Heart. MIT Press, Cambridge (1999)
9. Walton, M., Donner, J.: Public Access, Private Phone: The Interplay of Shared Access and the Mobile Internet for Teenagers in Cape Town (Global Impact Study Research Report Series), Cape Town (2012)
10. Higgs, N.T.: Measuring and Understanding the Well-being of South Africans: Everyday Quality of Life in South Africa. Social Indicators Research 81, 331–356 (2006)
11. Skuse, A., Cousins, T.: Getting Connected: The Social Dynamics of Urban Telecommunications Access and Use in Khayelitsha, Cape Town. New Media & Society 10, 9–26 (2008)
12. Skuse, A., Cousins, T.: Managing Distance: Rural Poverty and the Promise of Communication in Post-apartheid South Africa. Journal of Asian and African Studies 42, 185–207 (2007)

13. Oyedemi, T.D.: Social Inequalities and the South African ICT Access Policy Agendas. International Journal of Communication 3, 151–168 (2009)
14. Chetty, M., Banks, R., Brush, A.J.B., Donner, J., Grinter, R.E.: While the Meter Is Running: Computing in a Capped World. Interactions 18, 72–75 (2011)
15. Sey, A., Fellows, M.: Loose Strands: Searching for Evidence of Public Access ICT Impact on Development. In: Proc. iConference 2011, pp. 189–194. ACM, New York (2011)
16. James, T., Finlay, A., Jensen, M., Neville, M., Pillay, R.: Public Access ICT in South Africa. In: Gomez, R. (ed.) Libraries, Telecentres, Cybercafes and Public Access to ICT, pp. 429–451. IGI Global, Hershey (2011)
17. Pather, S., Gomez, R.: Public Access ICT: A South-South Comparative Analysis of Libraries, Telecentres and Cybercafés in South Africa and Brazil. In: AMCIS 2010, Lima, pp. 1–11 (2010)
18. ITU: Mobile-cellular Subscriptions Per 100 Inhabitants (2001-2011), `http://www.itu.int/ITU-D/ict/statistics/material/excel/2011/Mobile_cellular_01-11.xls`
19. World Wide Worx: Mobility 2011: The Mobile Internet in South Africa 2011 (executive Summary), Johannesburg (2011)
20. Bosch, T.: Wots Ur ASLR? Adolescent Girls' Use of Cellphones in Cape Town. Commonwealth Youth and Development 6 (2008)
21. Chigona, A., Chigona, W.: MXit up in the Media: Media Discourse Analysis on a Mobile Instant Messaging System. The Southern African Journal of Information and Communication 9, 42–57 (2009)
22. Schoon, A.: Raw Phones: The Domestication of Mobile Phones Amongst Young Adults in Hooggenoeg (2012), `http://eprints.ru.ac.za/2894/`
23. Walton, M., Donner, J.: Read-Write-Erase: Mobile-mediated Publics in South Africa's 2009 Elections. In: Katz, J.E. (ed.) Mobile Communication: Dimensions of Social Policy, pp. 117–132. Transaction Publishers, New Brunswick (2011)
24. Walton, M., Pallitt, N.: "Grand Theft South Africa": Games, Literacy and Inequality in Consumer Childhoods. Language and Education 26, 347–361 (2012)
25. Gitau, S., Marsden, G., Donner, J.: After Access – Challenges Facing Mobile-only Internet Users in the Developing World. In: Proc. CHI 2010, pp. 2603–2606. ACM, New York (2010)
26. Brown, C., Czerniewicz, L.: Debunking the "Digital Native": Beyond Digital Apartheid, Towards Digital Democracy. Journal of Computer Assisted Learning 26, 357–369 (2010)
27. Chigona, W., Beukes, D., Vally, J., Tanner, M.: Can Mobile Internet Help Alleviate Social Exclusion in Developing Countries? The Electronic Journal of Information Systems in Developing Countries 36, 1–16 (2009)
28. Kreutzer, T.: Generation Mobile: Online and Digital Media Usage on Mobile Phones Among Low-income Urban Youth in South Africa (2009), `http://www.tinokreutzer.org/mobile/MobileOnlineMedia-SurveyResults-2009.pdf`
29. Bloch, G.: The Toxic Mix: What's Wrong with South Africa's Schools and How to Fix It. Tafelberg, Cape Town (2009)
30. Fleisch, B.: Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics. Juta and Company (2008)
31. Statistics South Africa: Quarterly Labour Force Survey: Quarter 3 (July to September), Pretoria (2011)
32. Donner, J.: Research Approaches to Mobile Use in the Developing World: A Review of the Literature. The Information Society 24, 140–159 (2008)

33. Goldstuck, A.: The Mobile Internet Pinned Down, http://www.worldwideworx.com/the-mobile-internet-pinned-down/
34. Wyche, S.P., Smyth, T.N., Chetty, M., Aoki, P.M., Grinter, R.E.: Deliberate Interactions: Characterizing Technology Use in Nairobi, Kenya. In: Proc CHI 2010, pp. 2593–2602. ACM, Atlanta (2010)
35. Wang, T., Brown, B.: Ethnography of the Telephone: Changing Uses of Communication Technology in Village Life. In: Proc Mobile HCI 2011, pp. 1–10 (2011)
36. Foth, M., Hearn, G.: Networked Individualism of Urban Residents: Discovering the Communicative Ecology in Inner-city Apartment Buildings. Information, Communication & Society 10, 749–772 (2007)
37. Chiumbu, S.: Exploring Mobile Phone Practices in Social Movements in South Africa – the Western Cape Anti-Eviction Campaign. African Identities 10, 193–206 (2012)
38. Lindtner, S., Nardi, B.A., Wang, Y., Mainwaring, S., Jing, H., Liang, W.: A Hybrid Cultural Ecology: World of Warcraft in China. In: Proc. CSCW 2008, p. 371. ACM Press, New York (2008)
39. Pearson, M., Somekh, B.: Concept-mapping as a Research Tool: A Study of Primary Children's Representations of Information and Communication Technologies (ICT). Education and Information Technologies 8, 5–22 (2003)
40. Mavers, D.: Interpreting the Externalised Images of Pupils' Conceptions of ICT: Methods for the Analysis of Concept Maps. Computers & Education 38, 187–207 (2002)
41. StatsSA: Community Profiles (SuperWeb Interactive Format) (2012)
42. SAARF: All Media Products Survey (2012)
43. Walton, M., Marsden, G., Hassreiter, S., Allen, S.: Degrees of Sharing: Proximate Media Sharing and Messaging by Young People in Khayelitsha. In: Proc. MobileHCI 2012. ACM Press, New York (2012)
44. Gomez, R., Pather, S., Dosono, B.: Public Access Computing in South Africa: Old Lessons and New Challenges. Electronic Journal of Information Systems in Developing Countries 52 (2012)
45. Gurumurthy, A.: From Social Enterprises to Mobiles — Seeking a Peg to Hang a Premeditated ICTD Theory in an Ecology of Unequal Actors. Information Technologies & International Development 6, 57–63 (2010)
46. Bar, F., Best, M.L.: Assessing the Impact of Public Access to ICTs. Information Technologies and International Development 4 (2008)
47. Czerniewicz, L., Brown, C.: Born into the Digital Age in the South of Africa: The Reconfiguration of the "digital Citizen". In: Proc. Networked Learning 2010, pp. 859–865. Lancaster University, Lancaster (2010)
48. Bhattacharya, P., Thies, W.: Computer Viruses in Urban Indian Telecenters: Characterizing an Unsolved Problem. In: Proc ACM Workshop on Networked Systems for Developing Regions (NSDR 2011), pp. 1–6. ACM, New York (2011)
49. Forlano, L.: Anytime? Anywhere?: Reframing Debates Around Municipal Wireless Networking. The Journal of Community Informatics 4 (2008)
50. Daly, V.: Wi-Fi Grows to over 50% of Mobile Web Connections, http://news.bango.com/2011/02/02/wi-fi-grows-to-over-50-percent/

A Performance Review of Number Entry Interfaces

Patrick Oladimeji[1], Harold Thimbleby[1], and Anna L. Cox[2]

[1] Future Interaction Technology Lab, Swansea University, SA2 8PP, UK
{p.oladimeji,h.thimbleby}@swansea.ac.uk
[2] UCL Interaction Center, University College London, WC1E 6BT, UK
anna.cox@ucl.ac.uk

Abstract. Number entry is ubiquitous and there are several ways a number entry interface can be designed. Until recently, research has been focused mainly on one type of number entry interface: the numeric keypad. Various factors such as the range of values, and the space available for the design allows for several alternative interfaces to be used for number entry. In the design of medical devices such as those used for controlled drug delivery, accurate and timely entry of numbers are required in order to reduce any risk of harm to patients. This paper reviews five number entry interface styles and reports the result of an experiment conducted to evaluate the performance differences of the interfaces based on numbers used in infusion therapy in a hospital. The result shows a significant effect of interface style on speed and accuracy.

Keywords: Number entry interfaces, number entry error, user interface performance, safety critical devices.

1 Introduction

Number entry has been a part of human culture since humans learned to count and the use of numbers is evident in spoken languages around the world. Number entry interface design dates back as far as 2nd century B.C. as seen in the Salamis Tablet, an ancient Greek counting board [12, p.300]. Much later, starting in the mid 17th century, the advent of a series of mechanical calculators such as Pascal's Calculator and later, the Arithmometer, brought about a series of different design options for interacting with numbers and designing modern number entry user interfaces. Most of these interfaces are still in use today in interactive devices–although implemented in a variety of ways that account for technological advances both in software and hardware.

Tasks involving entering numbers are extremely common. For instance we enter or select numbers at the cash machine, we enter, select or modify numeric values on our microwave ovens to specify time, and we often change the volume on our music player. While performing any of these tasks, the user might be oblivious to the number entry aspect of the task, after all, you only wish to withdraw some money from the bank, warm up your food or increase the volume of music. Number entry is usually such a subtask in achieving a more primary goal that it is hardly noticed and very often perceived as trivial.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 365–382, 2013.

The reader could probably think of more than one type of interface for performing these tasks. For instance, a cash machine might use a 12-key numeric keypad, a microwave might use a dial and the music player might use a slider. There are several ways a number entry interface might be designed. Despite dating back many years, until recently, research on number entry has failed to identify a classification of number entry interfaces, and a review of the performance of the different styles of interfaces that might be beneficial to designers of interactive systems.

The most common number entry interface is the numeric keypad. It is found on telephones, calculators and keyboards. Its popularity is not surprising since it provides a direct mapping between interaction input and output and allows sequential entry of the digits that constitute the intended number much like spoken western language. Several constraints, factors and requirements in certain contexts may however restrict the adoption of a type of number entry interface. Examples of such constraints might be space requirements or footprint of the device in question or the range and precision of values intended to be addressed in the host application.

In the design of medical devices, these constraints become more compelling and designers ought to be able to make number entry interface design decisions with a clear understanding of the strengths and weaknesses of an interface. The use of many medical devices involves entry of numeric data that represent drug doses, duration of therapy or frequency of therapy. According to a report in 2007, about 7000 medicine doses are administered each day in each hospital in England and Wales [1]. Some drugs have to be administered intravenously due to the treatment requirements of patients. This sometimes involves multiple intravenous drugs to be administered simultaneously [9]. Devices such as infusion pumps, used for controlled delivery of drugs in hospitals, require timely and accurate programming in order to avoid patient harm [14]. Setting up an infusion pump requires entering numbers that correspond to the rate of infusion, the volume to infuse and duration of the infusion. Many adverse incidents in hospitals have been as a result of number entry errors in programming infusion pumps [8, 21, 22].

This paper presents the results of an empirical evaluation of five different number entry user interfaces using a customisable high-fidelity prototype. Our aim was to explore the performance difference across these interfaces with the intent of providing a summary of tradeoffs involved in choosing to implement one of the styles of interface over another.

2 Related Work

The majority of research in number entry interface performance has mainly obtained performance metric on a variety of configurations of the numeric keypad. Early research by Deininger [17] in the design of telephone keypads explored the performance differences of 16 layouts and the effects of keying behaviour of users on the keying entry speed. Deininger's experiment found that the entry speed was dependent on the participant's strategy for reading the numbers. Participants who memorized the numbers before starting the keying sequence, performed significantly better than those who referred back to the number during entry.

Further experiments on the effect of keypad layout by Conrad and Hull [5] initially suggested that the 3 x 3 grid of telephone keypad layout with 1, 2, 3 at the top was more accurate than the calculator layout with 7, 8, 9 at the top. Marteniuk et al. [11] later found that performance differences between different keypad layouts based on the two popular telephone and calculator layouts were as a result of the placement of the zero key, suggesting that the zero key be placed below the other keys.

Other studies have explored the effects of button size, button spacing and auditory feedback on number entry speed and accuracy on touch screen devices. Schedlbauer found better performance with larger button sizes [18] and Bender et al. [2] found that auditory feedback lasting between 50ms and 400ms only had a significant effect on accuracy for small targets of 10mm x 10mm.

Recently, research in number entry interface design has focused on trying to understand number entry error and improving design to reduce the risk for error. Wiseman et al. [23] built a taxonomy of number entry error based on an experiment designed to induce errors while people entered numbers on the numeric keypad. They identified 21 types of number entry errors and organised them into a framework based on the position of their causes in Norman's Action Cycle [13].

Oladimeji [15] proposes that number entry interfaces can be classified into the following groups: (1.) *Serial digit entry* describes interfaces that enforce sequential specification of the digits that make up a number typically using a numeric keypad. (2.) *Independent digit entry* describes interfaces that allow specification of the digits that make up a number in any order. (3.) *Incremental number entry* describes interfaces that allow number entry through widgets (such as dials, knobs or buttons) that are used to increase or decrease the number.

Thimbleby and Cairns [20] have shown that the probability of ten-fold errors, which are a significant risk to patient safety [10, 6], can be significantly reduced. They propose a method for parsing the input stream from a serial entry interface so that syntax errors such as multiple decimal points are correctly detected and alerted to the user.

Oladimeji et al. [16] compared a serial interface to an incremental interface to explore their effects on error detection. They found that the interface style had a significant effect on the parts of the interface on which users placed visual attention. While the incremental interface encouraged visual attention on the display, the serial interface encouraged visual attention on the input keys. Consequently, the serial interface had more undetected errors than the incremental interface.

Number entry interfaces can often be implemented in a variety of ways. For instance an independent digit entry interface such as that found in Figure 1(d) can be implemented in as many as 28 different ways. Variations in implementation might include whether or not changes to the digits wrap around. This means incrementing the digit '9' turns it to a '0' or decrementing the digit '0' turns it to a '9'. Another variation determines whether the action of increasing or decreasing a digit affects the neighbouring digit to the left. In order words, the implementation performs some arithmetic on the entire number. By running simulated trials of users making keying slips while entering numbers, Cauchi et al. [4] discovered that the differences in the implementation can have effects on the severity of error i.e., by how much an undetected error deviates from the intended number.

With a few exceptions, research in number entry has so far been based on the numeric keypad, usually testing the performance of different layouts. The serial interface offers very quick number entry and its performance scales well as the size of the number to be entered increases. However numbers used for tasks such as infusion therapy in hospitals are from a well-defined range with rules governing the allowed precision of numbers above certain thresholds. For instance, precision of numbers used for rate settings in a critical care unit might be two decimal places for numbers below 10 and only one decimal place for numbers that are between 10 and 100. In addition, from the analysis of interaction logs from infusion pumps used in 4 departments in a hospital, Wiseman et al. found that digits 0, 1, 2 and 5 were the most common digits used when programming infusion pumps [24]. These properties have made it feasible to use other number entry interfaces other than the serial interface in the design of medical devices.

For the rest of the paper, we present a detailed description of five interfaces, followed by an analytical evaluation of the speed of the interfaces using the Keystroke Level Model (KLM) [3] in section 4. We then present details of our user study in section 5, followed by the results and a discussion on the implications of the results. We conclude with some recommendations for number entry interface design.

3 Number Entry Interfaces

Based on Oladimeji's classification and implementations found on medical devices, we have implemented five exemplar interfaces: one instance of serial digit entry (*number pad*), two instances of independent digit entry (*up-down* and *five key*) and two instances of incremental entry (*chevrons* and *dial*). To evaluate the user interfaces, we built a prototype device with easily customisable keys.

Since previous researchers have explored the performance effects of different layout configurations of the serial interface, we evaluate only one instance of the serial interface in our study.

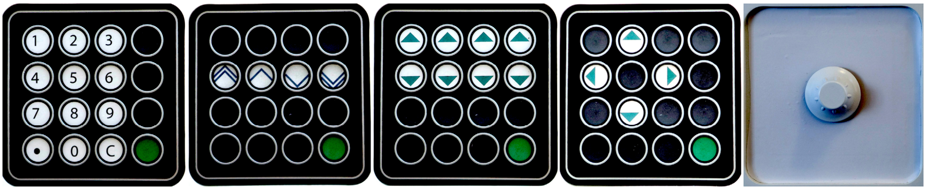


Fig. 1. The different configurations of interfaces used in our setup. From left to right (a) Number pad (b) Chevrons (c) Up-down (d) Five key (e) Dial

3.1 Number Pad

This interface allowed number entry using a 12-key numeric keypad in the telephone style layout (see Figure 1(a)). It had a decimal point and a cancel key. The decimal point key appends at most one decimal point to the number on the display. The cancel key deletes the rightmost character on the display.

3.2 Chevrons

This interface utilised four buttons in a single row. The two buttons on the left (i.e., the upward facing chevron buttons) increased the number displayed, while the buttons on the right (i.e., the downward facing chevron buttons) decreased the number. Within each pair of buttons, the double chevron buttons caused a change ten times more than the single chevron buttons. This interface allowed two modes of interaction. The user could press the buttons or they could press and hold the buttons. Pressing the buttons changes the displayed number as specified above. Pressing and holding the buttons changes the displayed number at a rate dependent on the duration of hold. Users were expected to press and hold for faster changes to the number.

3.3 Up-Down

This interface had eight buttons arranged in two rows and four columns. The top row buttons were used to increase the number and the bottom row buttons were used to reduce the number. Each column corresponded to a place value in the resulting number. For our set up, the rightmost column matched the hundredth place value and was used to increase or decrease the value by 0.01. This interface worked using the arithmetic configuration described by Cauchi et al. [4]. This means the effect of decreasing a digit from 0 or increasing a digit from 9 is carried over to the digit to the left.

3.4 Five Key

This interface had four[1] buttons arranged in a navigation style: up, down, left and right. The left and right buttons moved a cursor on the screen which selected a place value in the number and the up and down buttons increased or decreased the selected digit. Like the *up-down* interface, it worked using the *arithmetic* configuration.

3.5 Dial

This was a 24-step dial interface with unrestricted continuous rotations in both clockwise and anti-clockwise directions. Users entered numbers on this interface by turning the dial left or right to decrease or increase the number. Quicker turns on the dials caused bigger changes to the number.

---

[1] Although the reader might see this as a four key interface, we refer to this interface as five key to be consistent with naming conventions in literature, for example see Cauchi et al. [4].

4 Analytical Evaluation

Prior to running our experiment, we analysed the performance of the key based interfaces and estimated task completions times using the Keystroke-Level Model (KLM) for user performance [3]. This is a model for predicting error free expert performance and as such, we use this prediction as the best-case performance achievable by users of these interfaces. Moreover, we expected that the relative ranking produced by the KLM analyses should be maintained in the results of the experiment.

4.1 Numbers Used

We obtained log files of 60 syringe pumps from our affiliate University hospital in Swansea. The log files were completely anonymous and contained no personal information. We randomly selected 30 numbers used as rate and volume settings from the logs for our analysis. All the numbers had a decimal part and ranged from 0.26 to 83.3. A third of the numbers used had a precision of 2 decimal places.

4.2 Method

Based on simulations of the interfaces used in our experiment, we exhaustively explored the user interface model of each interface using the model discovery technique presented by Thimbleby and Gimblett [7, 19]. The user interface model discovery process produces a graph whose nodes represent the states in an interactive system and edges represent the user actions necessary to transition between the states.

To limit the number of states produced by the model discovery process, we limited the numbers addressable by the interfaces to a range covered by those used in the experiment. For each number entered in the experiment, we derived the optimal keying sequence for entering that number on the interface by searching for a shortest path from 0 to N, where N was the intended number. We ran a JavaScript implementation of the A* path finding algorithm, with cost functions that prioritised estimated time of execution over number of button clicks required to enter a number. We estimated the task completion time using standard KLM estimates for pointing and clicking [3]. As estimated by Card et al. [3], we used a value of 1100ms for the time (P) taken to point to a button and a value of 200ms for the time (K) taken to click a button. In our prediction of task time, we did not include the time (M) taken for mental preparation because we were interested in the execution time of each task. In our prediction, we do not include the initiation time, i.e., the time elapsed before the task is started or the commit time, i.e., the time taken to click the enter button to confirm the task. Using the keying sequence produced by this process, we derived an approximation for the performance time for executing the sequence.

Given that there are 24 steps in the rotary encoder used in the dial interface, to estimate the time T required to enter a given number N on this interface, we used the following expression:

$$T = \begin{cases} \frac{10t \times 100}{24} + \frac{(N-10) \times 10t}{24} & \text{if } N \geq 10 \\ \frac{N \times 100t}{24} & \text{otherwise} \end{cases}$$

Note that t is the time to perform one-step rotation on the rotary encoder. The value for t was set as 200ms. This is the value (K) taken to click a button.

4.3 Result

The Keystroke Level Model analyses produced the estimates displayed in Table 1. The predictions show that the *up-down* interface should be fastest with a slight performance edge over the *number pad* interface and the *chevrons* interface should be slowest. To validate these predictions, we designed and ran a user study.

Table 1. Approximation of the task times for the different interfaces KLM

Interfaces	Number pad	Chevrons	Up-down	Five key	Dial
Time (ms)	4875	9545	4600	6954	7855

5 Experiment

5.1 Design

The experiment was a two-way, mixed design. The within subjects independent variable was the type of number entry interface, and it had five levels: the five interfaces tested. The between subject independent variable was the instruction given to the participant: one group was instructed to enter the numbers as quickly as possible (the speed group) and the second group was instructed to enter the numbers as accurately as possible (the accurate group). We expected the speed group to exhibit more error due to time pressure. The order in which the interfaces were presented to the participants was randomized. The primary dependent variable was the speed of entry of correct numbers. Other dependent variables were the number of incorrect entries, the number of corrected errors.

5.2 Participants

There were 33 participants, 17 in the speed condition and 16 in the accurate condition. There were 22 females with 11 in the speed condition. Three participants were left-handed. The participants ranged in age from 18 - 43 with a mean age of 23.5 years (SD=4.86). The participants were undergraduate and postgraduate students in our University. Participants were randomly allocated to conditions.

Prior Experience with Interfaces. All participants were familiar with the *number pad* and reported using it on interfaces such as calculators, cash machines and telephones. Five participants (15%) were familiar with variations of the *chevrons* interface with experience using it in digital stop watches and alarm clocks, eight (24%) had prior experience with the *up-down* interface on medical devices and games, nine (27%) had prior experience with the *five key* interface on remote controls and game controllers and 19 (58%) had prior experience with the *dial* interface on microwave ovens and temperature controls.

5.3 Apparatus

To run our experiments we built a high fidelity prototype unit consisting of a colour display with a resolution of 800 x 480 pixels encased in a box with a configurable front panel and a pole clamp for mounting the device. The front panel served as the input user interface with two variations. The first variation was connected to a 4 x 4 membrane keypad with insert pockets for configuring what is displayed on the keys. The second variation was a dial powered by a 24 step rotary encoder with tactile feedback on rotation and a selection switch that is activated by pushing the dial. Both front panels were controlled by Arduino boards but with different components attached to them. With these two front panels we were able to configure five types of number entry interfaces; we configured four using the membrane keypad and one using the dial. The different configurations are shown in Figure 1.

We used a pole to mount the prototype unit as shown in Figure 2 and connected the unit to a laptop computer (a 15 inch MacBook Pro). The laptop displayed the instruction for the next trial. Instructions were displayed as numbers in the middle of the laptop screen using a white font color on a black background and a font size of 20px. We used a total of 30 different numbers in the experiment. We used 10 numbers in a practice session and 20 for the experiment. Numbers used were those described in section 4.1. We implemented the software for the experiment in JavaScript and HTML and we logged keystrokes with corresponding time values and the numeric value after each keystroke for all the trials in the experiment.

Fig. 2. The setup for the experiment showing the prototype mounted on a pole

5.4 Procedure

Each study session lasted about 45 minutes and we tested all participants individually. We informed each participant that the experiment involved entering numbers using five different number entry interfaces. Before the experiment started, the participants filled a short pre-experiment questionnaire containing demographic information about their age, gender, handedness and whether or not the participant was dyslexic.

The study itself was in five parts: one for each interface. Each part had a practice session followed by an experiment session. We randomly assigned participants to a speed or accuracy group. We instructed the speed group to enter the numbers in the instruction as quickly as possible and the accurate group to enter the numbers as accurately as possible. We randomised the order in which the users encountered the interface. A laptop computer displayed all study instructions. The instruction was a number displayed in the center of the computer screen. The next instruction was automatically displayed once the participant confirmed entry of the current trial. For the key-driven interfaces using the membrane keypad, participants confirmed entry by pressing a green button on the bottom right corner of the keypad. For the dial interface, participants pushed in the knob to confirm entry. A message signified the end of a session after a participant entered all the numbers required for that session.

Before starting each part of the experiment, the participants watched a video showing them how to use the interface they were about to test. They then had a training session where they tried out using the interface by entering 10 numbers. When they were confident with how the interface worked, they proceeded to the experiment.

The experiment session involved entering 20 numbers using the same interface they used in the training session. These numbers were different from those used in the training session. For each interface tested, each participant entered the same set of 20 numbers, although the order in which the numbers were encountered was randomised. Participants successively entered the numbers displayed in the instruction.

After the experiment, we conducted a short post-experiment semi-structured interview to find out prior experience with the interfaces and the participants relative preference for the interface styles. In return for their time, we gave the participants a gift voucher.

6 Results

6.1 Effect of Instruction

Table 2 shows means and standard deviations for each group across the five interfaces. Although we expected more errors in the speed group, an ANOVA showed that group had no statistically significant effect on the participants' speed of entry $F(1.67, 51.69) = 0.56$, $p = 0.55$ or accuracy of entry $F(2.96, 91.84) = 0.62$, $p = 0.60$. As a result, we combined both groups for the rest of the analysis.

We next summarise the results of the speed and accuracy of the interfaces.

Table 2. Mean and standard deviation for the speed and accuracy of entry between the groups

	Entry Accuracy				Entry Speed			
	Speed Group		Accurate Group		Speed Group		Accurate Group	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number pad	0.29	0.77	0.06	0.25	1906	423	2266	466
Chevrons	0.65	1.17	0.19	0.54	13355	3122	14471	2691
Up-down	0.65	1.69	0.38	0.62	3990	745	4783	1210
Five key	0	0	0	0	5231	908	5911	1213
Dial	0.94	1.92	0.44	0.81	9072	1211	10276	2024

For the rest of the results below, except the user interface preference statistic, we conducted post-hoc tests using multiple t-tests in order to find out which interfaces differed significantly from the others. For the user interface preference, we conducted post-hoc tests using multiple Wilcoxon Signed-Rank tests.

6.2 Speed of Number Entry

We separated the speed of entry of the interfaces into three constituent parts. The *initiation time* is the time elapsed between the display of the instruction and the participant's first key press. The *execution time* is the time elapsed between the participant's first key press and the last key press involved in setting the required number. The *commit time* is the time elapsed between the last key press in setting the required number and the key press for confirming the task.

Table 3. Mean and standard deviation for the initiation, execution and commit times

	Initiation		Execution		Commit		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number pad	1286	339	2080	474	730	223	4096	821
Chevrons	1535	433	13896	2931	1005	291	16436	3231
Up-down	1469	392	4374	1061	970	323	6813	1531
Five key	1463	354	5561	1105	1017	389	8040	1571
Dial	167	51	9655	1740	910	282	10732	1861

Initiation Time. A one-way repeated measures ANOVA with Greenhouse-Geisser correction found a statistically significant effect of interface style on initiation time $F(3.31, 105.85) = 200.08$, $p < 0.0001$. Post-hoc analysis showed that the *dial* interface had significantly less initiation time than all other interfaces, and the *number pad* had significantly less initiation time than the *chevrons*, *five key* and *up-down* interfaces. The *dial* interface had the shortest initiation time and the *chevron* interface had the longest. Table 3 shows the mean initiation time for all interfaces.

Execution Time. A one-way repeated measures ANOVA with Greenhouse-Geisser correction found a statistically significant effect of interface style on speed of entry

$F(1.69,54.02) = 425.5$, $p < 0.001$. Post-hoc test showed that the speed of entry of all the interfaces tested were significantly different for all pairs at the 0.001 level. The *number pad* had the shortest execution time while the *chevrons* interface had the longest. Table 3 shows the mean execution time for all the interfaces.

Commit Time. A one-way repeated measures ANOVA with Greenhouse-Geisser correction found a statistically significant effect of interface style on commit time $F(2.84, 90.83) = 24.35$, $p < 0.0001$. Post-hoc analysis showed that the *number pad* had significantly shorter commit time than all other interfaces and the *dial* had significantly shorter commit time than the chevrons interface. *five key* interface had the longest commit time. Table 3 shows the mean commit time for all interfaces.

6.3 Errors

We analysed both uncorrected errors and corrected errors. Uncorrected errors were trials for which the user transcribed and confirmed a wrong number whilst corrected errors were keying slips that the user recovered from before confirming the transcribed number.

Uncorrected Errors. A one-way repeated measures ANOVA with Greenhouse-Geisser correction found a statistically significant effect of interface style on uncorrected error $F(2.98, 95.25) = 4.68$, $p = 0.004$. Post-hoc test showed that the *five key* interface had significantly less errors than the *dial* interface $t(32)=-3.24$, $p=0.03$. The experiment elicited a total of 57 uncorrected errors, committed by 20 different participants. Only the *five key* interface was free of uncorrected errors. Table 4 shows the mean uncorrected errors on each interface.

Corrected Errors. A one-way repeated measures ANOVA with Greenhouse-Geisser correction found a statistically significant effect of interface style on number of corrections $F(2.55, 81.61) = 63.17$, $p < 0.001$. Post-hoc test showed a significant difference for all pairs of interfaces at a 0.01 level, with the exception of the pair *up-down/five key,* which did not differ significantly. Table 4 shows the mean corrected errors on each interface. The experiment elicited a total of 833 corrected errors. The *dial* interface had the highest number of corrected errors while the *number pad* had the least.

6.4 User Interface Preference

At the end of the experiment, each user ranked the interfaces in order of preference. We assigned a score of 1 to the lowest preference and a score of 5 for the highest preference. There was a statistically significant difference in the preference rating for the user interfaces $\chi 2\ (4) = 73.8$, $p < 0.0001$. The *number pad* was most preferred interface while the *chevrons* interface was the least preferred. Table 5 shows the mean ranks for all interfaces. Post-hoc test using multiple Wilcoxon Signed-Ranks test showed that the numeric keypad was preferred to all other interfaces, the up-down was preferred to chevrons, the up-down was preferred to five key, and the dial was preferred to chevrons.

Table 4. Mean and standard deviation for the corrected and uncorrected errors

	Corrected Errors		Uncorrected errors	
	Mean	SD	Mean	SD
Number pad	0.88	1.08	0.18	0.58
Chevrons	6.48	3.86	0.42	0.94
Up down	2.73	2.74	0.52	1.28
Five key	3.09	2.38	0	0
Dial	12.06	5.53	0.70	1.24

Table 5. Mean ranks for interface preference

	Number pad	Chevrons	Up-down	Five key	Dial
Mean Rank	4.81	1.69	3.50	2.44	2.56

7 Discussion

7.1 Relative Preference of Interfaces

Since all participants had prior experience using the *number pad*, it was not surprising that it was rated highest. This preference rating is also reflected in the speed exhibited by the interface. We were however surprised that the *dial* was not rated significantly worse than *up-down* and *five key* interfaces due to the number of corrected errors that occurred on the *dial*. One possible reason for this could be the significantly shorter initiation time for the *dial*. In addition, the simplicity of the interface which is based on increasing and decreasing the displayed number means the user has to do little thinking while executing the task. This was articulated by one participant, who said:

> *Dial was easier to turn the numbers. No need to move your hands from button to button.*

7.2 Types of Errors

The types of errors made during the study spanned across seven classes of errors previously reported in separate studies by Wiseman et al. [23] and Oladimeji et al. [16]. A summary of all errors is provided in Table 6.

The most common type of error was the Digit Added error. Thirteen different participants made this error on three different interfaces. Oladimeji et al. [16] reported this error in their experiment investigating the effect of interface style on error detection. While we classify this error as a member of the Digit Added error type in this

Table 6. Frequency of errors made during the experiment

Error type	Total	Interfaces	Example
Digit added	31	chevrons, up-down, dial	4.05 for 4.5
Wrong digit	8	chevrons, up-down, dial	60.5 for 62.5
Missing decimal	3	number pad	249 for 24.9
Out by ten	3	number pad, up-down	1.11 for 11.1
Missing digit	1	number pad	6.5 for 62.5
Skipped	4	number pad, chevrons, up-down,	‘’ for 62.5
No clear reason	9	chevrons, up-down, dial	56.7 for 3

paper, we believe the nature of the error makes it different from what the error type suggests. Syntactically, from the numerals that compose the intended number and the transcribed number, the error type suggests that an extra digit has been added to the number. This extra digit, in the case of errors in our experiment, is always zero. Semantically, however, this error appears to involve the inability to correctly understand the difference between the tenths and hundredths part of a number. It is possible that certain people mix up numbers matching the pattern. Indeed one participant transcribed 4.05 for 4.5 and in another trial transcribed 2.5 for 2.05. Over 50% of all unnoticed errors were of this form.

Despite featuring on the *chevrons, up-down* and *dial* interfaces, this error did not occur on the *number pad*. This could be because number entry on the *number pad* is a more direct transcription process of keying a sequence of digits that make up the intended number. Analysis of keystroke logs showed that an instance of this error occurred on the *five key* interface although it was noticed and corrected.

7.3 Difference in Speed Prediction and Study Results

We expected the absolute differences in the prediction of results and the actual study results since the participants that took part in the study were not expert users of all the interfaces. As a result, they could not match the optimal performance predicted by KLM. For the numbers used in the experiment, our prediction expected the *up-down* interface to be marginally faster than the *number pad*. Participants' familiarity with the *number pad* however meant that their performance was superior on this interface in comparison to the other interfaces. For the *number pad*, participants actually outperformed the expert model prediction and performed the task in less than half the predicted time. This could be due to the reduced target selection time exhibited in the experiment (554ms) for the *number pad* in contrast to the standard estimate used in the prediction model (1100ms).

The relative ranking in performance for the interfaces were preserved in the actual experimental data. For the *number pad*, *up-down* and *five key* interface, the observed task completion time for the experiment was less than the predicted time. On the other hand, the observed times for the *chevrons* and *dial* interfaces were higher than the predicted time. This difference in prediction could be due to the corrected error rates on the *chevrons* and *dial* interfaces which were higher than the corrected error rates

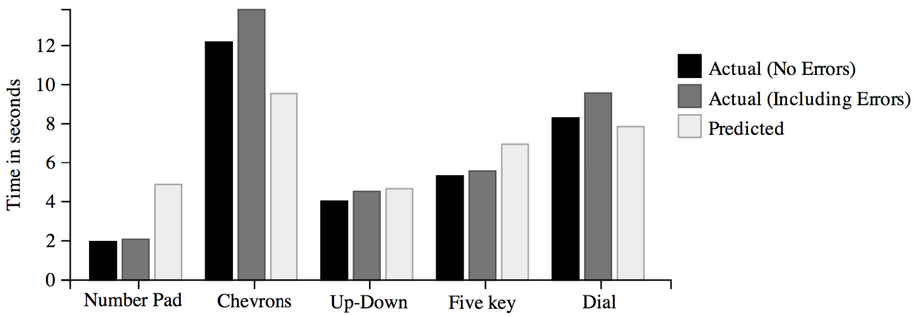


Fig. 3. A comparison of the actual and predicted performance for each interface

on the other interfaces. As a result, users spent a good portion of time correcting those errors which increased the task completion time and in some cases the frustration of users as evident in remarks during the post study interview:

> *With the dial and the chevrons, you don't really know when it switches to higher changes. I tend to stop just before I reach the value I want so that I can increase in one-step changes.*
> *For chevrons and dial, really had to time it right and let it go at the right time otherwise could be annoying.*

The corrected errors alone do not account for the deviation between the prediction and the recorded performance of the *chevrons* and *dial* interface. Another contributing factor is the strategy users develop to reduce the error rates. In the case of the *chevrons* interface, users employed discrete click interactions rather than the faster press and hold interaction, and for the *dial*, users made slower but more careful turns. Figure 3 shows that the cost of correcting an error is larger on both these interfaces in comparison to the others.

7.4 Effects of Interface Style on Number Perception

The different styles of interfaces had an effect on how numbers were perceived by the users. When users entered numbers on the serial interface (the *number pad*) or the independent digit interface (*up-down* or *five key*), they were more likely to think about the numbers as a sequence of digits without thinking much about the numeric quantity of the number as a whole. Whereas, using an incremental interface such as the *chevrons* or *dial*, participants were more likely to concentrate more on the number as a whole. One participant commented that:

> *For the number pad, up-down and the five key, I did not think of the number as a whole, just entered them digit by digit but for the chevron and dial, I had to understand the number.*

7.5 Severity of Errors Committed

The types of errors committed were closely related to the interface used to enter the number and consequently the severity of error, i.e., the deviation of the intended number from the transcribed number or the ratio between the intended and the transcribed number. Theoretically, the *number pad* and the *up-down* interfaces have the potential for producing the largest deviations from the intended number based on keying slips. This is due to the possibility of missing decimal points and missing digits on the *number pad* and the possibility of wrong place value on the *up-down* interface. We defined three levels of error severity based on the errors committed in our experiment. Low error severity referred to those errors where the ratio between the intended number and the transcribed number is at most 2, medium error severity refers to when the ratio is at most 10 and high error severity refers to when the ratio is greater than 10. Table 7 shows a summary of all errors committed and their severity.

Table 7. The severity of undetected errors committed on each interface

	Total Errors	Error Severity		
		Low	Medium	High
Number pad	4	0	3	1
Chevrons	11	10	1	0
Up-down	16	13	3	0
Five key	0	0	0	0
Dial	21	16	5	0

7.6 Incremental Interfaces and Varying Number Precision

As is typical of setting up some infusion devices used in hospital critical care, the set of numbers used for the study required that numbers below 10 were precise to two decimal places while numbers from 10 and above were precise to one decimal place. This factor meant that the display of incremental interfaces would only render numbers to the appropriate precision. As a result of this, button functions changed modes when the precision of numbers change on the display. For instance on the *chevrons* interface, when users change the value 9.99 to 10.0, the double chevron button changes meaning from ‘increase by a tenth’ to ‘increase by a unit’. Similarly on the *dial* interface, one turn on the dial changes meaning from ‘increase by a hundredth’ to ‘increase by a tenth’. We based our implementation of the *chevrons* interface on a medical device. It was also evident that some participant found the hold-down mode of the *chevrons* very difficult and challenging to predict. In this mode, the longer the buttons were held down, the larger the increments made to the number. This mode change was confusing for some users. Two participants expressed that:

> *For chevrons, the increments were very confusing. The same button did two jobs and the mode changes are confusing. Sudden changes were very confusing . . . for example you could go from 30 - 60 in a very short time span and then going back restarts the counter and climbs up rapidly. . .*

Chevrons, seem to jump quite a lot, took too long to get to intended number. Same problem with dial. It goes in sequential order rather than control individual digits.

The feature of varying precision described in this section is a requirement in infusion pumps used in critical care and intensive therapy units where low dose settings are common. It remains a design challenge to create an incremental interface that supports this form of varying precision in a way that is not confusing to the user.

8 Conclusions

Number entry is ubiquitous and number entry interfaces are very common in interactive devices. Unlike text entry, it is difficult to build a predictive model that suggests corrections for number entry error because a number entry error is usually more ambiguous than text entry errors.

In order to approach more dependable design for number entry interfaces, particularly those in use in safety critical scenarios like in medical devices, we have explored the performance differences between five different interfaces. Our results show that the *number pad* is the fastest and *five key* is the most accurate. With the results of our study, we make suggestions to designers concerning the trade offs to expect when choosing between different styles of number entry interfaces as well as the likely errors on an interface style.

The *number pad* offers the fastest mode for number entry but it comes with the risk of high severity errors such as tenfold errors. These high severity errors are mostly caused by unintentional repeated digits or unintentional missing digits. Designers should guard against repeated digits caused by overly sensitive keys, e.g., those caused by key-bounce errors [8]. Repeated decimal points should be properly parsed and alerted to the user as an error, as suggested by Thimbleby and Cairns [20]. Keypad hardware should be rigorously tested to guard against missing digits which might be caused by keys that provide tactile feedback even when they have not been completely activated electronically.

Incremental interfaces, like the *dial* or *chevrons*, have the advantage that users place their visual attention on the display of the interface so they are more likely to notice and correct any errors. They are however much slower for entering numbers and can be frustrating to users due to high likelihood of overshooting and undershooting the target value. Designers should explicitly indicate to the user, what place value in the number is being edited by using a cursor. Designers should consider offering a more direct control of the rate of change of values in order to reduce the error rate and consequently, task time for these interfaces.

For the *up-down* designers should be aware that numbers might be shifted to a wrong place value e.g., 24.5 might be transcribed as 2.45 since the decimal point is not explicitly set and the number entry keys might not be perfectly aligned to the digits they affect. Designers should consider labeling the keys on this interface or using this interface style on touch screens where input and output are on the same media and the cost of changing key layout is minimal.

Prior to this study, the majority of studies on number entry interfaces have been limited to the numeric keypad, mainly studying the effects of key layout on performance, both speed and accuracy. Our study covers a wider design space for number entry interfaces. The results provide information to designers about the tradeoffs involved in choosing one interface style over the other.

Acknowledgements. Funded as part of CHI+MED: Multidisciplinary Computer-Human Interaction research for design and safe use of interactive medical devices project EPSRC Grant Number EP/G059063/1. Carlos Monroy Aceves played a key role in the design of the prototype device.

References

1. National Patient Safety Agency: Safety in doses: medication safety incidents in the NHS. The fourth report from the Patient Safety Observatory (2007)
2. Bender, G.T., Dean, D.G., Dean, V.: Touch Screen Performance as a Function of the Duration of Auditory Feedback and Target Size. Liberal Arts and Sciences, Wichita State University: 7 (1999)
3. Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Commun. ACM 23(7), 396–410 (1980)
4. Cauchi, A., Gimblett, A., Thimbleby, H., Curzon, P., Masci, P.: Safer “5-key” number entry user interfaces using Differential Formal Analysis. In: Proceedings of HCI 2012 the 26th BCS Conference on Human Computer Interaction, pp. 29–38 (2012)
5. Conrad, R., Hull, A.J.: The preferred layout for numeral data-entry keysets. Ergonomics 11(2), 165–173 (1968)
6. Doherty, C., Donnell, C.M.: Tenfold Medication Errors: 5 Years’ Experience at a University-Affiliated Pediatric Hospital. Pediatrics 129(5) (2012)
7. Gimblett, A., Thimbleby, H.: User Interface Model Discovery: Towards a Generic Approach. In: Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, EICS 2010, pp. 145–154 (2010)
8. ISMP: Double key bounce and double keying errors. Medication Safety Alert! Acute Care Edition (2006)
9. Keohane, C.A., Hayes, J., Saniuk, C., Rothschild, J.M., Bates, D.W.: Intravenous medication safety and smart infusion systems: lessons learned and future opportunities. Journal of Infusion Nursing 28(5), 321–328 (2005)
10. Lesar, T.S.: Tenfold medication dose prescribing errors. The Annals of Pharmacotherapy 36(12), 1833–1839 (2002)
11. Marteniuk, R.G., Ivens, C.J., Brown, B.E.: Are there task specific performance effects for differently configured numeric keypads? Applied Ergonomics 27(5), 321–325 (1996)
12. Menninger, K.: Number Words and Number Symbols: A Cultural History of Numbers. Dover Publications (1992)
13. Norman, D.A.: The design of everyday things. BasicBooks (2002)
14. NPSA:Reducing harm from omitted and delayed medicines in hospital. Rapid Response Report NPSA/2010/RRR009 (2010)
15. Oladimeji, P.: Towards safer number entry in interactive medical systems. In: Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 329–332 (2012)

16. 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)
17. Deininger, R.L.: Human factors engineering studies of the design and use of push button telephone sets. Bell System Technical Journal 39(4), 995–1012 (1960)
18. Schedlbauer, M.: Effects of key size and spacing on the completion time and accuracy of input tasks on soft keypads using trackball and touch input. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 51(5), 429–433 (2007)
19. Thimbleby, H.: PressOn: Priciples of Interaction Programming. The MIT Press (2007)
20. Thimbleby, H., Cairns, P.: Reducing number entry errors: solving a widespread, serious problem. Journal of the Royal Society Interface 7(51), 1429–1439 (2010)
21. Vicente, K.J., Kada-Bekhaled, K., Hillel, G., Cassano, A., Orser, B.A.: Programming errors contribute to death from patient-controlled analgesia: case report and estimate of probability. Canadian Journal of Anesthesia 50(4), 328–332 (2003)
22. Westbrook, J.I., Rob, M.I., Woods, A., Parry, D.: Errors in the administration of intravenous medications in hospital and the role of correct procedures and nurse experience. BMJ Quality & Safety (2011)
23. Wiseman, S., Cairns, P., Cox, A.: A taxonomy of number entry error. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction, pp. 187–196. British Computer Society (2011)
24. Wiseman, S., Cox, A., Brumby, D.: Designing for the task: what numbers are really used in hospitals? In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems Extended Abstracts, pp. 1733–1738 (2012)

Predictive Input Interface of Mathematical Formulas

Yoshinori Hijikata, Keisuke Horie, and Shogo Nishida

Graduate School of Engineering Science, Osaka University
Toyonaka, Osaka, Japan
{hijikata,nishida}@sys.es.osaka-u.ac.jp,
horie@nishilab.sys.es.osaka-u.ac.jp

Abstract. Currently, inputting mathematical formulas into a document using a PC requires more effort by users than inputting normal text. This fact inhibits the spreading of mathematical formulas as internet contents. We propose a method for predicting user's inputs of mathematical formulas using an N-gram model: a popular probabilistic language model in natural language processing. Mathematical formulas are usually presented in hierarchical structure. Therefore, our method incorporates hierarchical information of mathematical formulas to create a prediction model. We try to achieve high prediction accuracy of inputting characters for mathematical formulas.

Keywords: mathematical input, probabilistic language model, predictive input, N-gram model.

1 Introduction

Mathematical formulas are helpful tools for representing knowledge in all research fields such as science, engineering, social science, and economics. The World Wide Web Consortium (W3C) has standardized the markup language for representing mathematical formulas, called MathML. Users have become able to present mathematical formulas on web pages. Most web browsers and formula manipulation software comply with MathML standards. Community sites related to mathematics are just beginning to become widespread. Those users share how to solve mathematical problems and teach each other their knowledge of mathematics. In the internet research field, search technologies are beginning to be applied to searching for mathematical formulas on web pages [4, 16]. It is believed that potential needs for presenting mathematical formulas are becoming greater on the web.

However, mathematical formulas have still not spread to the internet as a representation media to the same degree as other media such as text, graphics, sounds and movies. The currently available bothersome methods of inputting mathematical formulas are a deterrent against formula usage on the web. Although we usually use only a keyboard when inputting natural languages, one must use both a keyboard and a mouse when inputting mathematical formulas. Mathematical formulas are not simply presented as the sequence of numerical numbers, alphabet and other symbols. They

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 383–400, 2013.

are usually presented as a part of some structures such as fraction and exponent, which cannot be input using a keyboard. Users input them by clicking special buttons for the functions on the formula editor with a mouse. We believe that insisting on users using both a keyboard and a mouse engenders irritation when inputting mathematical formulas.

In this paper, we propose an input method incorporating a function to forecast mathematical characters that the user will input next. For predicting the subsequent characters that the user will input, the forecasting method uses the characters in the mathematical formulas that the user has already input. The target of the forecast is limited to mathematical structures or characters that cannot be input using a keyboard. Then, using keyboard, a user selects a prediction proposal given by our input method.

Predictive text entry is popular for inputting natural language, especially for handheld devices. In the research area of natural language processing, dictionaries are invariably used to realize predictive text entry [1, 5, 13]. The dictionary usually covers text elements with high appearance frequency. The forecasting method using the dictionary outputs prediction proposals when the text part the user has input most recently begins with some text in the dictionary. It shows the user the remainder of the matched text parts in a word or a phrase as a prediction proposal. It usually orders prediction proposals according to the most recently used order or the most frequently used order [3].

We need dictionaries specialized for mathematical formulas if we apply this method to mathematical formulas. However, unlike natural language, the subsequent characters are not narrowed down in mathematical formulas when a character is input. For example, “no see” usually comes after “Long time” in English. However, the connections among characters in mathematical formulas are not so definitive. It is difficult to create a dictionary that is effective for forecasting mathematical formulas. Therefore, we propose to make a prediction based on a probability model. We apply a probabilistic language model [7] that is popular in natural language processing, to mathematical formulas and to make a prediction using the probability output from this model. This method outputs prediction proposals in the order of the probability of the next input for the user.

We assume that our input method works on an ordinary formula editor such as Microsoft Office Formula Editor or MathType. These formula editors usually provide an input interface based on What You See Is What You Get (WYSIWYG)[1] . Users can check their input formulas on the screen immediately after they input them. They also provide buttons for inputting characters which cannot be input using a keyboard on the top of the screen. In this study, we implemented our original formula editor to evaluate our proposed input method. Our formula editor follows the characteristics of ordinary formula editors described above.

[1] WYSIWYG (What You See Is What You Get) is technology which provides an input interface where the content displayed on the screen matches the content of process (especially printed results).

2 Related Works

One popular input method of mathematical formulas is TeX. Users cannot see the input formula immediately after they input each character when using TeX. Formula editors are more popular for inputting mathematical formulas. Users can check the input formula directly because it complies with WYSIWYG. The formula editor in Microsoft Office, along with its enhanced version MathType and InftyEditor are popular formula editors. InftyEditor allows users to input a math structure by inputting a command. However, they must learn the commands in advance. Handwriting input has been studied for inputting formulas [8, 12, 17]. In these studies, the systems divide the input streams into a token using the user's input stroke (or gesture). Usually, they conduct character recognition by matching the extracted tokens to the characters in the database. Finally, they infer the structure of the formula using the stream of the identified characters. Although handwriting input requires a special input device like pen input, our proposed input interface does not require it.

Input word prediction has been studied since the 1980s in the field of natural language processing. Input characters have usually been predicted in a word unit [3]. Predicted words are usually provided when a user inputs a few beginning characters in the word [1, 5, 13]. Prediction is done by matching the input characters to the dictionary [9]. Especially when users input long characters, they want predictive word entry. However, the next characters to input become various in the mathematical formula. Therefore, it is difficult to apply the dictionary-based prediction for the predictive math entry. We introduce the N-gram model, a popular probabilistic language model, into predictive math entry. An N-gram model is usually used for predictive text entry in the research area of Augmentative and Alternative Communication (AAC) [14, 15]. Reactive Keyboard is a typical study of word prediction using an N-gram model for AAC [5]. In that study, a tree is built for the prediction, where one alphabetical character corresponds to a node. Priority is assigned to each node based on the number of occurrences of the N-gram. When a user inputs some characters, the method matches them with nodes in the tree, words in the children of the matched node are provided as prediction proposals. In another research area, Zweig et al. investigate methods for answering sentence completion questions using an N-gram model [18]. Although all of their methods do not consider the structure of a sentence, our method considers the structure of mathematical formulas.

In recent years, the predictive entry of natural language has been put to practical use in cell phones and smart phones. POBox is a major predictive entry system of Japanese [10]. It outputs a word that matches the starting few characters. It also outputs characters followed by the recent input words. T9 is a popular predictive entry for cell phones [6]. Characters are divided into nine groups in T9. Each group is assigned to one key in the cell phone. Matched words in the dictionary are shown to the user as prediction proposals if a user pushes a key. Predictive entry is useful for cell phones because they are equipped with limited number of buttons. Our assessment is that it is also useful for inputting mathematical formulas because current keyboards are not equipped with keys for numerous special mathematical characters.

3 Probabilistic Language Model

3.1 Introduction of a Probabilistic Language Model

The basic role of the probabilistic language model is to calculate the string generation probability $P(w_1^n)$ under a given string of words $w_1^n = w_1 \cdots w_n$. Each of $w_1, w_2, \cdots w_n$ stands for a word. $P(w_1^n)$ can be transformed to the following formula using the multiplication rule in probability theory [7].

$$P(w_1^n) = \prod_{i=1}^{n} P(w_i \mid w_1^{i-1}) \tag{1}$$

An N-gram model is a popular probabilistic language model in the natural language processing. We propose a method for predicting user's inputs of mathematical formulas using an N-gram model.

3.2 N-gram Model

Generally, when the probability of an event that might occur at some point in time is influenced only by the events which happened at the last N time point, we call this phenomenon an N-th order Markov Process [7]. An N-gram model is a model that approximates word occurrences as an N-1-th order Markov Process. In other words, it is considered that the occurrence of a word at some time point depends only on the last N-1 words. The general prediction model in the N-gram model becomes the following.

$$P(w_n \mid w_1^{n-1}) = P(w_n \mid w_{n-N+1}^{n-1}) \tag{2}$$

The cases of N=1, 2, 3 are respectively called unigram, bigram and trigram. The actual formula for calculating the probability of N-gram model becomes the following.

$$P(w_n \mid w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n})}{C(w_{n-N+1}^{n-1})} \tag{3}$$

We designate this probability as the N-gram probability. The calculation of this probability presents the process by which the user inputs the next word in the context that the user has input the recent words. This helps the prediction of the next input. Here, $C(w_1^n)$ stands for the number of occurrences of string of words w_1^n in the learning data.

4 Prediction Method of Math Input

4.1 Problems in Modeling Math Input

In predictive text entry, a user inputs characters sequentially. The system predicts a word to be input next, as inferred from the characters that the user has already input.

Natural language is a simple time series of data when we specifically examine the apparent sequences of characters. Therefore, it can be modeled appropriately by the probabilistic language model explained in the preceding section. Mathematical formulas have structures in their presentation. They are not simple symbolic sequences. For example, mathematical formulas including fractions or integrals have hierarchical structures. This indicates that mathematical formulas cannot be modeled using simple probabilistic language models.

4.2 Hierarchical N-gram Model

For solving the problem described above, we propose a hierarchical N-gram model. In this model, the user's log of math input (hereinafter, "log data of math input") is divided in hierarchical levels. A model is constructed in each hierarchical level.

The content and hierarchical level of the user's input are recorded in the log data of math input. Here is an example of mathematical formula for showing the actual log data.

$$\pi = \int_0^\infty \frac{sin^2 t}{t^2} dt \qquad (*)$$

In this formula, $\pi, =, \int, frac, d, t$ is in the first level, $0, \infty, t, sup, sin, sup, t$ is in the second level, 2 is in the third level. *frac* denotes fraction and *sup* signifies superscript. The hierarchical information is defined in advance in each character containing a hierarchical structure. For example, for the symbol of integral $\int$, it is defined that integral range exists in the lower level of $\int$. The previously defined hierarchical information helps to record logs with hierarchical levels.

For characters that can be input using a keyboard, the unit for recording the log is a variable, numerical value, operator, function like sin and log. The name of function is detected by preparing a dictionary in which popular function names are recorded in advance. For characters that cannot be input using a keyboard, the unit for recording the log is a character that is obtainable by clicking an input button in the formula editor. Characters that cannot be input using a keyboard are Greek alphabet characters, mathematical symbols such as differential ∂ and quantifier $\forall$, fraction, operators such as $\cap$ and $\subset$, large operators such as a summation symbol and integral symbol, accents such as tildes and circumflexes, script such as subscript and superscript. One of the log data of the above formula becomes the following. The number written in a parenthesis is the character's hierarchical level in the formula.

$$\{\pi(1), =(1), \int(1), 0(2), \infty(2), frac(1), t(2), sup(2), 2(3),$$
$$sin(2), sup(2), 2(3), t(2), d(1), t(1)\}$$

For modeling the math input, the log data are divided by the hierarchical level. The log data corresponding to the k-th level are designated as "k-th level log data". Not only the characters in the k-th level but also those in the higher level are used for representing the k-th level log data because, when predicting the next input in the lower level, the last characters in the upper level can be a trigger here. For example, in the sequence $\int(1), 0(2), \infty(2)$ appeared in the above log data, 0 and ∞ is often used for integral range. Therefore, these characters occurred by $\int$ as a trigger. The log data of each hierarchical level obtained from the above log data are shown below.

- 1-st level log data : $\{\pi, =, \int, frac, d, t\}$
- 2-nd level log data : $\{\pi, =, \int, 0, \infty, frac, t, sup, sin, sup, t, d, t\}$
- 3-rd level log data : $\{\pi, =, \int, 0, \infty, frac, t, sup, 2, sin, sup, 2, t, d, t\}$

The N-gram probability is calculated in each hierarchical level. The N-gram probability corresponding to the k-th level log data is designated as the "k-th level N-gram probability". The model of math input considering hierarchical level is designated as the "hierarchical N-gram model".

4.3 Prediction Using Hierarchical N-gram Model

A learning dataset is required for constructing a hierarchical N-gram model. It is ideal to create the model from the user's own log data. However, it is difficult to prepare a massive amount of log data of a target user (a user who will use the predictive math entry) in advance. Therefore, we prepare general log data of math input obtained from several users. The N-gram probability in each hierarchical level is calculated based on the general log data. To reflect the target user's input pattern in the model, the input log obtained while the user uses the predictive math entry is added to the log data. The hierarchical N-gram model is updated using the additional log. The prediction is made based on the N-gram probability according to the hierarchical level where the user sets focus by a keyboard or mouse. Characters with high probability to the last N-1 inputs are output as prediction proposals.

The method predicts characters to be input next in the same unit used for recording the log. It does not predict the combination of several input units at one time. The occurrence patterns become increasingly diverse and the prediction accuracy might become low if we predict the occurrence of the combination of input units. Prediction proposals are limited to characters that cannot be input using a keyboard. For math input, it is easier to input characters directly using a keyboard than to input them using predictive math entry if they can be input using a keyboard.

4.4 Smoothing

Generally, an N-gram model has a problem called the zero frequency problem. Because the N-gram probability is calculated using occurrence frequency, the probability for a word pair that does not occur in the learning dataset becomes zero. This fact suggests cases in which no prediction proposal is presented. To solve this problem, we conduct a smoothing of the probability value. We adopt a smoothing method called linear interpolation [7], which calculates the N-gram probability $P(w_n \mid w_{n-N+1}^{n-1})$ using not only the N-gram probability but also using the lower-order M-gram probability (M < N). Actually, it linearly interpolates N-gram probability $P(w_n \mid w_{n-N+1}^{n-1})$ and lower order M-gram probability in the following equation.

The N-gram probability is calculated as follows if N=2 (bigram).

$$P(w_n \mid w_{n-1}) = \lambda P(w_n \mid w_{n-1}) + (1-\lambda) P(w_n) \tag{4}$$

Therein, λ is the interpolation coefficient. We set $\lambda = 0.7$ from our experience. The N-gram probability is calculated as follows if N=3 (trigram).

$$P(w_n \mid w_{n-2}w_{n-1}) = \lambda_3 P(w_n \mid w_{n-2}w_{n-1}) + \lambda_2 P\left(w_n \mid w_{n-1}\right) + \lambda_1 P(w_n) \tag{5}$$

In that equation, λ_3, λ_2 and λ_1 respectively represent interpolation coefficients for a trigram, bigram, and unigram. We set $\lambda_3 = 0.5$, $\lambda_2 = 0.3$, $\lambda_1 = 0.2$ respectively from our experience. We used the values above for interpolation coefficients in a later experiment.

5 Interface of Predictive Math Entry

We implemented an interface equipped with our proposed prediction method. The interface was implemented in JavaScript. It runs on major web browsers. Figure 1 portrays our developed interface, which conforms to general formula editors. Therefore, the interface realizes WYSIWYG. It has buttons for inputting characters that cannot be input using a keyboard. The usage of our interface is the following. When a user moves a cursor to the place where the user inputs characters, predictive proposals are shown below the cursor in the predictive proposal display (see Figure 1). The user uses cursor buttons of the keyboard to move the cursor. The user moves a cursor to

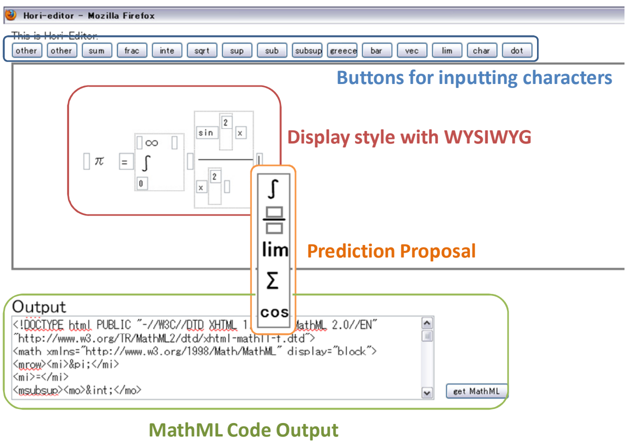


Fig. 1. Image of a mathematical editor with our prediction method of mathematical inputs

the predictive proposal display by hitting the space key if the user wants to select a character from the predictive proposals. To select a character from the predictive proposals, the user moves the focus to the character that the user wants to input by hitting the space key consecutively or using the cursor buttons. The selected character is input in the target place in the formula if the user pushes enter key. The user can obtain the input formula as in the form of MathML code by clicking the “get MathML” button.

6 Evaluation of Prediction Accuracy

The objective of this evaluation is to ascertain (1) whether our proposed prediction method outputs more highly accurate prediction results than those provided by baseline methods, and (2) whether the hierarchical N-gram model is effective for math input. As baseline methods, we use a method using only the user's recent inputs, a method using only the user's past inputs, which means that the method uses user's recent inputs, and a method using both methods. These baseline methods can be regarded as the simplest prediction methods. To evaluate the effectiveness of hierarchical modeling, we compare our hierarchical N-gram model to a general N-gram model that uses no hierarchical information.

6.1 Evaluation Data Set

We manually selected 1,000 mathematical formulas from a textbook of mathematical analysis [11] for evaluation. We invited six test subjects to input those mathematical formulas to build a dataset. The six test subjects were divided into two groups comprising three persons each. One group inputs half of the mathematical formulas. The other group inputs the remainder. Each test subject in the same group inputs the same 500 mathematical formulas. Consequently, three pieces of log data per mathematical formula are obtained. Our dataset contains 3,000 pieces of log data of math input.

We apply ten-fold cross validation to these datasets and evaluate our proposed method in terms of its prediction accuracy. In detail, the mathematical formulas are divided into 10 subsets. Then nine subsets are used as a learning dataset, another subset is used as an evaluation dataset. To calculate the prediction accuracy, we presume that a user inputs each formula of the evaluation dataset from its head to the tail. The prediction is made in each input position (see Figure 2). We examine which prediction proposal matches the actual character at the input position in the log data of the evaluation dataset. That is to say, at each input position in a mathematical formula in the evaluation dataset, we acquire prediction proposals using N-1 recent inputs extracted from the log data. We regard the character at each input position in the log data as a correct character and examine which prediction proposal corresponds to the correct character. We calculate prediction accuracy as a ratio of the input position where the top k candidates of the prediction proposals include the correct character in the log data to the entire input positions. We use the top 10 candidates for evaluation. This calculation is done for all input positions for all formulas in the test dataset.

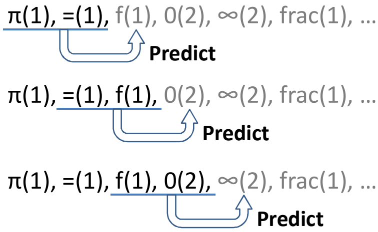


Fig. 2. Method for calculating prediction accuracy

6.2 Baseline Method

The following prediction methods are compared with our proposed prediction method. We consider these prediction methods as baselines.

- Prediction using recent inputs (Recent): This method considers the character existing at the k-th former position in the input log as a prediction proposal ranking at k-th order. It outputs prediction proposals to k=10. When k<10 at the current input, it outputs prediction proposals within the k-th rank.

- Prediction using recent inputs and their frequencies (Rct & Frq): This method orders the prediction proposals output by the recent input method described above according to their frequencies. The frequency of each candidate is calculated by counting its occurrences in the formula that the user is currently inputting.

- Unigram (N=1): This method makes a prediction based on the N-gram model (N=1). The prediction proposals become the top-k characters according to the frequencies in the whole dataset. It always outputs the same prediction proposals.

- Unigram (N=1) + Recent input method (Rct & N=1): This method is a hybrid method of unigram model and the recent input method. It calculates the appearance probability for all the characters using unigram model. For characters which exist in the former k characters in the log data of math input, it calculates the probability that a character at the j-th rank in the recent input method becomes a correct character (the prediction accuracy of the recent input method in Table 1 which can be calculated in advance). It adds the both probabilities and uses the added values for ordering the characters.

6.3 Comparison of Our Proposed Method and Baseline Methods

We show that our proposed prediction method achieves better prediction accuracy than the baseline methods. We examine our hierarchical N-gram model (N=2, 3) and their smoothing methods (hereinafter, "N-gram model (N=2, Smoothing(Smt))" and "N-gram model (N=3, Smoothing(Smt))". The implementation becomes difficult when building an N-gram model with N = 4 and higher because it requires vast memory capacity. The prediction accuracy is calculated using the mathematical formulas in the evaluation dataset. The results are presented in Table 1. Bold values represent the best prediction accuracy among the prediction methods, which is true also for the latter table. As the results show, the proposed method (N=3 with smoothing) apparently achieves the highest accuracy. Its accuracy becomes about 89% when the number of prediction proposals to show to the user is limited to the top five ranking. Its accuracy becomes about 95% when the number of prediction proposals to show to the user is limited to the top ten ranking.

We compared our proposed method (N=3) and another proposed method (N=2). When N=3, the prediction accuracies of the top three ranking and the smaller rankings (k=1,2) become higher than when N=2. The prediction accuracies of the larger rankings (the top four ranking and the larger rankings ($k = 5, 6, \cdots, 10$)) become worse because of the increase of the N-gram pairs in the learning dataset. However, when N=2, the prediction accuracies of the larger rankings become higher than when N=3. However, the prediction accuracies of the top three ranking and the smaller rankings become worse. The method can use only the latest input when N=2. Therefore, the prediction accuracies of the smaller rankings become worse than N=3. Regarding the result of N=3 with smoothing, the prediction accuracies are high both in the smaller rankings and in the larger rankings. It is apparent that the smoothing treatment compensates the above shortcomings by linear interpolation between the trigram and bigram.

Finally, we provide some insight into the results of the baseline methods. The prediction accuracy of the top five ranking is about 41% in the recent input method. Improvement of the precision is not apparent after the top five ranking. Therefore, characters that are repeated in one formula are limited to five kinds and fewer. Compared to the recent input method, in the recent and frequent input method, the accuracies become higher in the larger rankings. However, the degree of improvement is not great. Although unigram (N=1) always outputs the same prediction proposals, it achieves nearly 80% of accuracy in the top ten ranking. However, the accuracies in the smaller rankings are worse than those achieved using our proposed method. When combining unigram and the recent input method, the accuracies become better in the top three ranking and the larger rankings. However, the accuracies of the top one ranking become worse.

Table 1. Probability of the correct charcter appearing in the top k candidates

	k=1	2	3	4	5	6	7	8	9	10
Recent	0.154	0.286	0.353	0.393	0.410	0.416	0.416	0.416	0.416	0.416
Rct & Frq	0.181	0.295	0.359	0.394	0.411	0.415	0.416	0.416	0.416	0.416
N=1	0.287	0.496	0.571	0.595	0.638	0.681	0.702	0.728	0.758	0.793
Rct &N=1	0.242	0.498	0.614	0.668	0.710	0.736	0.787	0.810	0.826	0.844
N=2	0.546	0.703	0.778	0.826	0.856	0.880	0.898	0.912	0.925	0.935
N=3	0.627	0.730	0.780	0.806	0.819	0.826	0.833	0.839	0.843	0.845
N=2, Smt	0.541	0.689	0.758	0.798	0.838	0.873	0.899	0.916	0.930	0.940
N=3, Smt	**0.657**	**0.770**	**0.833**	**0.844**	**0.888**	**0.907**	**0.921**	**0.934**	**0.945**	**0.954**

6.4 Effectiveness of Hierarchical N-gram

We validate the effectiveness of introducing hierarchical information to an N-gram for predicting the math input. We compare the case which introduces hierarchical information to the model and the case which does not introduce hierarchical information to the model. The comparison is made for N-gram model (N=3, smoothing) that achieves the best accuracy in the previous subsection.

Hierarchical information works well for a case in which the user inputs characters at a shallower level after inputting characters at a deep level. We examine the prediction accuracy for this case. Those cases cover about 15% of input positions in our dataset. Table 2 shows results for those cases. They reveal that we can increase the prediction accuracy using the hierarchical information. It might decrease the overall usability if the prediction accuracy decreases for some input conditions. We infer that incorporating hierarchical information into the model is effective for predictive entry.

6.5 Effectiveness for Adding the Personal Log Data

This subsection validates the effectiveness for adding the target user's log data to the learning data. This might deal with the inconsistency problem of the input order among users. 400 log data are always used as learning data in this evaluation. One user's 500 input logs are divided in to five sets, each of which has 100 input logs. Each set is used as test data in turn. At first, 400 input logs consist of those of the other two users. As we explained, one formula has three user’s input logs. One user is randomly selected from the other two users. The selected user’s log is used here. We increase the ratio of the target user's log data to the all log data in the learning data set from 0% to 25, 50, 75, 100% and calculate the prediction accuracy. Note that we do not add the target user's log data to the original 400 learning data but replace the log data for the same formula in the original learning data set. The number of logs in the learning data stays constant. This eliminates the influence of the increase of the learning data to the prediction accuracy. Six users' logs are used for this evaluation.

The average of the six users' prediction accuracies are presented in Table 3. The results show that the prediction accuracy increases a little when the ratio of the personal logs increases. The improvement is not so large. However, we can also say that our prediction method achieves accurate prediction even if it does not use the target user's personal logs. It is expected that our method improves the prediction accuracy when many log data are used as learning data even if they are other users' logs.

Table 2. Probability of the correct character appearing in the top k candidates :comparing the case with hierarchical information(with) and that without hierarchical information(w/o)

	k=1	2	3	4	5	6	7	8	9	10
with	**0.603**	**0.712**	**0.790**	**0.816**	**0.845**	**0.872**	**0.889**	**0.915**	**0.928**	**0.940**
w/o	0.507	0.631	0.723	0.755	0.789	0.823	0.844	0.865	0.896	0.905

Table 3. Probability of the correct character appearing in the top k candidates :changing a ratio of log data inputted by a certain user in the learning dataset

	k=1	2	3	4	5	6	7	8	9	10
0%	0.477	0.596	0.647	0.690	0.729	0.772	0.793	0.815	0.834	0.844
25%	0.476	0.597	0.647	0.688	0.728	0.772	0.794	0.818	0.832	0.844
50%	0.478	0.597	0.648	0.688	0.728	0.775	0.797	**0.822**	0.837	0.849
75%	0.482	**0.600**	0.650	0.690	0.733	0.778	**0.800**	0.820	0.839	0.851
100%	**0.484**	**0.600**	**0.653**	**0.695**	**0.738**	**0.779**	**0.800**	0.820	**0.842**	**0.853**

7 Evaluation of Usability

The previous section showed that our proposed method outperforms other prediction methods in prediction accuracy. However, that fact does not mean that our proposed method helps users to input mathematical formulas. The objective of this section is to show that our interface of predictive math entry (an input interface incorporating our proposed prediction method) helps users' actual inputs by conducting user evaluation according to the interface's usability.

7.1 Experimental Condition

The evaluation of usability is accomplished according to quantitative indices and qualitative indices. The number of times of switching between a keyboard and a mouse (hereinafter, “#switching”), input time and the number of incorrect inputs (hereinafter, “#incorrect”) are used as quantitative indices. Questionnaires related to the usability of the interface, comprising multiple-choice questions and free description, were used as qualitative indices. In this experiment, 12 graduate and undergraduate students participated. We asked them about their experiences of inputting mathematical formulas on PCs. All users answered that they had some experiences on inputting mathematical formulas on PCs. They answered that they had used general formula editors or TeX. Among the participants, three users were good at inputting formulas using TeX and were able to input most formulas without seeing the reference.

7.2 Evaluation on the Differences among Interfaces

Experimental Method

This subsection shows the effectiveness of our interface of predictive math entry. Our interface was compared to a general formula editor and an input interface using TeX. Our interface of predictive math entry used in the experiment is the interface depicted in Figure 1. We eliminated the function of our predictive entry from the above interface. That is used as a general formula editor. We originally implemented an input interface using TeX on JavaScript. The screen shot is depicted in Figure 3. A user inputs mathematical formulas using TeX command in its text area. When the user clicks the compile button, the system outputs the compiled formulas with rendering. This interface does not comply with WYSIWYG.

The users input 60 formulas (20 formulas using each interface). The number of characters in each formula is set within some range. These formulas are selected from a textbook of mathematical analysis [11]. We carefully selected formulas different from the formulas in the learning dataset. All the 3,000 formulas used in evaluation of prediction accuracy are used as the learning data for our interface. The quantitative indices are measured in each input formula. After inputting 60 formulas, users answered the questionnaires for qualitative evaluation. The question items are shown in the 1-st column of Table 4. Q1 and Q3 are provided to ascertain which interface the users can use to input formulas with a good level of comfort. We set Q2 to find the time the user felt for inputting formulas. The actual time might differ from their sensory time. Q4 is provided to elicit how fast the user can learn the input method in each interface. Q5 is provided to find out which interface the user prefers with all evaluation viewpoints considered.

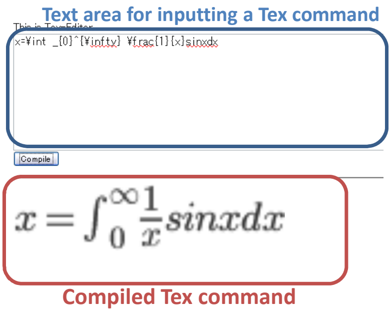


Fig. 3. Input interface using TeX

Results on Quantitative Indices

The results related to quantitative indices are presented in Table 5. The number of times necessary for key inputs and those of mouse clicks (hereinafter, #keyinputs and #clicks) are also presented in Table 5. Furthermore, we examined whether a statistically significant difference exists among the interface on these indices using a Student t-test. The results are presented in Table 6. In Tables 5 and 6, "ours" means our interface with predictive math entry, "w/o prediction" means the input interface without predictive math entry and "TeX" means the input interface using TeX. We confirmed the statistically-significant difference among interfaces in all indices except #incorrect between our interface and that without predictive entry.

The results showed on input time, it is apparent that our interface achieves the shortest input time. Input time is an important index to show that the user can input formulas smoothly. When comparing our interface and the interface without predictive entry, it is apparent that our interface decreases the #switching that is regarded as a reason that inputting mathematical formulas is bothersome. The reduction rate is about 89.1%. In the TeX interface, #incorrect becomes the three times that of the proposed interface and the interface without predictive entry. From these results, it is apparent that the incorrect inputs decrease when users can see the input formulas immediately after inputting characters. No significant difference was found between our proposed interface and the input interface without predictive math entry. However, the value of #incorrect is lower in the proposed method than the interface without predictive entry. A clear difference might be found if we increase the number of users in the experiment. #keyinputs is higher in the proposed method than the interface without predictive entry because users input characters by selecting the prediction proposal with a keyboard and move the cursor with cursor key.

Results on Qualitative Indices

The results of questionnaires for qualitative evaluation are presented in Table 4. The value for each item is the number of users who selected the item. Values shown in the parentheses are the numbers of users who can input mathematical formulas without seeing any references (hereinafter “TeX users”). Our interface achieves the best evaluation for all questionnaire items (Q1 - Q5). Especially for Q3, all users answered that they can input formulas most intuitively using our interface. In our interface, users can see the input formulas right after they input each character. Our interface also provides the prediction proposals that the users want to input. These characteristics engender the users' high evaluations to our interface. When we checked users' free descriptions, many users gave the opinion that switching a keyboard and a mouse took a burden in inputting mathematical formulas. This result supports our proposed interface that decreases the mouse input using the predictive entry.

However, some users supported the TeX interface in answers to the questionnaires. They gave the opinion that they are comfortable with the interface, and that they are used to it because they usually use TeX for inputting formulas. Actually, they were able to input formulas smoothly when using the TeX interface in the experiment. In the TeX interface, users can input formulas only by a keyboard when they learn the TeX command. Therefore, users who learn the TeX command tend to assign positive opinions to the input interface using TeX.

Table 4. Results of questionnaires for finding differences between each interfaces (The value for each item is the number of users who selected the item. The values shown in parentheses are the numbers of users who can input mathematical formulas without seeing any references).

	ours	w/o prediction	TeX
Q1. By which interface could you input the most smoothly?	10(2)	0	2(1)
Q2. By which interface could you input the most quickly?	10(2)	0	2(1)
Q3. By which interface could you input the most intuitively?	12(3)	0	0
Q4. Which interface could you get familiar with the most quickly?	11(3)	0	1
Q5. Which interface do you prefer most?	9(1)	0	3(2)

Table 5. Results of the experiment for finding differences between interfaces

	ours	w/o prediction	TeX
#clicks	1.69	13.2	0
#keyinputs	86.8	52.2	121
#switching	1.79	16.5	0
input time	66.0	74.0	89.6
#incorrect	1.38	1.57	4.43

Table 6. Results of *t*-test for the experiment for finding differences between interfaces

	Ours vs. w/o prediction	Ours vs. TeX	w/o prediction vs. TeX
#switching	***	***	***
input time	***	***	***
#incorrect		***	***

*** : p<0.01

7.3 Evaluation on the Differences in Prediction Accuracy

Experimental Method

This subsection presents results of an examination of the influence of the differences in prediction accuracy on usability. Prediction methods of three types were selected considering the difference in prediction accuracy. The N-gram model (N=3, smoothing) was selected as a method with high prediction accuracy. The recent input and unigram method and the recent input method were selected as a method with medium prediction accuracy and a method with low prediction accuracy, respectively. Hereinafter, we designate these methods as “high-accuracy method (high)”, “medium-accuracy method (medium)”, and “low-accuracy method (low)”. We conducted a user experiment using our input interface depicted in Figure 1. We changed the prediction method in this interface for the experiment.

The experimental method is the same as that used in the evaluation on the differences among interfaces. Only question items for the qualitative evaluation differ from those used in the previous evaluation. Table 7 presents the question items. Q3 is provided to know that the users noticed the difference in the prediction accuracy. Q1, Q2, and Q4 are the same questions as those used in the prior evaluation.

Table 7. Results of questionnaires for finding differences between different prediction methods

	high	medium	low
Q1. By which interface could you input the most smoothly?	11	0	1
Q2. By which interface could you input the most quickly?	12	0	0
Q3. Which interface’s prediction did you feel the most accurately?	11	0	1
Q4. Which interface do you prefer most?	12	0	0

Table 8. Results of the experiment for finding differences between different prediction methods

	high	medium	Low
#clicks	1.32	4.00	6.99
#keyinputs	87.5	96.3	73.4
#switching	1.46	4.20	8.70
input time	54.3	62.7	61.9
#incorrect	1.21	1.43	1.53

Table 9. Results of *t*-test for the experiment for finding differences between different prediction methods

	high vs. medium	high vs. low	medium vs. low
#switching	***	***	***
input time	***	***	
#incorrect	**	*	

*** : p<0.01, ** : p<0.05, * :p< 0.1

Results on Quantitative Indices

Table 8 presents the results on quantitative evaluation. #keyinputs and #clicks are also shown in Table 8. We conducted Student t-test to assess differences among the prediction methods. The results are presented in Table 9. We confirmed statistically significant difference among the methods in all indices except input time and #incorrect between the medium-accuracy method and the low-accuracy method. When particularly addressing #switching in Table 8, #switching decreases when the prediction accuracy becomes high. #incorrect decreases in the high-accuracy method compared to the other methods. The input time becomes the shortest in the high-accuracy method. From these results, it is apparent that prediction accuracy influences the usability for inputting formulas.

It is particularly interesting that the input time becomes shorter in the low-accuracy method than in the medium-accuracy method. The reason for this result is the following. The low-accuracy method (recent input method) shows only those characters which the user has input before as prediction proposals. In this case, the user can expect what characters are given as prediction proposals while inputting characters. We think that users use predictive entry by considering what characters are given as prediction proposals next. In fact, when we observed the users' input activities when they used the N-gram model (N=3, smoothing), we found that some users moved the input cursor to the predictive proposal display before checking what predictive proposals are shown there. In N-gram model (N=3, smoothing), the users input formulas under the expectation in which their target character exists in the higher rank in the predictive proposal list because the prediction accuracy is highly sufficient. Actually, they input formulas very smoothly by forecasting the prediction results. With the recent input and unigram method, the users had difficulty forecasting the next prediction proposals because unigram does not achieve the high-accuracy prediction and the prediction proposals change according to the most recent inputs. Based on this result, it is important for users to forecast the next prediction proposals for predictive entry.

Results on Qualitative Indices

The results of questionnaires for qualitative evaluation are presented in Table 7. Most users selected the high-accuracy method as the best method for all question items. This result corresponds to the results on the quantitative evaluation. For Q1 and Q3, one user selected the low-accuracy method as the best method. The reason is related to the user's input activities explained in the previous subsection. The user preferred the prediction that is easily forecasted before shifting the input cursor to the prediction proposal display.

8 Conclusions

As described in this paper, we proposed an input method that predicts the next input characters for mathematical formulas. N-gram model is applied to our method, which is a popular probabilistic language model. We incorporated hierarchical information in mathematical formulas into an N-gram model. The calculated probability was used for predictive math entry. The proposed method is evaluated using prediction accuracy. Results showed that the prediction accuracy of our method is higher than that of other baseline methods. The proposed input interface was evaluated using a user experiment. Results showed that our interface outperforms the input method without predictive entry and the input method using Tex in the usability. We expect that our interface for predictive math entry shall contribute to the spread of math contents. The ease with which users can forecast the prediction proposals is important for usability. We will examine the relation between user predictability and usability for the predictive math entry.

References

1. Copestake, A.: Augmented and alternative nlp techniques for augmentative and alternative communication. In: Proceedings of the ACL Workshop on Natural Language Processing for Communication Aids, pp. 37–42 (1997)
2. Darragh, J., Witten, H., James, L.: The reactive keyboard: A predictive typing aid. IEEE Computer 23, 41–49 (1990)
3. Garay-Vitoria, N., Abascal, J.: Text prediction systems: A survey. Univers. Access Inf. Soc. 4(3), 188–203 (2006)
4. Hashimoto, H., Hijikata, Y., Nishida, S.: Incorporating breadth first search for indexing mathml objects. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2008, pp. 3519–3523 (2008)
5. Hunnicutt, S.: Input and output alternative in word prediction. In: STL/QPRS, pp. 15–29 (1987)
6. King, M.T., Grover, D.L., Kushler, C.A., Grunbock, C.A.: Reduced keyboard disambiguating system patent (1998)

7. Kita, K.: Gengo to Keisan – Kakuritsuteki Gengo Model (Language and Computation - Probabilistic Language Model). University of Tokyo Press, Tokyo (1999)
8. LaViola Jr., J.J., Zeleznik, R.C.: Mathpad2: a system for the creation and exploration of mathematical sketches. ACM Trans. Graph. 23(3), 432–440 (2004)
9. Mackenzie, I.S., Felzer, T.: Sak: Scanning ambiguous keyboard for efficient one-key text entry. ACM Trans. Comput.-Hum. Interact. 17(3), 11:1–11:39 (2010)
10. Masui, T.: POBox: An efficient text input method for handheld and ubiquitous computers. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 289–300. Springer, Heidelberg (1999)
11. Miyake, T.: Introduction of Derivation and Integration. Baihukan, Tokyo, Japan (1992)
12. O'Connell, T., Li, C., Miller, T.S., Zeleznik, R.C., LaViola, J. J.J.: A usability evaluation of algosketch: a pen-based application for mathematics. In: Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling, SBIM 2009, pp. 149–157. ACM, New York (2009)
13. Swiffin, A., Arnott, J., Pickering, J., Newell, A.: Adaptive and predictive techniques in a communication prosthesis. In: Augmentative Aleternative Communication, pp. 181–191 (1998)
14. Vertanen, K., Kristensson, P.O.: The imagination of crowds: conversational aac language modeling using crowdsourcing and large data sources. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, pp. 700–711. Association for Computational Linguistics, Stroudsburg (2011)
15. Wandmacher, T., Antoine, J.-Y., Poirier, F.: Sibylle: a system for alternative communication adapting to the context and its user. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility, Assets 2007, pp. 203–210. ACM, New York (2007)
16. Zanibbi, R., Blostein, D.: Recognition and retrieval of mathematical expressions. International Journal on Document Analysis and Recognition (IJDAR), 1–27 (2011)
17. Zeleznik, R., Bragdon, A., Adeputra, F., Ko, H.-S.: Hands-on math: a page-based multitouch and pen desktop for technical work and problem solving. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, pp. 17–26. ACM, New York (2010)
18. Zweig, G., Platt, J.C., Meek, C., Burges, C.J.C., Yessenalina, A., Liu, Q.: Computational approaches to sentence completion. In: ACL (1), pp. 601–610. The Association for Computer Linguistics (2012)

Selection-Based Mid-Air Text Entry on Large Displays

Anders Markussen, Mikkel R. Jakobsen, and Kasper Hornbæk

Department of Computer Science, University of Copenhagen
Njalsgade 128, Building 24, 5. Floor, 2300 Copenhagen, Denmark
{amark,mikkelrj,kash}@diku.dk

Abstract. Most text entry methods require users to have physical devices within reach. In many contexts of use, such as around large displays where users need to move freely, device-dependent methods are ill suited. We explore how selection-based text entry methods may be adapted for use in mid-air. Initially, we analyze the design space for text entry in mid-air, focusing on single-character input with one hand. We propose three text entry methods: H4 Mid-Air (an adaptation of a game controller-based method by MacKenzie et al. [21]), MultiTap (a mid-air variant of a mobile phone text entry method), and Projected QWERTY (a mid-air variant of the QWERTY keyboard). After six sessions, participants reached an average of 13.2 words per minute (WPM) with the most successful method, Projected QWERTY. Users rated this method highest on satisfaction and it resulted in the least physical movement.

Keywords: Text entry, mid-air interaction techniques, large high-resolution displays, Huffman coding, multitap.

1 Introduction

Devices and interaction techniques for text entry are much researched [24], and it is clear that the effectiveness of text entry is shaped by the context of use. For instance, mobile text entry is different from desktop text entry [22,30], and typing on a tactile keyboard requires little or no visual attention, whereas text entry on a touch surface requires visual attention. Thus, text entry in non-desktop settings presents new challenges and requires new methods [39].

The present paper is motivated by a need to support text entry in one such setting, users working with a large high-resolution display. Large high-resolution displays have been shown to improve productivity [11] and, in contrast to desktop displays, they promote physical movement [3]. Around large displays, users can move in order to navigate, explore, and make sense of data on the display. We seek to design text entry methods that allow users to move in front of the display, without having to hold a device or move to a fixed location to be able to enter text.

Recent research has helped users interact with large displays by supporting object selection and manipulation (e.g., [5,14,19,35]). Mid-air interaction [16], based on tracking of users' hands, may work well for interaction in the context where users move in front of a large display. Vogel and Balakrishnan [35], for instance, used

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 401–418, 2013.

Vicon-tracking to let users point to a large display from a distance and manipulate the cursor; Nancel et al. [27] showed how mid-air gestures can be used to navigate a large display.

Whereas mid-air interactions have been explored for selection and manipulation, they are rarely used for text entry. Prior work approximates mid-air interaction by using devices such as the Nintendo Wiimote [9,33]. Other mid-air text entry techniques include AirStroke [28], a glove- and vision-based method using the Graffiti unistoke alphabet [10]. AirStroke provided a text entry rate 6.5 words per minute (WPM) without word completion. Kristensson et al. [20] demonstrated continuous recognition of mid-air gestures for writing Graffiti letters using a Kinect sensor to detect gestures within a predefined input zone.

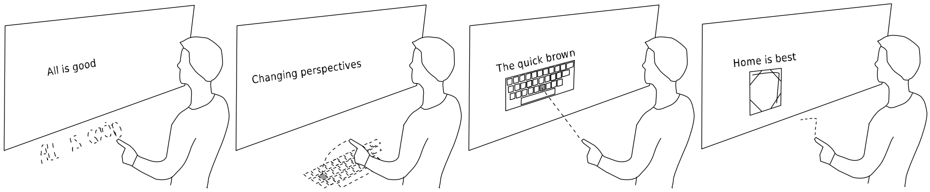
We adapt existing selection-based text entry methods to mid-air interaction with large displays. Selection-based methods rely on series of movements and activations of UI components to facilitate text entry. We do so for several reasons: (1) Leveraging familiarity with existing techniques help users learn the techniques faster, which is preferable for walk-up-and-use contexts of large displays. (2) Although mid-air text entry can potentially benefit from the increased expressiveness and additional degrees of freedom of spatial 3D input, simple and effortless techniques is recommended when the user's goal is simple [7]. (3) Despite the potential of more expressive input, the most successful mid-air text entry method to date has to our knowledge been the ray-casting selection-based QWERTY method of Shoemaker et al. [33]. More studies of adaptations of text entry methods from other contexts, such as desktop or mobile computing, are needed in order to establish a base line for mid-air text entry. In order to simplify comparison, we have chosen to focus on single-character input (rather than predictive input) and on one-handed input.

In this paper, we contribute an analysis of the design space for mid-air text entry using a structured approach that enables researchers to relate future analyses to ours. Further, we contribute an evaluation of three mid-air text entry methods that match the context of using large high-resolution displays. The methods we propose are adapted versions of previously successful methods from three different domains; game controller text entry, mobile phone text entry, and a previously successful mid-air text entry method. The methods provide a solid baseline for comparison of future mid-air text entry methods.

2 Design Space for Mid-Air Text Entry

Many considerations in designing for mid-air text entry are similar to those encountered when designing text entry in other contexts; previous work describes them thoroughly (e.g., [18]). Below we therefore focus on design considerations specific to mid-air text entry, aiming to sum up earlier mid-air text entry work in the process; Fig. 1 shows some initial design ideas that we also discuss.

A guiding context of use for the present analysis is work around large high-resolution displays. The scope of our analysis is text entry methods that support input of single characters through hand movement. Although predictive methods can



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

Fig. 1. Some initial ideas for mid-air text entry. (a) shows handwriting in mid-air; (b) shows typing on an imaginary keyboard directly in front of the user; (c) shows ray-casting to a QWERTY keyboard; (d) shows EdgeWrite gestures in mid-air.

perform significantly better, we consider single-character entry a baseline that supports a variety of text entry needs (e.g., entering a code or acronym). Even though writing with coarse body movements is certainly possible (e.g., the photographer Howard Schatz's Body Type), we follow earlier work and focus on movement of the fingers and hands (though only one hand at a time).

We structure the discussion using the design space analysis methods of MacLean et al. [25]. They distinguished questions (about what a design should do), options (answers to questions), and criteria (ways of assessing designs) as three key components for mapping a design space.

2.1 Questions and Options

Q: What Type of Movement? Earlier work has two uses of user's hand movement in mid-air. In gesture-based techniques, users write either freely or using a set of gestures. This is the idea in Fig. 1d and in many other studies [9,18,20]. GesText [18], for instance, uses accelerometer data for text entry. The most successful version used single-depth vertical and horizontal gestures to achieve 5.4 WPM.

In selection-based techniques, users point at symbols laid out in either 2D (e.g., using a QWERTY layout, see Fig. 1c) or 3D (e.g., as in [33]). Shoemaker et al. found better performance and satisfaction with techniques using 2D layouts (QWERTY and circular) compared with a 3D technique where symbols were laid out in a cube [33].

Q: 2D or 3D? Whereas many text entry methods use some form of interaction with a 2D surface (i.e., work on touch screen devices), mid-air interaction provides pitch, yaw and roll [1,9] in addition to position in 3D.

2D-approaches can mimic typing on a surface. Fig. 1b has the users imagine a QWERTY keyboard floating in front of them and use that plane for input; handwriting in mid-air (Fig. 1a) also creates an imaginary surface on which the user writes. Such designs are simple; Benko [6] suggested that we try to achieve the simplicity of touch-enabled devices when designing mid-air interaction techniques. In Kristensson's work [20], Grafitti is used in free-air, but depth (z-distance) does not appear to be used in classifying gestures: effectively, users are writing on a plane.

3D-approaches can use all of the six degrees-of-freedom. However, there seems to be a trade-off between the richness of 6-DoF gestures and an increase in complexity. For instance, the distance to the screen (or away from ones body) could be used for making selections. Research on GesText [18], however, suggested that for accelerometer input, using depth was not efficient.

Q: Typing in Relation to What? Another question is whether to use an explicit point of reference for making gestures or selections. Touch and mid-air interaction differ in that the touch surface can implicitly maintain a point of reference for the user, whereas this is not the case for mid-air interaction. Several options exist:

- Absolute point of reference, such as the display surface (Fig. 1c). Many mid-air input techniques use this approach [27].
- Relative point of reference, which could include the other hand (as in imaginary interfaces, [15]) or the location of ones feet.
- Kinesthetic point of reference, that is, a remembered hand position. For selection-based input using a QWERTY layout (see Fig. 1b), the user might initiate text entry by placing both hands on an imaginary plane; the position of left and right index fingers map to f and j on a virtual keyboard that is transformed to fit the finger placement. While this is attractive, it is well know that human hands drift [26].

Q: Visual Feedback or Not? Given the lack of tactile feedback, typing on a touch surface is primarily supported by visual feedback. In mid-air, visual support is even more challenging to provide, as mid-air text entry at large displays uses indirect input, that is, the input space is separated from the output space [17]. Users may need feedback on tracking of their movements, feedback on movements in relation to recognized gestures or characters, and feedback on production of characters.

Q: How to Initiate and Finish Writing? A well-known challenge in gesture-based input is to identify when gestures start and stop [4]. Specific gestures, pinches, input zones, and so forth has been used to delimit gestures (see for instance [35,36]). A similar challenge for selection-based input methods is to determine when a symbol is activated. In Kristensson's work [20], Grafitti input was delimited to an input zone and gestures were ignored outside this zone.

2.2 Criteria

Intuitiveness, Efficiency and Learnability. For some use contexts, the method for entering text must be easy to learn. For instance, a goal for "walk-up-and-use" systems might be that novices can enter text with minimal introduction, and perform acceptably without practice. Wobbrock et al. [37] and North et al. [29] evaluated intuitive gestures for multi-touch surfaces. The design of mid-air interaction techniques could benefit from similar studies. One approach is to draw on users' experience with widespread text entry methods. For instance, Fig. 1b and Fig. 1c benefit from users' knowledge of the QWERTY layout.

Multi-user Support. With multiple users around a large high-resolution display, text entry methods must satisfy additional criteria. First, users may physically interfere with each other's use of the display (e.g., by blocking the view of the display). Second, physically-based interactions must be social acceptable, else users might avoid physical movement because of fear of looking "silly" [31].

Distance- and Visibility-Dependence. Shoemaker et al. [33] argued that mid-air text entry methods differ in how they are affected by the distance and visibility of the display used for entering text. For instance, Fig. 1b is not distance-dependent, but Fig. 1c is.

Tracking Sensitivity. Many tracking technologies have been used for mid-air interaction, including optical tracking, gyroscopic sensors, and magnetic sensors. Some design options require accurate tracking (e.g., handwriting recognition), whereas others can do with very low tracking precision (e.g., 2D gesture-based input like Fig. 1d).

Effort and Fatigue. The motor effort needed to perform mid-air text entry (e.g., due to imprecise tracking) can be relatively large compared to typing on a keyboard. Extended periods of large movements in mid-air can cause fatigue. One approach to dealing with fatigue is to extend methods for movement minimization [22] to include the full range of body motions involved in mid-air text entry. For instance, text entry methods could be compared a priori on the effort they induce on hands, elbows, and shoulders.

3 Three Candidate Methods

The design space for mid-air text entry methods just outlined is huge. In order to identify candidate text entry methods in the space that are relevant to large high-resolution display interaction, we made two overall decisions. First, we have chosen to focus on selection-based input. Although gesture-based text input may potentially be intuitive and efficient for entering text mid-air, it is difficult to develop competitive text entry performance using existing gesture-based techniques. For instance, [28] reported a mean entry speed of 6.5 WPM for AirStroke without word completion. Second, we limit body movement to reduce fatigue, focusing on movement of hands and fingers.

For all methods, hand tracking is implemented using a glove with reflective markers attached to the back of the hand. A marker tracks the location of the index finger. Differences in angle between the hand's orientation and the vector connecting the location of the hand and the fingertip are used to detect taps. An increase in the angle of more than 5 degrees followed by a decrease of more than five degrees within 500 ms is interpreted as a tap. These thresholds were identified during pilot studies.

Based on the design space presented earlier, we can compare our methods to earlier work. A key design decision is that we use orthogonal projection instead of ray casting, as used for instance by Shoemaker et al [33]. With ray casting, users' movements are magnified as they move away from the display, which results in distance-dependent performance [33]. Given our aim of supporting users in moving in front of a large display, an important criterion was to make text entry performance distance-independent. Our methods work by projecting the hand's location orthogonally onto the display. Motor control is thus unaffected by the user's distance to the display.

3.1 H4 Mid-Air

H4 Mid-Air is an adaptation of H4 Writer [21], a text entry method operated with the thumb using four buttons on a physical game controller. H4 Writer allow users to type 20 WPM after sufficient training [21]. H4 Mid-Air uses the character selection mechanism of H4 Writer, which is based on Huffman coding: To produce a character the user produces the sequence that encodes the character. Visual feedback helps users learn the sequence for different characters.

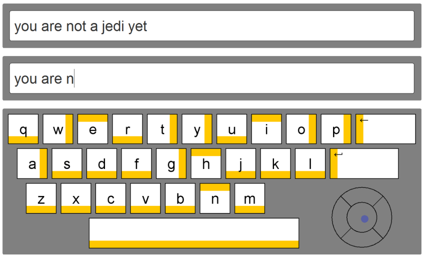


Fig. 2. The user interface for H4 Mid-Air

We adapted the method to mid-air text entry in two ways. First, without a physical controller, the user enters the sequence for a character by moving the hand, which is projected onto the display, to one of four zones and tapping (described in a later section). The motor space is divided into four slices, for each of the four zones, surrounding a "dead" center (4cm in diameter). The slices are open-ended in motor space to make targeting easier. The position of the user's hand relative to the center is shown in an on-screen radar. Second, H4 Mid-Air differs from H4 Writer by providing visual feedback on a QWERTY layout (see Fig. 2). In pilot studies, we tested the visual feedback used by MacKenzie et al. [21]. However, in this method the characters are relocated between the four zones for each step of the selection sequence and we found

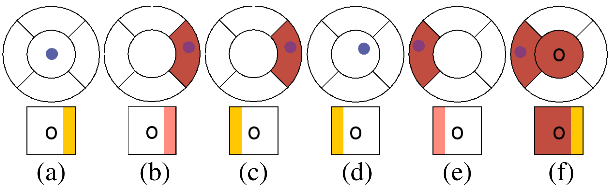


Fig. 3. Production of an 'o' with H4 Mid-Air. In order to produce an 'o', the two step Huffman code "right, left" needs to be completed. This is done through the following sequence of steps: (a) text entry has been initialized (b) hover in the indicated area. Indication is highlighted (c) a tap results in an update of the feedback and the indication changes to the next Huffman code (d) move to next action (e) hover in the indicated area (f) tap produces character. Key and center of radar is highlighted for 500 ms.

that much time was spent scanning for characters. We experimented with visual feedback that minimize scanning time. The resulting feedback is designed to improve learning and to minimize the time spent scanning for characters: (a) the user locates characters in the well-known QWERTY layout instead of the unfamiliar four-zone layout; (b) once a character is found, the user can maintain focus on the character to get visual feedback on the required sequence, without having to rescan after each input action in the sequence. The visual feedback consists of highlighting one of the sides of the key to indicate which zone that needs to be activated next to produce the character. The production of the character 'o' is shown in Fig. 3.

The motor space can be recentered by closing the hand. After 500ms the cursor is removed from the radar. Reopening the hand will define the current location of the hand as the center of motor space and put the cursor at the center of the radar. This allows clutching to redefine the motor space if a position becomes uncomfortable.

3.2 MultiTap

MultiTap shows a reduced keyboard with nine keys and a dot cursor that can be controlled by moving the hand (see Fig. 4). Three or four characters are mapped onto each key. To produce a character, the user taps the corresponding key once or multiple times—the number of taps corresponding to the character's index on that key. The character is produced when a different key is tapped or after a delay of 800ms. For example, tapping twice on "ABC", followed by tapping another key, produces 'B'. The location of the user's hand is projected onto the display plane. However, cursor movement is relative to the center key, which measures 4cm×4cm in motor space. This size was determined through pilot studies. The center is the only key with a fixed size in motor space; other keys are infinitely sized. Clutching results in resetting the motor space and positioning of the cursor at the center of keyboard. As visual feedback, a key is highlighted in orange when the cursor hovers over the key (as shown in Fig. 4). On activation of the key, the background of currently selected character is highlighted in red until the tap timeout expires or until the next activation of the key. A tap timeout of 800ms was found to be appropriate through pilot studies. This timeout is a bit lower than the usual tap timeout of 1-1.5 seconds [8] for similar methods on mobile devices.

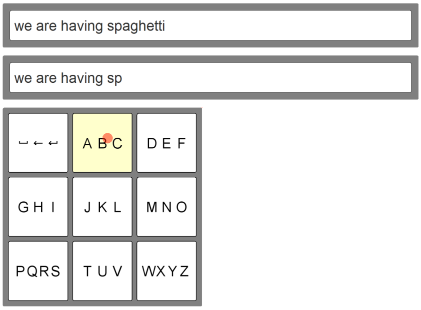


Fig. 4. The user interface for MultiTap

The method aims to provide a visibility-independent text entry method. We thus adjusted the size of the center key so that users are able to home in on the key without visual feedback during a pilot study, we tested that users were able to tap a button of this size with their eyes closed. Our intent is to leverage proprioception in homing in on the center key. Also, this method is inspired by a method that has been used on mobile phones and may thus be familiar to many users.

3.3 Projected QWERTY

Projected QWERTY shows a standard QWERTY keyboard layout on the display, with a dot cursor that can be controlled by moving the hand (see Fig. 5). A character is produced by moving the hand to a key and tapping. The location of the user's hand is projected onto the display plane. The motor space is 20cm×10cm, about the size of a physical QWERTY keyboard of a small to medium laptop. The keyboard layout was shown in about 50cm×25cm. Pilot studies found these dimensions appropriate. Tracking, tapping and clutching are implemented in the same way as for H4 Mid-Air and MultiTap. Clutching results in resetting the motor space and positioning of the cursor at the center of keyboard. Visual feedback is similar to MultiTap: hovering over a key highlights the key in orange (see Fig. 5); when activated, the key is highlighted for 500ms or until the cursor leaves the key; when producing a character, it is highlighted in red in the transcribed string for 500ms.

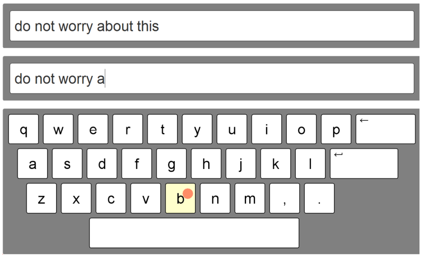


Fig. 5. The user interface for Projected QWERTY

Projected QWERTY resembles the QWERTY keyboard technique of [33]. Instead of ray casting from the hand to the screen based on the hand's orientation, Projected QWERTY projects the hand's location onto the display plane. Thereby cursor movement relative to hand movement is independent of distance: in contrast, moving the cursor using ray-casting leads to magnified cursor movement as the user moves away from the display. Another benefit is that the motor space is absolute. Projected QWERTY is thus theoretically visibility-independent, since key locations in motor space are constant. However, taking the number of keys on a QWERTY keyboard into account, actual visibility-independence is not expected.

3.4 Differences among Methods

Next, we describe the main differences among the methods. First, the methods vary in the need for visual feedback on hand movements. Projected QWERTY has small activation areas in motor space, which makes it difficult to target a key without visual feedback. Both H4 Mid-Air and MultiTap use open-ended activation areas, located around the point of reference, which should allow users to point without visual feedback. Text entry without visual feedback assumes that the user has memorized the activation sequences for all characters ([21] suggested it possible for H4 Mid-Air).

Second, the number of buttons/activation areas is varied among methods, which may impact the visibility-dependence of the method. However, it also has an impact on the physical movement needed to operate the methods. H4 Mid-Air requires a sequence of movement and taps, MultiTap requires one movement followed by a sequence of taps, and Projected QWERTY requires only one movement and one tap in order to produce a character. We investigate how these differences affect the amount of hand movement required and the usability of the methods.

Third, the methods aim to ease adoption by novices. Using the well-known alphabet layouts for Projected QWERTY and MultiTap should help novices adopt the methods faster. Although H4 Mid-Air leverages the QWERTY layout, it does not benefit from previous user experience as well as the other methods. However, the H4 Huffman coding has proven fast in previous longitudinal studies [2,21].

4 Empirical Study

To evaluate the three mid-air text entry methods, we conducted a controlled experiment. The experiment spanned six sessions in which participants used the methods to transcribe sentences.

4.1 Participants

Six participants (one female) were recruited; ages ranged from 21 to 28. One participant performed the experiment left-handedly. None of the participants were native English speakers, but all participants rated their level of English between good (2) and fluent (6).

4.2 Apparatus

A 2.80m×1.20m display containing 7680×3240 pixels was used. The display is back-projected by 12 projectors that are arranged as tiles in a 4×3 layout. Participants stood 2 meters away from the display while transcribing sentences.

For tracking, we used the OptiTrack (http://www.naturalpoint.com/optitrack) motion capture system equipped with 24 V100:R2 cameras. The system provides tracking data at 100 fps. The tracking precision was ±4 mm over the entire tracking volume; participants were located in a part of the volume with higher precision. Although the OptiTrack system is expensive, we decided to use it for several reasons. First, affordable tracking systems available at the time of this study have low precision. High-precision tracking reduces noise in the data and thus gives confidence that we are measuring the performance of the techniques, and not effects of noise caused by current tracking equipment. Second, affordable systems (e.g., Microsoft Kinect) have limited fields of view and require the user to interact at certain, constrained distances. This limits users' ability to move freely, which is needed around large displays. Third, tracking technology is improving at a high rate. The present study can be replicated and the text entry methods practically applied with widespread equipment within a few years; use of high-precision tracking thus ensures a better baseline for future research.

4.3 Tasks

Users were asked to transcribe randomly selected phrases from the MacKenzie and Soukoreff corpus [23]. Sentences were transcribed as unconstrained text entry [38].

Consequently, users were allowed, but not forced, to delete previously entered text and correct any errors that they noticed. Participants were instructed to complete sentences as quickly and accurately as possible.

4.4 Design

The experiment was conducted as a within-subjects design with the three text entry methods (H4 Mid-Air, MultiTap, and Projected QWERTY) and text entry session as independent variables. Dependent measures were text entry speed, error rate, physical hand movement, and subjective satisfaction: they are detailed in the next subsection.

Participants completed 6 sessions. During each session, participants transcribed text with all text entry methods, completing 2 blocks of 5 sentences with each method. The order in which the methods were used was fully counterbalanced between participants and sessions. In all, 1080 phrases were transcribed (6 participants × 6 sessions × 3 text entry methods × 2 blocks × 5 phrases).

4.5 Data Collection and Analysis

We collected data describing participants' interaction with the methods, including data on the location of their hand. From these data we used StreamAnalyzer [38] to calculate text entry speed and error rate, and derived a measure of hand movement. Text entry speed was calculated using equation 1, where |T| is the length of the transcribed string and S is the time in seconds from the entry of the first character to the entry of the last character.

$$WPM = \frac{|T|-1}{S} \times 60 \times \frac{1}{5} \ , \tag{1}$$

Error rate was calculated using the methods described by Soukoreff and MacKenzie [34], as Minimum String Distance (MSD), Uncorrected Errors (ErrUC) and Corrected Errors (ErrC).

To provide a quantitative measure of the physical effort put into typing, we defined Hand Movement Per Word (HMPW). HMPW is calculated as the sum of the distances travelled by the hand between tracking frames. Calculating the sum of distances over data containing noise may result in erroneous values for HMPW. We therefore ran the Douglas-Peucker algorithm [12] on the movement data with a threshold of 2mm in order to minimize noise. HMPW is measured in meters per word in the transcribed string; one word is five characters including whitespaces. As with WPM, HMPW is measured from the entry of the first character to the entry of the last character and is calculated as follows:

$$HMPW = \frac{|HM|*5}{|T|-1} \tag{2}$$

|HM| is the sum of distances between consecutive tracking frames and |T| is the length of the transcribed string. We removed the first phrase with each text entry method from this calculation because it was typed slower and with more hand

movements than subsequent sentences, and because participants said they used the first phrase to get used to a method.

Subjective satisfaction was measured using three instruments at various stages throughout the experiment: (1) To get an estimate of how much effort participants had to put into the operation of a text entry method, SMEQ [32] was administered to participants each time they had finished using a text entry method. (2) We adapted thirteen questions from the ISO-9241-9 standard [13] to evaluate physical operation, fatigue and comfort, speed and accuracy, and overall usability. We administered these questions at each participant's first and last session to gauge their experience with the methods after little training and after some training. (3) After the first and last session, participants ranked the three text entry methods (1 being the one they liked the most, 3 being the one they liked the least).

4.6 Procedure

At the first session, participants were introduced to the concept of mid-air text entry and to the three methods being evaluated. Participants were then allowed to practice with each method. They were asked to practice until they felt confident with using the method, and felt they would be able to reproduce any randomly chosen character. No participant entered more than 4 sentences per text entry method during practice.

At the beginning of sessions 2 to 6, participants were asked to practice with each method until they felt confident that they would be able to reproduce any randomly chosen character. Typically, participants practiced only one sentence to reacquaint themselves with a method. Session lasted 45 minutes on average.

In each session, participants completed two blocks of five sentences with each text entry method. Participants were allowed a short break after each block. After having transcribed the two blocks of sentences with a method, participants were administered an electronic SMEQ; at the end of session 1 and 6 participants answered the ISO questionnaire and ranked interfaces by order of preference.

5 Results

A 3 (text entry method) × 6 (session) repeated measures analysis of variance was performed on the entry speeds, the error rates, and the measure of hand movement. Significant effects were examined using Bonferroni-corrected pairwise comparisons.

5.1 Text Entry Speed

Fig. 6 shows the text entry speed in words per minute (WPM) for the text entry methods across sessions. We found a main effect for text entry method, $F(2, 10) = 109.63$, $p < .01$. Pairwise comparisons showed that Projected QWERTY ($M = 11.63$, $SD = 2.29$) was faster than MultiTap ($M = 8.38$, $SD = 2.45$), $p < 0.01$, which again was faster than H4 Mid-Air ($M = 4.19$, $SD = 1.25$), $p < 0.01$. In the final session, Projected QWERTY achieved 13.2 WPM (*SD* = 1.55), MultiTap almost 9.5 WPM (*SD* = 2.19), and H4 Mid-Air 5.2 WPM (*SD* = 1.31).

A main effect was also found for session, $F(5, 25) = 176.22$, $p < .01$, showing that users improved over sessions. Speed improved from first to last session by 39% for Projected QWERTY, 47% for MultiTap, and 80% for H4 Mid-Air, all significant at the $p < .001$ level. No significant interaction was found between method and session.

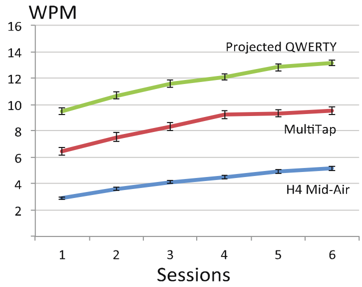


Fig. 6. Mean words-per-minute (WPM) for methods over sessions. Error bars show standard error of the mean

5.2 Error Rate

Fig. 7 shows the error rate measured as Minimum String Distance (MSD), Uncorrected Errors (ErrUC), and Corrected Errors (ErrC). Text entry method was found to have a significant effect on ErrC, $F(2, 10) = 9.14$, $p < .01$, but not on MSD, $F(2, 10) = 1.52$, $p = .27$, or ErrUC, $F(2, 10) = 1.46$, $p = .28$. Pairwise comparisons showed more corrected errors with MultiTap (*M* = 5.7%, *SD* = 6.6 %) than with Projected QWERTY (*M* = 2.8%, *SD* = 4.6 %), $p < .05$. No significant different were found between MultiTap and H4 Mid-Air (*M* = 4.0%, *SD* = 4.9 %), $p = .35$, or between H4 Mid-Air and Projected QWERTY, $p = .20$. Session was also found to have a significant effect on ErrC, $F(2.175, 10.874) = 4.19$, $p < .05$, but not on MSD, $F(1.853, 9.264) = 2.205$, $p = .17$, or ErrUC, $F(1.922, 9.609) = 1.962$, $p > .19$. Overall, ErrC declines over sessions, which was expected because participants make fewer errors as they become increasingly familiar with a text entry method.

5.3 Subjective Measures

Fig. 8 shows the SMEQ scores across input methods and sessions; recall that lower SMEQ scores represent lower mental effort. Mental effort differs significantly across input methods, $F(2,10) = 6.25$, $p < .05$. Pairwise comparisons showed that mental

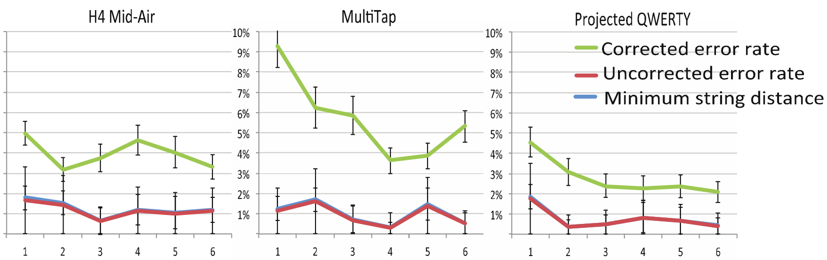


Fig. 7. Error rates (mean) over sessions and text entry method. Error bars show standard error of the mean.

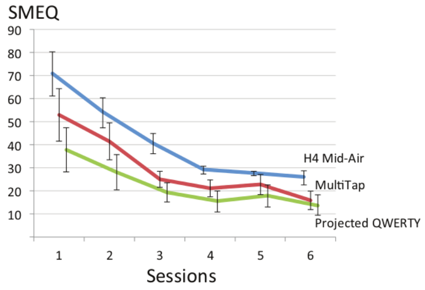


Fig. 8. Mean SMEQ scores over sessions for the three text entry methods. Error bars show standard error of the mean.

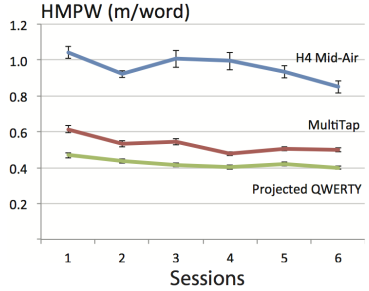


Fig. 9. Mean HMPW for the three text entry methods over sessions. Error bars show standard error of the mean.

effort with Projected QWERTY is lower than with H4 Mid-Air. In the final session, MultiTap and Projected QWERTY lead to comparable mental effort. As expected, we found a significant effect of session, $F(5, 25) = 17.19$, $p < .01$. In particular H4 Mid-Air improved; from session 1 to 6, its SMEQ ratings decreased with 63%.

Overall, we found a significant effect of input method on the 12 satisfaction questions, $F(2, 10) = 5.63$, $p < .05$; session and interaction between input method and session were not significant. Key factors in the main effect of input method were perceived difficulty of use (H4 Mid-Air is perceived as more difficult to use and more uncomfortable to use) and mental effort (H4 Mid-Air is perceived as requiring more effort).

Participants' ranking of input methods showed a clear pattern. In the first session, five participants ranked Projected QWERTY first, four ranked MultiTap second, and four ranked H4 Mid-Air last. This is a significant difference, $\chi 2$ (4, N = 6) = 12.5, $p < .05$. This pattern changed only minimally in the last session, where just one participant changed ranking, giving Projected QWERTY first for four participants (H4 Mid-Air first for two). The ranking was still significant, incidentally with the same $\chi 2$ score.

Participants' comments after each session and in a comparison of methods after session 6 support the above data. Generally, users commented that H4 Mid-Air was hard to use and tired them. Two participants, however, noted that H4 Mid-Air was fun and made them think of typing as a game.

5.4 Hand Movement

Fig. 9 shows average hand movement (HMPW) for the text entry methods across sessions. We found a main effect for text entry method, $F(1.018, 5.092) = 31.95$, $p < .01$, but not for session, $F(2.339, 11.696) = 3.020$, $p = .08$. Pairwise comparisons showed more hand movement for H4 Mid-Air ($M = 0.96$ m, $SD = 0.28$ m) than for MultiTap ($M = 0.53$ m, $SD = 0.12$ m), which in turn was more than for Projected QWERTY ($M = 0.42$ m, $SD = 0.09$ m), all with $p < .05$.

The three methods differ in their use of motor space. Projected QWERTY uses absolute mapping of all buttons into rectangular areas in motor space; H4 Mid-Air

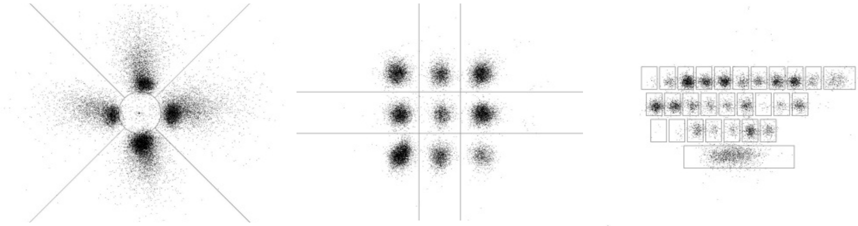


Fig. 10. Distribution of taps in motor space for each text entry method: H4 Mid-Air (left), MultiTap (middle), Projected QWERTY (right). Motor space proportions are identical for the interfaces. Activation zones are indicated by gray lines.

divides the motor space into four slices surrounding a 4cm dead area; MultiTap separates the motor space into a central 4cm x 4cm square activation zone, surrounded by eight infinite activation zones.

Fig. 10 shows a plot of all taps for the three text entry methods. Even though H4 Mid-Air and MultiTap both have open-ended activation spaces, they used motor space differently. For H4 Mid-Air, the shape of the tap-cloud indicate that users are aiming for the activation space close to the center, but often overshooting takes place, resulting in taps in the parts of the activation area further away from the center. Perhaps users may be relying less on the visible feedback of the radar and more on proprioception. However, tap-clouds for MultiTap indicate that participants targeted the center of the keys in the on-screen keyboard. We hypothesize that this difference is primarily related to the visual design: using buttons may discourage the use of open-ended activation areas. This suggests that the radar feedback of H4 Mid-Air facilitates fast, but inaccurate movements that could potentially be based on muscle memory rather than visual feedback. In contrast, the button-based design of the MultiTap keyboard may motivate users to perform accurate pointing and tapping rather than quick movements based on muscle memory.

During the experiment, several participants commented that the movement required could be reduced for H4 Mid-Air and MultiTap without loss of text entry performance. In Fig. 10, we see that the tap-clouds for both H4 Mid-Air and MultiTap are clearly separated by areas with few taps. This is less pronounced for Projected QWERTY. It seems that taps on the keys of Projected QWERTY occurs everywhere on the keys. Taking the low error rate and relatively good performance of Projected QWERTY into account, we see no reason to suspect that the motor space of Projected QWERTY is too small for participants to tap. Rather, this pattern could indicate that the button size in Projected QWERTY is approaching a lower limit for the current tracking precision; using this technique with lower-precision equipment (e.g., Kinect) could result in poor performance. Instead, we hypothesize that a reduction of the motor space movement required for H4 Mid-Air and MultiTap could reduce HMPW and potentially improve text entry performance. We do however note that a reduction of motor space would potentially impact the level of visibility-independence of the text entry methods.

6 Discussions

We first discuss our results in relation to design options and criteria in mid-air text entry. Then we discuss potential improvements to each of the methods.

6.1 Results and Design Space

The criteria that the methods were designed for impacts their performance. First, MultiTap and H4 Mid-Air, both designed to work without visual feedback, perform significantly worse than Projected QWERTY in terms of speed; MultiTap also has a higher error rate. This suggests a trade-off between visibility-dependence and performance. However, further empirical studies are needed to actually show whether skilled users can use MultiTap and H4 Mid-Air without feedback. Surprisingly, we did not see any benefit from the open-ended activation areas of MultiTap. On the contrary, users seemed to perform accurate pointing and tapping within the motor space of the keys.

Second, the number of buttons is likely to affect the results. MultiTap had fewer but larger buttons (in motor space) than Projected QWERTY, which might explain why more hand movements were found for MultiTap. We would expect an increase in performance with MultiTap if the activation areas were smaller. However, this could have a detrimental effect on the method's visibility-dependence. Given the wide distribution of taps with H4 Mid-Air (see Fig. 10), we hesitate to make similar speculations about reducing the activation areas for H4 Mid-Air.

6.2 Improving the Methods

The H4 Mid-Air technique did not work well in the present study. It achieved only an average of 5.2 WPM in the last session, compared to 20.4 WPM in the paper describing H4 Writer [21], and 14.0 WPM in a glove based study [2]. It is worth noting, however, that the number of sessions and transcribed phrases per participant in our study were significantly lower than in these two studies. Users' satisfaction was the lowest for H4 Mid-Air among the methods we explored, though SMEQ scores dropped by 63%.

In our view, H4 Mid-Air performance may be improved in several ways. First, the distribution of tap points is elliptical (rather than circular as for the other methods). Thus, users moved their hands much more than they had to, as also shown by the HMPW measure. Second, the original H4 Writer used Huffman coding on four possible choices because a four-button device was used. We can do a HX Mid-Air, where X is the number of discrete zones that the user can actuate in mid-air; it is not clear that four is the right number. For instance, Fig. 10 suggests plenty of space to do H8, resulting in reduced input zones in motor space and shorter Huffman codes for each character. Third, tapping was used to write a character, but as mentioned in the section on design space, many other options exist (e.g., pinching, using depth). Fourth, we reiterate that the feedback method for H4 Mid-Air was designed to facilitate walk-up-and-use. We have not compared it to the feedback in the H4 Writer system [21].

The MultiTap method performed quite good, achieving an average of 9.5 WPM on the last session; almost identical to the 10 WPM performance of typical MultiTap implementations for mobile phones (without text prediction) [24]. One way to improve MultiTap is to minimize the time spent on timeouts between taps; earlier studies have attempted to do this by using an extra button to skip the timeout. However, adding buttons to the MultiTap interface would reduce some of the potential benefits with regards to visibility-independence. Other improvements that do not impact visibility-independence could be the use of bimanual interaction or depth information, even though that was found too complicated with an accelerometer in GesText [18]. Interestingly, MultiTap seems to be successful in generating a feeling of buttons-in-the-air, which means that users make less movement to hit an area (in contrast, tap distributions in H4 Mid-Air were elliptical). As previously mentioned, a reduction in used motor space could also result in performance improvements for MultiTap.

The Projected QWERTY method achieved high text entry rates (13.2 WPM in the final session) and was the preferred system among users. We have a few ideas for further improving this technique. First, the size of the motor space for controlling Projected QWERTY was determined through pilot studies, but it might be further optimized. Second, Projected QWERTY could easily be extended to support input for two hands, which could dramatically increase performance.

6.3 Next Steps

Our results indicate that the adaptions of successful text entry methods from other domains are indeed possible, and that some of these adapted techniques can provide acceptable text entry speeds. This paper provides a performance baseline of how a set of adapted text entry methods from other domains may perform in mid-air. Based on the experiences from our experiments, we agree with Shoemaker et al. [33] that the terms distance- and visibility-dependence describe important properties of mid-air text entry well, and we see a need to further study how aspects of the mid-air design space affect these properties.

We are aware that the availability of 6-DoF and precise tracking could potentially open up for new and innovative text entry methods. The guidelines for designing character-entry systems in 3D user interfaces presented by Bowman et al. [7] combined with detailed empirically based analyses of the mid-air design space, such as that attempted in the present paper, provides starting points for future research.

Acknowledgements. This work has been supported in part by the Danish Council for Strategic Research, grant 10-092316.

References

1. Bailly, G., Vo, D.-B., Lecolinet, E., Guiard, Y.: Gesture-aware remote controls. In: Proc. ICMI 2011, pp. 263–270. ACM, New York (2010)
2. Bajer, B., MacKenzie, I.S., Baljko, M.: Huffman Base-4 Text Entry Glove (H4 TEG). In: Proc. ISWC 2012, pp. 41–47. IEEE (2012)

3. Ball, R., North, C., Bowman, D.A.: Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In: Proc. CHI 2007, pp. 191–200. ACM, New York (2007)
4. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. Communications of the ACM 36(7), 28–35 (1993)
5. Baudisch, P., Cutrell, E., Robbins, D., et al.: Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch- and Pen-operated Systems. In: Proc. Interact 2003, pp. 57–64 (2003)
6. Benko, H.: Beyond flat surface computing: challenges of depth-aware and curved interfaces. In: Proc. MM 2009, pp. 935–944. ACM, New York (2009)
7. Bowman, D.A.: 3D user interfaces: theory and practice. Addison-Wesley (2005)
8. Butts, L., Cockburn, A.: An evaluation of mobile phone text input methods. In: Proc. AUIC 2002, vol. 7, pp. 55–59. IEEE, Los Alamitos (2002)
9. Castellucci, S.J., MacKenzie, I.S.: Unigest: text entry using three degrees of motion. In: Proc. CHI 2008, pp. 3549–3554. ACM, New York (2008)
10. Castellucci, S.J., MacKenzie, I.S.: Graffiti vs. unistrokes: an empirical comparison. In: Proc. CHI 2008, pp. 305–308. ACM, New York (2008)
11. Czerwinski, M., Smith, G., Regan, T., et al.: Toward characterizing the productivity benefits of very large displays. In: Proc. Interact 2003, pp. 17–24 (2003)
12. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica 10(2), 112–122 (1973)
13. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In: Proc. CHI 1999, pp. 215–222. ACM, New York (1999)
14. Forlines, C., Vogel, D., Balakrishnan, R.: HybridPointing: fluid switching between absolute and relative pointing with a direct input device. In: Proc. UIST 2006, pp. 211–220. ACM, New York (2006)
15. Gustafson, S., Bierwirth, D., Baudisch, P.: Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback. In: Proc. UIST 2010, pp. 3–12. ACM, New York (2010)
16. Hilliges, O., Izadi, S., Wilson, A.D., Hodges, S., Garcia-Mendoza, A., Butz, A.: Interactions in the air: adding further depth to interactive tabletops. In: Proc. UIST 2009, pp. 139–148. ACM, New York (2009)
17. Hinckley, K., Wigdor, D.: The Human-Computer Interaction Handbook, pp. 95–132. CRC Press (2012)
18. Jones, E., Alexander, J., Andreou, A., Irani, P., Subramanian, S.: GesText: accelerometer-based gestural text-entry systems. In: Proc. CHI 2010, pp. 2173–2182. ACM, New York (2010)
19. Khan, A., Fitzmaurice, G., Almeida, D., Burtnyk, N., Kurtenbach, G.: A remote control interface for large displays. In: Proc. UIST 2004, pp. 127–136. ACM, New York (2004)
20. Kristensson, P.O., Nicholson, T., Quigley, A.: Continuous recognition of one-handed and two-handed gestures using 3D full-body motion tracking sensors. In: IUI 2012, pp. 89–92. ACM, New York (2012)
21. MacKenzie, I.S., Soukoreff, R.W., Helga, J.: 1 thumb, 4 buttons, 20 words per minute: design and evaluation of H4-writer. In: Proc. UIST 2011, pp. 471–480. ACM, New York (2011)
22. MacKenzie, I.S., Soukoreff, R.W.: Text Entry for Mobile Computing: Models and Methods, Theory and Practice. Human–Computer Interaction 17(2), 147–198 (2002)

23. MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: Proc. CHI 2003, pp. 754–755. ACM, New York (2003)
24. MacKenzie, I.S., Tanaka-Ishii, K.: Text Entry Systems: Mobility, Accessibility, Universality. Morgan Kaufman (2007)
25. MacLean, A., Young, R., Bellotti, V., Moran, T.P.: Questions, Options, and Criteria: Elements of Design Space Analysis. Human–Computer Interaction 6(3), 201–250 (1991)
26. Makin, T.R., Holmes, N.P., Ehrsson, H.H.: On the other hand: dummy hands and peripersonal space. Behavioural Brain Research 191(1), 1–10 (2008)
27. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., Mackay, W.: Mid-air pan-and-zoom on wall-sized displays. In: Proc. CHI 2011, pp. 177–186. ACM, New York (2011)
28. Ni, T., Bowman, D., North, C.: AirStroke: bringing unistroke text entry to freehand gesture interfaces. In: Proc. CHI 2011, pp. 2473–2476. ACM, New York (2011)
29. North, C., Dwyer, T., Lee, B., Fisher, D., Isenberg, P., Robertson, G., Inkpen, K.: Understanding Multi-touch Manipulation for Surface Computing. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 236–249. Springer, Heidelberg (2009)
30. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-second bursts. In: Proc. CHI 2005, pp. 919–928. ACM, New York (2005)
31. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. In: Proc. CHI 2010, pp. 887–896. ACM, New York (2010)
32. Sauro, J., Dumas, J.S.: Comparison of three one-question, post-task usability questionnaires. In: Proc. CHI 2009, pp. 887–896. ACM, New York (2009)
33. Shoemaker, G., Findlater, L., Dawson, J.Q., Booth, K.S.: Mid-air text input techniques for very large wall displays. In: Proc. GI 2009, pp. 231–238. Canadian Information Processing Society, Toronto (2009)
34. Soukoreff, R.W., MacKenzie, I.S.: Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified. In: Proc. CHI 2003, pp. 113–120. ACM, New York (2003)
35. Vogel, D., Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays. In: Proc. UIST 2005, pp. 33–42. ACM, New York (2005)
36. Wilson, A.D.: Robust computer vision-based detection of pinching for one and two-handed gesture input. In: Proc. UIST 2006, pp. 255–258. ACM, New York (2006)
37. 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)
38. Wobbrock, J.O., Myers, B.A.: Analyzing the input stream for character- level errors in unconstrained text entry evaluations. TOCHI 13(4), 458–489 (2006)
39. Zhai, S., Kristensson, P.-O., Smith, B.A.: In search of effective text input interfaces for off the desktop computing. Interacting with Computers 17(3), 229–250 (2005)

Evaluating Multivariate Visualizations as Multi-objective Decision Aids

Meirav Taieb-Maimon[1], Lior Limonad[2], David Amid[2], David Boaz[2], and Ateret Anaby-Tavor[2]

[1] Information Systems Engineering, Ben-Gurion University, Israel
meiravta@bgu.ac.il
[2] IBM Research – Haifa, Israel
{liorli,davida,davidbo,atereta}@il.ibm.com

Abstract. When choosing a solution, decision makers are often required to account for multiple conflicting objectives. This is a situation that can result in a potentially huge number of candidate solutions. Despite the wide selection of multivariate visualizations that can potentially help decide between various candidates, no designated means exist to assess the effectiveness of these visualizations under different circumstances. As a first contribution in this work, we developed a method to evaluate different types of multivariate visualization. The method focuses on the visualization's ability to facilitate a better understanding of inter-objective trade-offs as a proxy to more sensible decision making. We used the method to evaluate two existing visualization aids: Parallel-Coordinates and an adaptation of Self Organizing Maps (SOM). Both visualizations were compared with tabular data presentation. Our results show that the first visualization is more effective than a plain tabular visualization for the purpose of multi-objective decision making.

Keywords: Multi-criterion decision making, Multivariate visualizations, Information Visualization, Usage experience evaluation.

1 Introduction

There are many circumstances where a person has to decide among competing alternatives. From a consumer trying to choose the right car, house, or cell phone to a business executive who must decide upon a new portfolio of product offerings, and on to elected representative voting on national health policies—all of them must choose among multiple, often competing, alternatives. Many studies in the area of *multi-criterion decision making* (MCDM) attempt to help decision-makers reach better conclusions in a more efficient way [18]. The multi-criterion decision making process typically examines various alternatives with respect to their values for each criterion. In most practical scenarios, the number of options is too large to be examined by a human, and decision makers aspire to examine a limited set of options. Towards this end, the MCDM research community has identified two major challenges: (a) reducing the number of options by means of an optimization process that yields a smaller

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 419–436, 2013.

set of optimal solutions, termed Pareto Frontier [6], and (b) effectively visualizing these solutions (namely, the solution space) to help users select the solutions that best satisfy their subjective criteria. The task of visualizing the Pareto Frontier is based upon multivariate visualization, and is generally considered a hard problem for more than three objectives [19].

The decision making process becomes even more challenging when some of the objectives being considered conflict with each other. In this conflicting relationship, the performance in one objective is seen to deteriorate as performance in another is improved. In such settings, actual choices should ideally reflect the tradeoffs that are in keeping with the decision maker's priorities. For example, given a certain budget, a person who travels frequently and wants to buy a new laptop would tend to give weight a higher priority over performance, while a "gamer" would aim for increased performance and settle for a heavier weight.

We distinguish between *sensible* and *non-sensible* choices. Sensible choices faithfully correspond to the decision-maker's subjective perception of inter-objective trade-offs and non-sensible choices do not. According to Lotov et al. [22], "Tradeoff information is extremely important for the decision maker since it helps to identify the most preferred point along the tradeoff curve." Hence, it is expected that the better the decision-maker's understanding of the various tradeoffs in the solution space, the more sensible their choices will be. Correspondingly, we introduce the following as an underlying premise:

Assumption 1: A better understanding of the inter-objective conflicts results in more sensible decision making.

Derived from the above assumption, our main objective in this work is *to provide decision makers with a concrete means to improve the understanding of inter-objective conflicts*.

Ideally, identifying a sensible choice should rely on first discovering the perceived trade-offs the decision maker may have with regard to a given set of objectives [10, 11, 26]. Pragmatically, extracting even a single trade-off function seems to be highly cumbersome, and in some cases very inaccurate [12]. Furthermore, in some cases, the decision maker may become aware of subjective preferences only after confronting the actual solution space [12].

Since a purely analytic approach is deemed unfeasible, we chose to focus on the area of 'visual analytics,' in which visual interfaces are used to facilitate interactive reasoning [14]. Having decision makers closely involved in data exploration and processing cycles eliminates the need to extract and formalize their subjective preferences. Despite the plethora of multivariate visualization aids that exist in this domain (e.g., SOM [25], Interactive Decision Maps [21], Parallel Coordinates (PC) [13]), we found no prior work on user-experience evaluation to direct our selection of a designated visual interface that would facilitate multi-objective decision making. As a result, we decided to investigate and test the effectiveness of existing visual interfaces in promoting the selection of sensible choices. The evaluation of information visualization is an area in which the development of designated metrics and benchmarks has been identified as invaluable [28]. Our work is also unique in tackling the special circumstances in which any choice made is deemed acceptable since all competing alternatives are quantitatively equivalent (all being on the pareto frontier). It is the sole (unknown) preference of each decision maker that can determine the subjective ranking of the solution space.

We consider our contribution as twofold: (1) providing a pragmatic method and a carefully adapted measurement scale for testing the effectiveness and usage-experience of multivariate visualization aids for MCDM, and (2) providing a first glance into the actual performance of two existing visualization techniques for promoting effective MCDM. In the following sections we present the preliminary process of adopting and refining our experimental instruments, followed by a complete report of our experimental settings and results.

2 The Effectiveness of Visualization Aids

The presented work is driven by the growing desire evident in the HCI literature for the development of alternative methods to evaluate visualization and encourage more widespread adoption of visualization [31]. Specifically, we focused on the unique nature of interaction with multivariate information visualization to more effectively benefit from and measure insights discovered [30] rather than extracting efficiency measures such as time to complete a task. In our experimental configuration, we did include a measurement of task completion time for which no significant differences had been found across the different visualization types[1]. Hence, we considered effectiveness according to the following proposition:

Proposition 1: A (multi-dimensional) *visual representation* of the solution space promotes better *decision making sensibility* than a *plain representation* of the solution space.

For the purpose of a *visual representation*, we used two alternative techniques: Parallel Coordinates (PC), which is considered the prevalent technique for the visualization of multi-objective data [1, 7, 32], and Self-Organizing Map for Multi Objective (SOMMOS), a recent adaptation of a Self-Organizing Map (SOM) [17] visualization being developed by IBM as a proprietary visualization for exploring and visualizing a Pareto Frontier.

Figure 1 illustrates an example for the two visualizations, showing three different apartments being considered for rent. Each alternative is measured based on three objectives: number of roommates, price, and distance from the university. On the left hand side, the PC visualization is depicted with each axis assigned a corresponding objective. Along each axis, an arrow indicates the aspiration of the decision maker to either minimize or maximize the specific objective. In Figure 1, the decision maker's goal is to minimize all three objectives. Each choice is depicted as a line that intersects every axis in accordance with the associated objective's value. The decision maker can observe the extent of conflict between objectives by searching for a cross (an 'x' pattern) between axes. For instance, based on the cross pattern apparent

[1] $p=.186$ and $p=.144$ for the two experimental tasks RENT and TAM.

between the ‘Roommates’ axis and the ‘Price’ axis, the decision maker can infer a strong conflict between the corresponding objectives. In contrast, by examining the lines between the ‘Price’ axis and the ‘Distance from University’ axis, the decision maker can conclude that there is no conflict between the corresponding objectives based on the absence of any crossing.

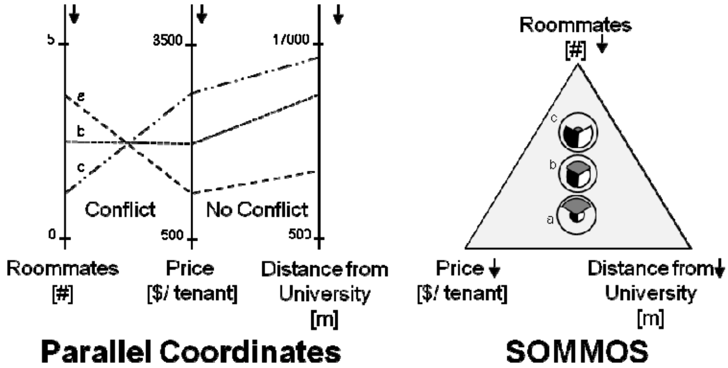


Fig. 1. The PC and SOMMOS visual representations evaluated

On the right hand side, the same solution space is depicted using SOMMOS. SOMMOS is a variation of the well-known self-organizing map (SOM) algorithm [17] adapted for visualizing Pareto Frontiers. The SOMMOS visualization comprises three pillars: a map layout, a visual representation of individual choices, and interactive capabilities. The map layout enables the user to explore and draw insights based on its structure. The objectives are represented by the vertices of the map. For example, the three objectives in Figure 1 form a triangular layout. One of SOMMOS' virtues is the ability to navigate through the map. Specifically, the decision maker can expect that the closer the choice is to a vertex of an objective, the closer its value is towards the goal of that objective. Each choice on the map is encoded using polar area chart glyphs. Each objective is assigned a corresponding colored slice. By noting the level of colored filling, decision makers are able to understand the nature of the compromise in an area, spot an area of choices that maximize\minimize one of the objectives at the expense of the other, and identify cases in which objectives are correlative. In Figure 1, the decision maker can clearly identify that the Roommates objective is in conflict with both Price and Distance from University such that as the slice filling for Roommates decreases, the slice filling for Price and Distance increases. The decision maker can also observe that across all choices, the filling pattern for both Distance from University and Price is consistent. Hence, there is no conflict between the two.

In addition to testing each of the visualizations, we suspected that the two visualization types may complement one another. Because each has its unique strengths and weaknesses, allowing decision makers to simultaneously interact with both could potentially yield better results. Nevertheless, it was unclear whether the cognitive effort associated with toggling between two visualizations would impair the potential impact of combining the two. As a result, our experimental settings included a visual interface that combined PC and SOMMOS.

We used a simple table for the purpose of a *plain representation*, being synonymously referred to as the “no visualization” in this work. We structured the table with columns designating objectives and rows designating possible choices. Table 1 shows a plain representation of the same data underlying the solution space illustrated in Figure 1.

Table 1. Plain representation of the solution space in Figure 1

	Roommates (#)	Price ($/tenant)	Distance from University (m)
a	4	400	700
b	3	600	1500
c	2	800	2300

As an antecedent to the notion of *decision making sensibility*, we developed a designated scale (see next section) to measure *subjective conflict understanding*. This helped us adhere to Assumption 1, according to which sensible choices rely first and foremost on the decision maker's ability to understand the nature of trade-offs between the objectives.

Correspondingly our experimental hypothesis is as follows:

H1: A (multi-dimensional) *visual representation* of the solution space promotes better *subjective conflict understanding* than a *plain representation*.

In addition to the measures that relate directly to the above hypothesis, we considered several background factors that may interfere with the effectiveness of the visualization being used. We extended our research model as illustrated in Figure 2 to consider three types of moderating factors:

Subjective usage experience – Several measures were employed to reflect the way users feel about using the visualizations. Such experience may be a potential moderator to the main effect hypothesized above. However, the secondary effect of each of the visualizations on usage experience was in itself unclear. Hence, this led to a second hypothesis:

H2: A (multi-dimensional) *visual representation* of the solution space promotes better *usage experience* than a *plain representation*.

We adopted a set of concrete preliminary measures from previous literature and further refined them as part of a card sorting procedure, explained in the next section.

Problem characteristics – The particular *problem domain* and *complexity* may also be a factor affecting the degree to which the visualization means can facilitate better objective trade-off understanding. The complexity, for example, may affect its capacity to scale in size of data [33]. As a result, we decided to manipulate these factors to account for their influences.

Participant factors – As noted by Gemino et al. [9], individual factors can affect the effectiveness of a visualization. To account for potential interference, we included questions to measure participants' familiarity with the corresponding problem domain and their experience. The latter included general questions about experience using visual analytics and specific questions regarding experience with multi-objective decision making.

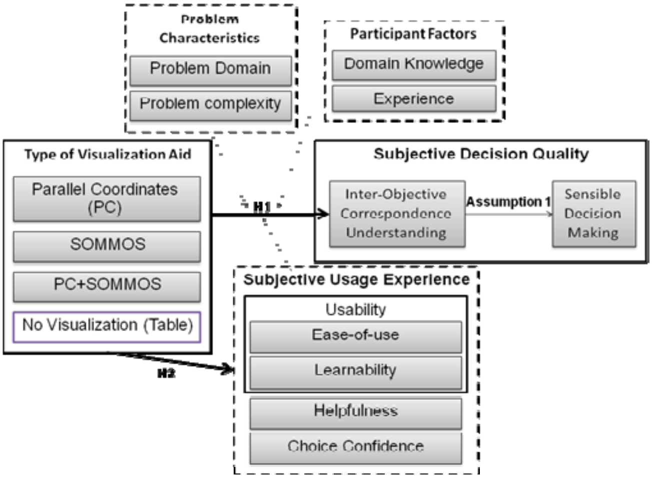


Fig. 2. Extended research model

3 Scale Development

Aside from the main effectiveness construct, all other scales have been originally adopted from prior literature having solid foundations in the HCI literature. To ensure scale validity and reliability in the operationalization of our research model, we produced a preliminary set of questions based on existing scales, followed by a *card sorting* procedure [3, 4, 24]. This led us to the scale items detailed in Table 5. Card sorting[2] is an iterative scale development technique in which a panel of judges is asked in several rounds to sort a set of scale items into separate categories, based on similarities and differences among them. This technique was previously used by Davis et al. [3, 4] to assess the coverage of an intended domain of constructs. Later, it was further refined [24] to generally account for scale reliability (i.e., content and construct validity). Prior to conducting the first card sorting round, we populated items for each of the constructs as described next.

As mentioned in the preceding section, we operationalized the high level notion of ***subjective decision quality*** through the construct of **inter-objective correspondence understanding** (i.e., as an antecedent to sensible decision making). This construct is both innovative to this work and related to the particular decision problem domain

[2] We used OptimalSort by OptimalWorkshop.com as a platform for card sorting.

being explored. As such, no prior measurement scales were available for adoption. We produced a set of seven items based on our experience with the notion of inter-objective trade-offs. Specifically, all items were generated in accordance with the general definition by Purshouse et al. [29] according to which inter-objective trade-off understanding is defined as the degree to which an individual believes he or she were able to understand the 'conflict' between the objective. The 'conflict' is interpreted as a relationship between any two objectives in which as performance in one is improved, performance in the other is seen to deteriorate. Each related question was worded such that its most correct answer corresponds to one of the extreme ends in a 1-to-9 Likert scale [23].

To operationalize the high level notion of *subjective usage experience* associated with the decision making task, we adopted a preliminary set of post-test concrete constructs from prior literature as follows:

(1) **Usability** – being conventionally used to capture the general appropriateness of an artifact to a particular purpose [2]. In our study, the artifact being evaluated corresponds to the various visualization interfaces mentioned earlier. The concrete purpose corresponds to the facilitation of multi-objective decision making. We refined the common 10-item SUS scale [2] to match such circumstances. In addition, we eliminated two questions with very similar phrasing to two of the other questions that were intended to measure ease-of-use.

(2) **Ease-of-use** – adopted from Moore and Benbasat [24], originally aimed at measuring the degree to which users believe that using a particular system would be free of physical and mental effort. Correspondingly, we adopted a preliminary set of six items and slightly re-worded the phrasing to match the intended usage purpose.

(3) **Cognitive-load** – the actual cognitive capacity required for the decision making task [27]. The measurement of the actual cognitive capacity is typically considered a multi-dimensional construct. In our study we adopted a corresponding measurement scale from DeLeeuw and Mayer [5], to measure two items reflecting: (a) **mental-effort** – the self reported effort subjects indicate as required to reach a certain decision, and (b) **task-difficulty** – the self reported difficulty subjects indicate as being associated with the decision making task. Stemmed from the basic notion of task difficulty, and the particular nature of the task in this work, we added two additional items. These items take into account the capacity of the visualization to assist in the recognition of similar choices (i.e., clusters), and its capacity to gradually narrow down the set of preferred choices until a single choice is identified.

(4) **Decision-confidence** – in its broader sense reflects the subject's confidence in the certainty of estimates or predictions of future performance based on their perceptions of their knowledge and experience. Adopted from the work by Kidwell et al. [16], this construct was measured by a unitary item which we slightly re-worded to reflect the confidence between the choice that is made and subjective preferences.

We operationalized the two remaining construct groups: *problem characteristics* and *participant factors* in a more straightforward manner. We incorporated problem characteristic factors into the experimental design by manipulating the problem's domain and the problem's complexity. Its exact levels of manipulation are further explained in the following section. We incorporated participant factors by including three questions on background knowledge: familiarity with the problem domain, previous experience using visual-analytics interfaces, and previous experience in solving multi-objective problems.

The set of all items being produced for the measurement of the main effect (i.e., H1), together with all other items populated for the measurement of the resulting user experience (i.e., H2), were jointly considered as the input pool for the first card sorting round. We fragmented this preliminary set as illustrated in Table 2.

Table 2. Preliminary construct-wise item pool for card-sorting input

Conflict understanding	7	Usability	8	Ease of use	6
Mental effort	1	Task difficulty	3	Decision confidence	1

To eliminate ambiguous items, we conducted three separate card sorting rounds with 21 independent judges. As noted by Moore et al. [24], this was aimed at increasing construct validity and our confidence in the developed scale. The procedure was an ‘open card sort’ style, implying that the judges in the first round were not told what the underlying constructs (i.e., item categories) were, rather they were asked to provide their own construct labels and definitions. We used three indicators to identify construct validity and the convergence of the sorting procedure: (1) Inter-rater agreement between the judges for each construct in each round, (2) Hit ratio (“convergent” validity) – item-wise portion of true item classifications within its “target” construct, and (3) Missed ratio (“discriminant” validity) – construct-wise portion of wrong item classification within each construct. In each round, a “doesn’t fit” category was also permitted for items that did not match any of the other clusters. A summary of results is illustrated in Table 3.

After the first sorting round, we observed a significant cross loading between ease-of-use and usability. As noted above, this can be traced back to [20] having the SUS scale being composed of two sub-factors to measure usability and learn-ability, the former having conceptual overlap with ease-of-use. In addition, two items had poor loading on either ease-of-use or usability. We dropped these items and split the remaining items as being measures of either ease-of-use or learn-ability, considering usability as the combination of both (i.e., a second degree construct). **Learn-ability** is defined as the degree to which the tool enables the user to learn how to use it.

Table 3. Card sorting results

	Round 1 (8 judges)				Round 2 (6 judges)				Round 3 (7 judges)			
	# items	Inter-rater agree	Hit Ratio	Missed Ratio	# items	Inter-rater agree	Hit Ratio	Missed Ratio	# items	Inter-rater agree	Hit Ratio	Missed Ratio
Conflict Understanding	7	0.75	1.00	0.05	7	0.76	0.76	0	7	0.71	0.8	0.03
Usability	8	0.61	0.28	0.50	-	-	-	-	-	-	-	-
Ease of use	6	0.55	0.50	0.59	8	0.57	0.75	0.25	7	0.51	0.88	0.14
Learnability	-	-	-	-	4	0.46	0.75	0.14	3	0.67	0.86	0.1
Mental effort	1	0.67	0.25	0.50	-	-	-	-	-	-	-	-
Task difficulty	3	0.52	0.71	0.41	-	-	-	-	-	-	-	-
Helfulness	-	-	-	-	3	0.38	0.55	0.57	3	0.5	0.86	0.14
Decision Confidence	1	1.00	0.25	0.00	-	-	-	-	-	-	-	-
Choice confidence	-	-	-	-	1	0.33	0.67	0.67	1	0.24	0.86	0.45
Doesn't fit		0.42	-	-		1	-	-		None	-	-
Average		**0.68**	**0.50**	**0.34**		**0.50**	**0.70**	**0.33**		**0.53**	**0.85**	**0.17**

The single item measuring mental-effort had a significantly stronger load on ease-of-use than on itself. This was also explained by the definition of ease-of-use being: "the degree to which an individual believes that using a particular system would be free of physical and mental effort". Hence, we decided to drop the entire construct, since it was sufficiently captured in the measurements for ease-of-use.

Task-difficulty had very poor inter-rater agreement and inconsistent definitions across the judges. Consequently, we concluded that the actual construct being measured should be labeled as **decision-making helpfulness.** This captured the extent to which the visualization aid helped in the decision making process.

Decision-confidence also had very low hit ratio. Hence, we reworded the item to better reflect the certainty of choice made based on the perception of one's own knowledge and experience. We labeled this construct as **choice confidence**. This concluded all modifications made according to the results of the first sorting round.

After the second card sorting round, we dropped two additional items due to loading ambiguities. One was originally intended to measure ease-of-use but appeared to load equally on choice-confidence. The second item was originally intended to measure learn-ability and appeared to load equally on ease-of-use. To further increase item loading, we reworded two additional items (7 and 20) to better reflect their intended constructs. This concluded the modifications for the second sorting round.

Finally, after the third sorting round, the only apparent problem was a relatively low discriminant validity for choice confidence. Due to its high convergent validity, and no consistent classification to any of the items wrongly classified as its potential measures, no further actions seemed necessary. All remaining constructs demonstrated reasonable inter-rater agreements, alongside good convergent and discriminant validities. Hence, we decided to keep all items unchanged and concluded sorting rounds with the final loadings as illustrated in Table 4. A summary of all scale items is depicted in Table 5. The developed scale is our first contribution in this paper and can be used to assess any multivariate visualization aimed to promote MCDM.

Table 4. Item loading matrix after the sorting procedure

	Choice confidence	Conflict understanding	Decision making helpfulness	Ease of use	Learn-ability
EOU 1				100.00%	
EOU 2			14.29%	71.43%	14.29%
EOU 3	14.29%			85.71%	
EOU 4				85.71%	
EOU 5				100.00%	
EOU 6				100.00%	
EOU 7				71.43%	14.29%
LBT 8				14.29%	85.71%
LBT 9				14.29%	85.71%
LBT 10				14.29%	85.71%
CON 11		85.71%		14.29%	
CON 12		85.71%		14.29%	
CON 13		85.71%	14.29%		
CON 14	14.29%	71.43%		14.29%	
CON 15	14.29%	71.43%		14.29%	
CON 16	14.29%	85.71%			
CON 17	14.29%	71.43%	14.29%		
HLP 18			85.71%		
HLP 19	14.29%		85.71%		
HLP 20		14.29%	85.71%		
CNF 21	85.71%				

Table 5. Final items after the sorting procedure (EOU – ease-of-use, LBT – learnability, CON[3] – conflict understanding, HLP – decision-making helpfulness, CNF – choice confidence)

Item	Item Description
EOU1	The visualization aid was cumbersone to use
EOU2	Using the visualization aid required a lot of mental effort
EOU3	using the visualization aid was often frustrating
EOU4	using the visualization aid was clear and understandable
EOU5	Overall, I believe that the visualization aid is easy to use
EOU6	I found the usage of the visualization aid unnecessarily complex
EOU7	I think there was too much inconsistency in the usage of the visualization aid
LBT1	Learning how to use the visualization aid was easy for me
LBT2	I would imagine that most people would learn to use the visualiztion aid very quickly
LBT3	I needed to learn a lot of things before I could start using the visualization aid
CON1	Based on all presented alternatives, an increase in the value of <objective1> requires a decrease in the value of <objective2>
CON2	Based on most presented alternatives, an increase in the value of <objective1> requires a decrease in the value of <objective2>
CON3	In most presented alternatives, the two objectives <objective1> and <objective2> seem to be conflicting
CON4	With regard to the conflict between <objective1> and <objective2>, there are solutions that denote a good compromise between them
CON5	With regard to the conflict between <objective1> and <objective2>, there are solutions that denote a bad compromise between them
CON6	Solutions <1> and <2> are strong indicators to having a conflict between <objective1> and <objective2>
CON7	Solution <1> denotes a good compromise between <objective1> and <objective2>
HLP1	The visualization aid did not help me find the selection solution
HLP2	Using the visualization aid helped me focus the selection to the finding of a preferred solution
HLP3	The visualization aid eased the decision making task be helping me identify groups of solutions that have similar characteristics
CNF1	I have strong confidence that my choice fits my personal preferences

4 Experimental Design

We pursued a 4X2 (visualization type, problem domain) mixed experimental design to evaluate our hypotheses. In the laboratory setting, each participant was randomly assigned to one of four visualization interface groups: Tabular, PC, SOMMOS, PC and SOMMOS. The participants were then asked to make two consecutive choices, each corresponding to a different problem domain.

The first problem domain was in the area of property rental, and the task was to select an apartment for rent (namely, RENT). We selected this domain because it was an area with which most participants are familiar with. The corresponding visualization of the solution space was composed of 20 apartments to be examined according to 3 objectives: number of roommates, price, and distance from university; all these objectives were intended to be minimized.

The second problem domain was in the area of transportation asset management (namely, TAM). We chose this domain because it was an area of less familiarity.

[3] Note that in all CON items, the phrases appearing in brackets were populated by concrete objectives/solutions in the context of the concrete problem domain.

The corresponding visualization of the solution space was composed of 100 possible transportation plans (e.g., a portfolio of transportation projects), to be examined according to 5 objectives: air quality, safety, foreseen economic growth, traffic scarcity, number of pedestrian/cyclist trails; all these objectives were intended to be maximized. The order in which the two problems were presented to each participant was counter balanced across participants.

5 Method

Procedure. The entire experimental procedure took an hour on average. In the beginning of the experiment, each participant was given a 15 minute computer-based training session, which included: a primer to multi-criteria decision making and a video demonstration of the visualization type corresponding to the participant's group. The students were then given a quiz with 10 questions to determine whether they had understood the concepts and the usage of the tool presented in the training. A passing score of 70% was targeted to ensure all participants understood the visualization presented to them and its operation. Following the quiz, each participant was given a questionnaire booklet designed to facilitate the experimental procedure. The booklet had the following sections: participation consent, a set of questions about the participant's background knowledge and experience, and two question sets, corresponding to the two decision-making problems. In each such set, participants were first instructed to use the visualization aid for the purpose of choosing a single alternative out of the presented set, and then record both their selection, and a short explanation justifying their choice. Next, participants were presented with a randomly shuffled set of questions about usage experience and objective trade-off understanding.

Pilot. We carried out a pilot with 30 undergraduate students to further test the procedure and all corresponding experimental materials. Results indicated a significant effect of the visualizations on the perceptions of ease-of-use ($F_{(3,11)}$=3.94, p=.039) and usability ($F_{(3,11)}$=3.76, p=.044) for the RENT problem domain. Although an impact on conflict understanding was not yet apparent, it was decided to go ahead with the main experiment as the sample size in the pilot was fairly limited. We incorporated some minor re-wording after the research team reviewed the pilot responses.

Participants. A group of 93 undergraduate students volunteered to participate in the main experiment, which followed a similar scheme to the pilot. As an incentive for participation, students were offered 2 bonus points in a related academic course. After eliminating students who failed the quiz (i.e., 3 in total who scored below the 70% score threshold), and questionnaires that had been improperly filled out not adhering to the instructions, 172 useful responses remained in total: 85 for the RENT problem domain and 87 for the TAM problem domain. These responses were analyzed and all results are detailed in the following section.

6 Results Analysis

Data Preparation. Prior to running the statistical analysis, we explored the data using BOXPLOT to determine extreme responses. We concluded that 5 responses were outliers, scoring more than 1.5 times the inter-quartile range[4] (i.e., 4 in RENT, and 1 in TAM). Correspondingly, we executed all subsequent analyses twice, with and without the outliers. We found that outlier elimination does not affect any of the significant findings. Hence, final results reporting utilized 80 responses for the rental problem domain, and 85 for the transportation problem domain.

Scale Reliability. Our analysis of the results checked the scale reliability for the measurement of usage experience constructs. Item-wise reliability scores for each construct are illustrated in Table 6, showing that the reliability level indicated by Cronbach's alpha met the suggested tolerance (> 0.7, [23]).

We skipped the reliability check for the scale developed for conflict understanding. Unlike the other constructs that are traditionally considered reflective [8], the items used to measure conflict understanding were not expected to be correlated to one another (i.e., a formative construct). Indeed, this was corroborated by exploratory factor analysis that has been attempted for each problem-domain question set, yielding 3 different principle components associated with each question set in each problem domain. Furthermore, no conceptual grounds could have been attributed to the underlying components, supporting the irrelevance of this test [8]. That is, the inherent content validity of the scale developed for conflict understanding is shaped by how the questions complement each other.

Table 6. Reliability scores (Cronbach's alpha)

Variable	Ease of use	Helpfulness	Learnability	Usability	Confidence
Alpha	0.87	0.71	0.70	0.88	Single-item

6.1 Subjective Decision Quality

We conducted a separate one-way ANOVA for each problem domain to examine the effect of the visualization interface (between groups independent variable) on how well participants understood the relationship between objectives (dependent variable). The results depicted in Table 7 indicate a significant effect of the visualization type only for the case of RENT ($F_{(3,77)}$=2.86, p=.042). We performed the analysis for each problem domain separately since the actual measurements for correspondence understanding (i.e., items in questionnaire) were different between the two problem domains.

---

[4] The inter-quartile range (IQR) contains the middle 50 percent of the distribution. If the data is normally distributed, a range that is 1.5 times the IQR covers ~99.3% of the distribution.

Table 7. Inter-objective correspondence understanding (scores ranged from 1 = least understanding to 9 = highest level of understanding)

	PC			SOMMOS			PC+SOMMOS			Table			ANOVA		
Case	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*df*	*F*	*Sig.*
Rent	7.51	0.72	22	7.06	0.85	22	7.13	0.56	16	6.85	0.82	21	(3,77)	2.86	0.042*
TAM	6.51	0.76	23	6.65	0.97	22	6.41	0.85	20	6.30	0.88	21	(3,82)	0.636	0.594

Post-hocs. We conducted pair-wise comparisons between the visualization interfaces once the significant main effect was apparent for the case of RENT. Consistent with the ANOVA analysis, LSD post-hoc comparisons indicated a significant difference between the PC visualization (*M*=7.51, *SD*=.72) and the plain representation (*M*=6.85, *SD*=.82), p=.005. No further significant pair-wise differences were evident.

6.2 Subjective Usage Experience

We conducted a two-way (visualization type, problem domain) MANOVA to test the possible effects of the visualization type and problem domain on usage experience. In this case, it was possible to integrate problem domains in the analysis as the questions pertaining to all usage experience factors were the same in both problem domains. The two-way interaction (visualization type*problem domain) was not found to be significant for any of the usage experience factors, allowing us to interpret each of the main effects separately.

Descriptive statistics for the effect of visualization type on the usage experience factors are illustrated in Table 9. The main effect of the visualization type on the perception of ease-of-use was significant ($F_{(3,157)}$=2.81, p=.04). Similarly, the effect of the visualization type on the perception of usability was also significant ($F_{(3,157)}$=2.81,p=.04). The effect of the visualization type on the remaining usage experience factors was not found to be significant. However, it might be worth noting a somewhat moderately significant effect of the visualization type on learn-ability ($F_{(3,157)}$, p=.06).

Descriptive statistics for the effect of problem domain on the usage experience factors are illustrated in Table 10. As in the case of visualization type, the main effect of the problem domain was also significant with regard to its impact on the perception of ease-of-use ($F_{(1,157)}$=9.51, p=.002), and on the perception of usability ($F_{(1,157)}$=7.66, p=.006). In addition, choice confidence was also significantly affected by the problem domain ($F_{(1,157)}$=4.64, p=.03). All remaining usage experience factors were not found to be significantly affected by the problem domain.

Post Hocs. We conducted pair-wise comparisons between the visualization interfaces following the conclusion of the significant main effects for ease-of-use and usability. With respect to the effect of the visualization type on ease-of-use, LSD post-hoc comparisons indicated two significant differences: between the PC visualization (*M*=6.13, *SD*=1.63) and the plain representation (*M*=7.06, *SD*=1.36), p=.005, and also between the latter and the combined visualization of PC & SOMMOS (*M*=6.37, *SD*=1.47), p=.048. No further significant pair-wise differences were evident.

With respect to the effect of the visualization type on usability, LSD post-hoc comparisons indicated three significant differences: between the PC visualization (*M*=6.55, *SD*=1.28) and the plain representation (*M*=7.32, *SD*=1.18), *p*=.008; between the SOMMOS visualization (*M*=6.74, *SD*=1.53) and the plain representation, *p*=0.046; and between the combined visualization of PC & SOMMOS (*M*=6.67, *SD*=1.26) and the plain representation, *p*=0.032. There were no further significant pair-wise differences evident in the effect on usability.

Table 8. Usage experience - between problem domains and between visualization types

Effect source	Dependent Variable	*(df,157)*	*F*	*Sig.*
VisualType	Ease of use	3	2.81	0.04*
	Learnability	3	2.49	0.06
	Helpfulness	3	0.73	0.54
	Choice confidence	3	0.29	0.83
	Usability	3	2.81	0.04*
Problem domain	Ease of use	1	9.51	0.002*
	Learnability	1	0.97	0.33
	Helpfulness	1	2.09	0.15
	Choice confidence	1	4.64	0.03*
	Usability	1	7.66	0.006*

Table 9. Descriptive statistics for usage experience factor per visualization type (1-9 scale)

	PC			SOMMOS			PC+SOMMOS			Table		
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Ease of use	6.13	1.63	45	6.53	1.65	43	6.37	1.47	36	7.06	1.36	41
Helpfulness	7.17	1.34	45	7.35	1.30	43	7.57	1.41	36	7.26	1.21	41
Learnability	7.53	.99	45	7.24	1.56	43	7.37	1.15	36	7.93	1.14	41
Usability	6.55	1.28	45	6.74	1.53	43	6.67	1.26	36	7.32	1.18	41
Confidence	7.56	1.25	45	7.40	1.76	43	7.33	1.47	36	7.61	1.39	41

Table 10. Descriptive statistics for usage experience factor per problem domain (1-9 scale)

	Rent			TAM		
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*
Ease of use	6.90	1.48	80	6.16	1.56	85
Helpfulness	7.48	1.10	80	7.19	1.47	85
Learnability	7.63	1.17	80	7.42	1.30	85
Usability	7.12	1.29	80	6.54	1.34	85
Confidence	7.74	1.25	80	7.24	1.62	85

6.3 Moderator Analysis

With respect to each of the factor groups proposed as possible moderators to the influence of visualization type on inter-objective correspondence understanding (i.e., H1), we hereby note the following results:

Problem Characteristics. We incorporated both problem domain and problem complexity factors into the experimental design, along with the manipulation of the four visualization types. As implied from Table 7, there was an interaction between problem characteristics and the main effect of the visualization interface. This was apparent in having a significant impact only for the case of lower complexity (i.e., RENT). However, since the manipulation of problem-characteristics was simultaneously facilitated by changing the number of alternatives in the solution space (20 vs. 100), the number of objective (3 vs. 5), and the problem domain (RENT vs. TAM), it is impossible to conclude at this point which one (or more) of the three individual factors were responsible for the moderation.

Participant Factors. Background knowledge in any of the two problem domains, and prior experience in using visual interfaces and in solving multi-objective problems were all incorporated as pre-test questions. We statistically tested the potential effect of these factors as possible covariates. Specifically, we incorporated the following factors as potential covariates: *knowledge in problem domain*, *experience in using visual interfaces*, and *experience in multi-objective decision making*. None of the participant factors were found to significantly moderate the main affect of the visualization type on conflict understanding.

Subjective Usage Experience. It is apparent in the results that the manipulation of problem domain also had a significant impact on *ease-of-use*, *choice confidence*, and *usability* as depicted in Table 6. Similar to the moderation of problem characteristics, this impact may have influenced the potential main effect of the visualization interfaces in the case of the more complicated problem domain. Hence, we statistically tested the potential effect of these factors as possible covariates. None of the usage experience factors were found to significantly moderate the main effect of the visualization type on conflict understanding. Correspondingly, we conducted all post-hoc analyses illustrated in the previous section without integrating the covariates.

7 Conclusions

Based on the results and the research model in Figure 2, we concluded that:

- The visualization of PC is more effective in terms of inter-objective conflict understanding than a tabular visualization when used in the context of a multi-objective decision problem that is not overly complicated and in a problem domain with which the decision maker has higher familiarity. Beyond a certain complexity threshold and\or when the problem domain is less familiar to the user, the gain from using a visual interface diminishes. This finding is in line with prior research [15] according to which domain familiarity should be taken into account as a significant background factor that can interact with the effectiveness of treatments in the context of problem solving tasks. Interestingly, this impact is not affected by the decision maker's background knowledge and experience with using multivariate visualizations.
- The type of the visualization aid has a significant impact on subjective usage experience, supporting H2. Specifically, the perceptions of ease-of-use and usability

show that a plain representation typically scores higher than a graphical visualization aid. This may be explained by the widespread use of tabular visualizations.
- The decision making problem at hand may also facilitate different perceptions of ease-of-use and usability. Further, decision making confidence is also affected by the problem characteristics. This effect should be carefully taken into consideration as it may also indirectly mask the effectiveness of the visualization aid in promoting better inter-objective correspondence understanding. Our experimental design employed a confounded manipulation, simultaneously tweaking problem domain (RENT vs. TAM), and problem complexity in two dimensions: size of solution space (20 vs. 100), and number of objectives (3 vs. 5). Thus, further work is required to determine which of the concrete dimensions may actually be a source for effect moderation.

Aside from the above conclusions inferred directly from the statistical analysis, our research team deduced some qualitative insights while administrating the experimental procedure. The most prominent observations included the following:

- Despite our expectations regarding the superiority of the combined visualization (PC+SOMMOS) over the individual constituents, it did not appear to have fulfilled this expectation. Observing user actions revealed that most participants who were assigned to this group used each of the two visualization interfaces independent of the other, practicing a fairly limited number of transitions between them. The root cause we observed for this behavior was the mere use of visual 'tabbing' which did not allow participants to simultaneously watch the two visualizations on a single screen. This may also be attributed to the content of the training session presented to this group.
- It was observed that users do not immediately rush into making a choice. Rather, the decision making process progressed through several phases. This included: high-level visual skimming of the entire solution space, prioritizing between the presented objectives, identification of choice groups that seem to possess similar characteristics, flipping back and forth between the upper and lower bounds for each objective range and observing its effect on the visible solution space while doing so, and gradually iterating through all previous phases while eventually converging on a single choice. Based on this observation our conclusion is twofold: (a) the suitability of concrete visualization means with regard to each of the above phases may vary, and (b) the multi-objective decision making process itself should be further investigated to more diligently learn about its methodological constituents.

8 Future Work

Different levels of complexity revealed the possible existence of a certain cognitive threshold beyond which the effectiveness of the visualization means is impaired. Our immediate intention is to pursue an experiment aimed at testing the exact complexity parameters that may play a significant role in this regard (e.g., number of objectives, solution space size).

As concluded from the observed methodological aspects of the decision making process, and in congruence with [31], a longitudinal investigation is necessary to further discover the actual cognitive stages associated with more effective decision making in multi-objective settings.

Our preliminary experiment was focused on the understanding of trade-off as the main outcome variable. Understanding the potential impact of visual interfaces on other outcome variables (e.g., cluster specific trade-off understanding) requires further investigation.

Finally, it is clear that certain features in the visualization interface can promote better decision making performance. For example, the ‘filtering’ feature allowing users to restrict the accepted range per objective was used extensively, regardless of the actual multi-dimensional visualization type. There are still a handful of additional features that need to be tested for similar impacts, including visual clustering, objective layout shuffling, glyph shapes, visualization layouts, and more.

Acknowledgements. Our thanks go to Dr. Ofer M. Shir for the insights and help in the field experiment.

References

1. Bagajewicz, M., Cabrera, E.: Pareto optimal solutions visualization techniques for multiobjective design and upgrade of instrumentation networks. Industrial & Engineering Chemistry Research 42, 5195–5203 (2003)
2. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry (1996)
3. Davis, F.: A technology acceptance model for empirically testing new end-user information systems: Theory and results (1985)
4. Davis, F.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly (1989)
5. DeLeeuw, K., Mayer, R.: A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. Journal of Educational Psychology (2008)
6. Ehrgott, M.: Multicriteria optimization (2005)
7. Fleming, P.J., Purshouse, R.C., Lygoe, R.J.: Many-objective optimization: An engineering design perspective. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E., et al. (eds.) EMO 2005. LNCS, vol. 3410, pp. 14–32. Springer, Heidelberg (2005)
8. Freeze, R., Raschke, R.L.: An Assessment of Formative and Reflective Constructs in IS Research. In: ECIS 2007 Proceedings (2007)
9. Gemino, A., Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering 9(4), 248–260 (2004)
10. Green, P., Srinivasan, V.: Conjoint analysis in consumer research: issues and outlook. Journal of Consumer Research (1978)
11. Green, P.E., Wind, Y.: New Way to Measure Consumers’ Judgments. Harvard Business Review 53, 107–117 (1975)
12. Horn, J.: Multicriteria decision making and evolutionary computation (1996)
13. Inselberg, A., Dimsdale, B.: Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: Proceedings of the 1st Conference on Visualization 1990, pp. 361–378 (1990)

14. Keim, D., et al.: Visual Analytics: Definition, Process, and Challenges. Neurosciences, 154–175 (2008)
15. Khatri, V., et al.: Understanding conceptual schemas: Exploring the role of application and IS domain knowledge. Information Systems Research 17(1), 81–99 (2006)
16. Kidwell, B., et al.: Emotional calibration effects on consumer choice. Journal of Consumer Research (2008)
17. Kohonen, T., et al.: Self-Organizing Maps. Springer, Berlin (2001)
18. Köksalan, M., et al.: Multiple Criteria Decision Making: From Early History to the 21st Century (2011)
19. Korhonen, P., Wallenius, J.: Visualization in the multiple objective decision-making framework. In: Branke, J., Deb, K., Miettinen, K., Słowiński, R. (eds.) Multiobjective Optimization. LNCS, vol. 5252, pp. 195–212. Springer, Heidelberg (2008)
20. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009)
21. Lotov, A., et al.: Interactive decision maps: Approximation and visualization of Pareto frontier (2004)
22. Lotov, A.V., Miettinen, K.: Visualizing the Pareto frontier. In: Branke, J., Deb, K., Miettinen, K., Słowiński, R. (eds.) Multiobjective Optimization. LNCS, vol. 5252, pp. 213–243. Springer, Heidelberg (2008)
23. Trochim, W.M.K.: Research methods knowledge base (2007)
24. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Information Systems Research 2(3), 192–222 (1991)
25. Obayashi, S., et al.: Multi-objective design exploration for aerodynamic configurations. AIAA Paper (2005)
26. Orme, B.: Getting started with conjoint analysis (2006)
27. Paas, F., et al.: Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist 38, 63–71 (2003)
28. Plaisant, C.: The challenge of information visualization evaluation. In: Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2004, vol. 11(4), pp. 109–116 (2004)
29. Purshouse, R.C., Fleming, P.J.: Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation. In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele, L. (eds.) EMO 2003. LNCS, vol. 2632, pp. 16–30. Springer, Heidelberg (2003)
30. Shneiderman, B., et al.: Creativity support tools: Report from a US National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction 20(2), 61–77 (2006)
31. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, pp. 1–7. ACM (2006)
32. Witowski, K., et al.: Decision making in Multi-Objective Optimization for Industrial Applications–Data mining and visualization of Pareto data. In: Proceedings of the 7th European LS-DYNA Conference (2009)
33. Yang, J., et al.: Interactive hierarchical displays: a general framework for visualization and exploration of large multivariate data sets. Computers & Graphics (2003)

Homestead Creator: Using Card Sorting in Search for Culture-Aware Categorizations of Interface Objects

Kasper Rodil[1], Matthias Rehm[1], and Heike Winschiers-Theophilus[2]

[1] Dept. Architecture, Design and Media Technology, Aalborg University, Denmark
[2] Dept. Software Eng. Polytechnic of Namibia, Namibia
kr@create.aau.dk, matthias@create.aau.dk,
hwinschiers@polytechnic.edu.na

Abstract. Designing intuitive interfaces for rural African users requires us to understand the users' conceptual model. We acknowledge differences in categorization approaches based on cultural factors, among others. In the absence of comprehensive literature and theories, we explore card sorting as a means to derive a local categorization of interface objects for one of our prototypes. Results indicate a locational-relational categorization scheme among Herero elders in Namibia.

Keywords: categorization, indigenous knowledge, 3D visualization, card sorting, HCI, interface design.

1 Introduction

Since late 2010 we have investigated 3D graphics as means of contextualizing indigenous knowledge in digital representations. Besides the numerous challenges of technical constraints and understudied implications of interactions, perception and recognition by African rural dwellers we have had a major breakthrough in terms of dialogical design communication. The graphical representations have triggered uncountable discussions, narrations and design suggestions in the community. Thus at this point we further explore visual communication as a form of design dialogue to overcome language and cultural barriers. Kostelnick describes visual communication design as a continuum between global and culture-focused design, where hybridity of the two ends can exist [1]. He argues that the former 'view' is a product of cognitive perception having universal empirically testable characteristics. The latter is an argument also carried in linguistics that culture is an influencing factor. That the characteristics of visual communication are experiential and socially constructed only fully transferred within their contexts of origin. It is the balancing act between sensitivity to locally negotiated methods and artifacts of visual communication and introduction to technology that currently occupy our research and must be understood from this viewpoint. The notion we wish to raise is not to argue one or the other, but argue for the sensitivity and hybridity of these two ends. That nothing can be taken for granted – especially not appraisal of universal metaphors in interface design or knowledge management structures.

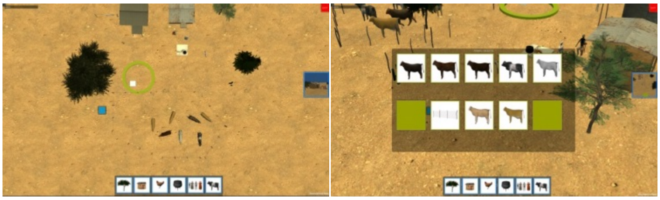
P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 437–444, 2013.

Since the conception of this project, elders from the pilot community in Namibia have been co-designing prototypes with external designers. The elders are well-respected and knowledgeable males, who have acquired much confidence in the usage of novel technologies as well as the ability to critically evaluate and suggest design improvements.

2 The HomeStead Creator

In 2011 we developed a prototype termed the HomeSteadCreator (HSC), which was received with enthusiasm by the Herero community members [2]. The HSC is running on an Android powered tablet. The village elders consider the tablet as being less intimidating and the touch interactions to be more intuitive than the usage of laptops and prototypes we have evaluated earlier.

HSC is a tool for rural Herero community members to re-create their own environment or any imaginary context with familiar 3D objects. These re-created scenarios are then combined with locally recorded IK videos and audio files to provide a digital context. Fig. 1 shows an example of the 3D re-construction of a homestead from two different cameras.



(a) (b)

Fig. 1. A re-created homestead seen from (a) 90-degree tilt or (b) free-roam camera perspective. (a) shows the six categories as default. (b) shows the cattle category expanded.

When the HSC is launched, a textured terrain in a 3D environment seen from a 90-degree tilt perspective (perpendicular to the terrain) is displayed.

The user of the HSC is presented with a row of icons representing categories defined by us. When they are being activated by a single-touch they expand and show virtual 3D objects within the ‘category’ (visible on Fig. 1(b)).

The objects are selected by single-touch and subsequently they are instantiated in the center of the terrain. The user can then position and rotate the object by common finger gestures (two-finger drag, two-finger circular motion etc.). The prototype was refined in 2012 based on the co-designers suggestions. From the introduction of the HSC, the tool has enhanced the design dialogue substantially while promoting cross-cultural learning. E.g. the elders have shown us the correct position of the holy fire and the elder's house or ‘the courtesy route around the homestead to greet the man of the house’ by visualizing it on the HSC. Thus it has shown to be able to facilitate dialogue on specific issues of interaction design and as an ethnographical tool using visualization as a communication tool. In subsequent sessions the HSC was even expanded to be a storytelling device between elders and youths.

3 Current Categorization Attempts

One of the focal points in the 2012 revision was the organization of the interface with the categorization of virtual objects. This grew out of the continuous demand for new graphical objects by the elders. The 7-10 inch tablet screen can only display so many objects at a time. The current prototype has six categories with up to ten objects (see figure 2). Upon suggestions by the elders, we did a number of rearrangements of objects. In 2011, we had pre-grouped the objects according to our own categorization scheme. During the usability evaluation session of the prototype, we also explored the appropriateness of the categorization as well as the completeness of objects represented. For example the 'fence' which was previously grouped under 'objects' was requested to be moved to the 'cattle' category. The cattle were previously grouped with the other animals yet in the Herero tradition, cattle are attributed a different status than other animals thus the request for a separate category. Furthermore, the 'fire place' was requested to be part of the 'homestead items' together with the pots and washing line. See Table 1 for the updated categories.

Table 1. The figure shows the six groups with corresponding icons and objects within each group

group	icon	objects	Count
local flora	Tree 	trees, cacti	8
larger structures	House 	houses, large solar panel, water pump, water tank	9
smaller animals	Chicken 	goats, sheep, dogs, chicken	7
objects	Cooking pot 	cooking pot, table, washing line, fireplace, chair	5
people	People 	people in various age groups, attires and postures	10
cattle	Bull 	bull, cows, fences	8

These seemingly minor re-arrangements reminded us of the underlying differences of conceptualization and categorization between the co-designers in the village and us.

We acknowledge that with the increasing number of objects, neither scrollable lists nor our pre-ordered categorization system is adequate. Instead of continuously re-iterating the categorization of newly added objects we attempt to conceptualize the underlying structuring of objects.

4 Situated Concepts in a Local Ontology

From a practical technology design point of view, we need a guiding ontology that captures the concepts and relations between them for a re-contextualization of the

Herero homestead. Defining an ontological representation will allow for sharing and re-using the captured knowledge [3]. Assuming a culture-specific categorization of the domain, the ontology will not only be useful for the obvious task of structuring menu items in the HSC. It will also capture parts of the intangible knowledge of the user group [4], thus allowing to create knowledge sharing applications that can partially construct relevant scenarios automatically based on the encoded knowledge.

Rehm [5] has focused on a situated acquisition of concepts and has shown the usefulness of combining nature and nurture views as described by Rosser [6]. On the one hand he has taken into account information processing and learning routines that have developed over time and are shared by all humans (nature perspective). On the other hand, based on these processing abilities, categorizations and concepts are developed by individual experiences, thus shaping an individual conceptual system that is in accordance with one's experiential history, taking into account environmental as well as social factors (nurture perspective). For the sake of the task we are interested in defining the relevant categories and their relations in the domain of the HSC. The 'fences and cows' example highlights the fact that there is no 'universal' conceptualization. Instead, we have taken into account Lakoff's ideas of situated concepts that cannot be viewed independent of the context of their use [7]. And as Ingold explains: "…to individuals who belong to different intentional worlds, the same objects in the same physical surroundings may mean quite different things. And when people act towards these objects, or with them in mind, their actions respond to the ways they are already appropriated, categorized or valorised in terms of a particular, pre-existent design." [8].

Not being members of the Herero culture we lack crucial information about the use of objects rendering the categorizations that we create from our own experiential history useless in the given cultural context. Studies have investigated these cultural differences. For instance Nisbett et al. found a difference between being either holistic (East-Asians) and analytical (Westerners) [9]. Hunn describes how the Tzeltal peoples (a Mayan ethnic group) classify butterflies and moths [10]. They distinguish butterfly larvae into sixteen terminal groups due to characteristics of being edible, dangerous to crops etc. The adult butterflies do not have these characteristics (although being visually distinguishable) thus they are (locally) not important for categorization. Similarly, do members of the Herero community distinguish cattle from other animals and require a larger variety of the same type. Local trees are by the elders also not necessarily classified by species, but by characteristics that make them appropriate for the homestead or the kraal. To determine a local categorization scheme of virtual objects we have turned to card sorting as a means to find a local ontology for further developing the object structures in the HSC [11].

5 Card Sorting as a Method for Establishing the Local Classification Scheme

Card sorting is a traditional HCI method to organize information in web pages into meaningful categories and for intuitive retrieval of information.

The method is usually either an ‘open’ (generative type) or ‘closed’ (evaluative). One of the benefits with the ‘open’ version is that participants can categorize objects without pre-defined sorting, themes or adhering to an overlying taxonomy. Conversely, the ‘closed’ version is suitable for establishing at what level participants agree with the categorization and terms used for pre-sorted groups.

Card sorting is valuable for gaining insights into the participants’ mental models by eliciting how they sort, group and organize items. Petrie el al. showed that card sorting can illuminate cultural differences in the mental models behind information architectures [12]. The product of the card sorting method is a snapshot of a subjective categorization and developers might be fixed on the final sort to implement information architecture from. However we argue that the method itself (especially the open version) is also effective in creating a dialogue around the user’s standpoint. Thus in the process of sorting the researcher gains an insight on the users’ world views.

6 Method

In total we conducted the study at three different Herero sites in Namibia. With a total number of participants being 5 females and 9 males. The presentation of detailed results is beyond the scope of this paper, therefore we will only present results from the 5 elders from our long-term collaboration village.

In preparation for the card sorting sessions a set of 47 cards was printed. The set consists of laminated cards with images of the 3D objects from the HSC, which are rendered with a white background. The cards are printed as they appear as icons in the six interface categories in the HSC.

The village elders who are the future users of the HSC and have been co-designers on the project from its conception were recruited for the sessions. The sessions were decided to be facilitated individually to investigate each elder’s viewpoint. The elders were prompted to say what they saw on the cards and to talk aloud while sorting, this was to confirm recognition of the represented 3D objects and to understand the rationale for the sorting decisions.

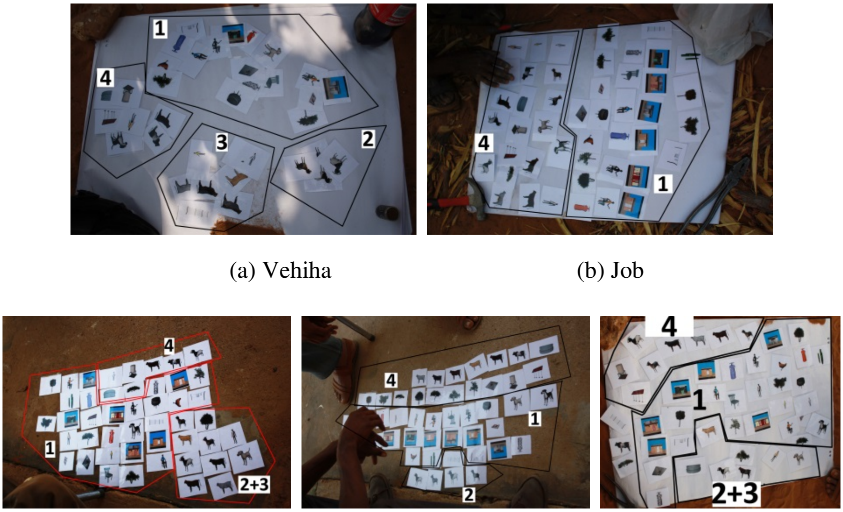
The agenda was explained before the sessions began and we stressed the openness of the studies by emphasizing that nothing was regarded wrong or right and their help would aid in improving the prototype. An open card sorting was chosen for not to impose any overlaying structure potentially overriding the local way of categorization. Thus the elders were instructed to group the cards as they preferred.

The sessions were documented with video recordings, still photographs, observation and interview notes. In the case of a participant not being English speaking the interviews were facilitated by a local Herero co-researcher. An independent Herero translator translated the videos after the field trips to minimize translation bias.

7 Results

We were able to recruit 5 elders in the village. Fig. 2 displays the 5 final card sorts from the study conducted in the village.

The first participant (Vehiha) began by looking at all the cards he was given. After about five minutes, he began placing cards into five, laid out, distinguishable groups with all the cards facing up. He said that the two groups representing the homestead could be piled together since they represent the same. The cards were completely mixed when relating the now four card groups (see Fig. 2 (a)) to the six interface categories we implemented in the HSC. It puzzled us that he had actually made a map of the village. The four groups were each representing a place within the village. Marked on Fig. 2 as: 1: [homestead(s)], 2: [group of goats], 3:[cattle kraal] 4: [community water pump area].



(a) Vehiha (b) Job

(c) Gerrard (d) Beau (e) Benji

Fig. 2. The pictures show the final card sorts by the 5 village elders. Annotations are added to visualize the categorization of cards based on a location-relational categorization scheme.

For instance, one card group contained cards depicting the community water pump (where they bring their cattle for water), a young boy, a bull, a dog, a cow, a solar panel and a water tank. He said that there would always be a young man taking care of the cattle and being responsible for herding them between the kraal (animal enclosure) and the water pump. Thus, the young man belonged to that water pump group. Then Vehiha added a boy to the cattle group (3) and said (trans.): “while the big man is busy at the homestead the others are at the kraal.” So the ‘boy’ card belongs there too.

We asked where we should place a certain tree. He said that specific trees in the bush do not belong in the groups he sorted, but there should be a tree for the boy to rest under at the water pump group, which he then added. The other elders followed the same categorization scheme, and where consistent each in making a group for the homesteads (1) and the water pump (4). Participant 2 (Job) emphasized the layout of the homestead and the order of objects relational to each other situated there. For instance by showing that the man of the homestead per tradition sits to the right of the entrance between the house and the holy fire (see Fig. 2 (b)). This explanation was consistent with similar explanations given by the elders, thus the card sorting method was also a trigger for sharing local customs.

All cards were placed according to location of where the real objects/people represented as cards are functioning or located in relation to each other. Unsurprisingly none adhered to our categorization scheme. It must be stated that all findings across the three sites were showing the same categorization scheme as the results as from the five elders reported here. And for instance when we prompted on where to place the snake (this was a hypothetical card mentioned in the discussion), the participants stated that the snake would not belong in any group, since snakes have no purpose within the village.

8 Discussion

The open card sorting proved very useful since it was possible to uncover an unexpected categorization scheme. The exploratory nature allowed external researchers to ask follow-up questions ensuring that the rationale behind the categorization was transferred from the participants. It also accommodated through dialogue, that the categorization would need duplicates of some objects per group. This finding is highly interesting since card sorting is often used to categorize an item only one place based on participants' suggestions. Here it was established -contrary to our categorization scheme- that some objects should be multiple places based on participants' suggestions. The set of cards did not have duplicates for all objects, which actually shows how our method was influencing the study with a preconception of dichotomous thinking. For instance if there were more 'fence' cards the participants would have been able to separate the groups more visibly as the places are separated by fences in the village. Conversely, the elders were able to place more than one 'boy' card since we made three 3D models of a boy (thus more cards). Although we gained the feedback in the sessions on this matter, further studies should have more cards of the same type to accommodate the possibility to place many of the same type.

9 Conclusion

All findings from the categorization tasks point towards a shared local way of grouping the cards. We observed a 'place' based pattern, where objects meaningful to a location are grouped together, such as the 'boy' that looks after the 'cattle' at the 'kraal'. The open card sorting method supported the discovery of a new categorization scheme which we call location-relational. The pattern confirms prior research on Herero's prioritization of place as a meaningful location with activities and protocols attached to [13]. The findings dramatically changes our previous design ideas and future retrieval of interface objects will follow this new categorization scheme for ensuring intuitiveness and preservation of tacit knowledge in the interface.

The full data set is sufficient for the purpose of informing our local design since it is in the hands of this study's elders the HSC will be used. We proceed to refine the study for robustness, investigate the many details left out for this contribution and conduct follow-up studies. We strongly encourage researchers to pursue the data collection and analysis for a sound theory building in this under-researched area.

References

1. Kostelnick, C.: Cultural adaptation and information design: two contrasting views. IEEE Transactions on Professional Communication 38(4), 182–196 (1995)
2. Rodil, K., Winschiers-Theophilus, H., Jensen, K.L., Rehm, M.: Homestead creator: a tool for indigenous designers. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, NordiCHI 2012, pp. 627–630. ACM, New York (2012)
3. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human Computer Studies 43(5), 907–928 (1995)
4. Unesco: The UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage, `http://www.unesco.org/culture/ich/en/convention/` (accessed May 6, 2013)
5. Rehm, M.: Being seaward-handed: a computational model of the acquisition of language-specific spatial references. Cognitive Processing 5(1), 15–30 (2004)
6. Rosser, R.A.: Cognitive development: Psychological and biological perspectives. Allyn and Bacon Boston (1994)
7. Lakoff, G.: Cognitive models and prototype theory. In: Neisser, U. (ed.) Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization, vol. 1. Cambridge University Press (1989)
8. Ingold, T.: The Perception of Environment. Essays on Dwelling, Livelihood and Skill. Routledge, London (2000)
9. Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A.: Culture and systems of thought: holistic versus analytic cognition. . Psychological Review 108(2), 291 (2001)
10. Hunn, E.: The utilitarian factor in folk biological classification. American Anthropologist 84(4) (1982)
11. Ross, J.: Comparing user research methods for information architecture (2011), `http://www.uxmatters.com/mt/archives/2011/06/comparing-user-research-methods-for-information-architecture.php` (accessed March 24, 2013)
12. Petrie, H., Power, C., Cairns, P., Seneler, C.: Using card sorts for understanding website information architectures: Technological, methodological and cultural issues. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 309–322. Springer, Heidelberg (2011)
13. Bidwell, N.J., Winschiers-Theophilus, H., Kapuire, G.K., Rehm, M.: Pushing personhood into place: Situating media in rural knowledge in africa. International Journal of Human-Computer Studies 69(10), 618–631 (2011)

The Influence of Website Category on Aesthetic Preferences

Eleftherios Papachristos and Nikolaos Avouris

University of Patras, Human-Computer Interaction Group,
Electrical and Computer Engineering Dept.,
265 04 Rio Patras, Greece
epap@ece.upatras.gr, avouris@upatras.gr

Abstract. This paper investigates whether users' aesthetic impressions about websites vary considerably across different domains. The assumption that aesthetic judgments about websites that belong to diverse domains are based on different visual design aspects has been investigated in three distinct studies in healthcare, tourism, and web design business. In these studies participants expressed their overall preference as well as their judgments on the constructs of visual appeal, perceived usability and novelty. In addition, descriptions about the test websites were obtained by expert panel and objective measures. Preference Mapping (PM), which is a data summarization and visualization technique, has been performed in each study. Attribute projection into the preference maps allowed for the identification of important driver of preference for each individual domain. Even though, visual appeal was the most important predictor of overall preference in all studies, appealing websites had different visual characteristics in each domain. Furthermore the importance of the evaluation constructs varied considerably among studies, indicating that aesthetic perceptions differed considerably across domains. These findings emphasize the need for flexible evaluation methods that can be used to identify important visual design factors within a specific website domain.

Keywords: Website design, aesthetic evaluation, website categories, visual appeal, preference mapping.

1 Introduction

Since Tractinsky et al. [1] in 2000 published their seminal paper "What is beautiful is usable" much has changed in the HCI community in regard to aesthetics research. The lively debate that was initiated with the controversial suggestion that aesthetic design could influence perceptions about pragmatic qualities of user interfaces shifted attention to more subjective aspects of interaction. Particularly for websites the construct of *visual appeal* (used by some authors interchangeably with beauty) has been proven to be a very important factor determining users' overall impressions [2]. Furthermore, a series of experiments (e.g. [3]) have sown that users could form stable visual appeal judgments in time periods of as short as 50 msec. These judgments are based mainly on visual design since other aspects (e.g. content) cannot be recognized

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 445–452, 2013.

during such sort time periods. Surprisingly, relatively stable judgments in the same exposure times could also be found for the constructs of perceived usability and credibility [3]. These results demonstrate the general importance of websites visual design. Although, actual website use could influence users' perceptions, it is nonetheless important to create positive first impressions, considering that it is more difficult to overcome negative ones.

These research findings emphasize the importance of website's visual design and therefore the need for appropriate evaluation methods. The most common evaluation approach is using one of the aesthetics oriented multiple-item instruments that are gaining steadily acceptance in the HCI community; for example the "Classical-Expressive" aesthetic scale [3], AttrakDiff [2] or the more recent visAWI [6] questionnaire. Among the advantages, of those multiple-item questionnaires is that they ensure fairly reliable and valid measurements and that they provide common ground for results communication and for between study comparisons. These questionnaires have been created to be sufficiently generic in order to be applicable to most - if not all - websites. This generality has been achieved by the inclusion of a variety of websites from different domains as test stimuli during questionnaire creation. Both the VisAWI and the "Classical–Expressive" aesthetic scale, for example, reported a similar website sampling procedure during questionnaire creation and validation. The result of using websites of different domains as test stimuli is that design factors pertinent to specific domains are canceled out. Visual design evaluation with one of the aforementioned questionnaires means having participants rate a design on a predefined set of factors that have been identified to be important for websites in general.

However, websites vary in terms of purpose, target user groups, and therefore visual styles. Although, there is no commonly agreed upon taxonomy, various categorizations schemes of websites have been proposed. Studies have shown that users have distinct mental models for different kinds of websites [7]. According to Norman [8] users form internal mental models of things with which they have interacted, which in turn creates expectation about similar objects they may encounter. These expectations are becoming stronger as the number of encounters increases and users are gaining more experience with a particular website domain. For example, users may have different expectations about the visual design of news websites or online shops. Tuch, et al. [9] showed that users' aesthetic judgment can be strongly influenced by their perceptions of website prototypicality. Prototypicality refers to the amount to which an object is representative of a class of objects and depends heavily on each individuals mental models that are build through experience. Designs that contradict what users typically expect of a website may lead to a negative first impression [9]. The fact that prototypicality is an important influencing factor in aesthetic judgments has been shown repeatedly in various empirical studies [9]. Thus, relying on predefined questionnaires could mean ignoring the visual design aspects that in the users mind have particular importance in the specific website domain.

Other evaluation methods such us Repertory grid technique (RGT), Multidimensional Scaling (MDS), or Preference Mapping (PM) are based on multiple website evaluations and do not impose a set of predefined evaluation criteria on participants. In addition, conducting evaluation studies with one of these techniques allows for the identification of design factors that are important in specific website domain by using multiple websites from the same domain as study stimuli. In the studies presented in this paper Preference Mapping (PM) has been used for evaluation of visual design of a number of examples of websites of a specific domain. This was repeated in the domains of healthcare, tourism and web design businesses. The main objective of this paper is to demonstrate that different design factors can be important driver of preferences in websites that belong to different domains, based on comparison of the findings of evaluation of the websites of these domains. This research is part of a broader research project that attempts to define guidelines for design and evaluation of various kinds of websites including social media presence of small medium organizations in various fields.

2 Method

Three evaluation studies have been conducted involving website designs from three distinct domains. In the first study (Healthcare domain) 15 hospital websites were evaluated by 34 participants (29 male, 4 female, mean age = 22.2). In the second study (Tourism domain) 32 participants (21 male, 11 female, mean age = 23.3) were asked to evaluate 18 hotel websites. In the third study (Web design business domain) 12 websites of web design companies were evaluated by 30 participants (17 male, 13 female, mean age = 28). These particular websites types were selected because they represent domains that differ both in terms of characteristics and in purpose. However, given that the goal was to identify important visual design characteristics that can shape user first impressions in each of these categories it was important to make sure that none of the test websites were previously known to our participants. Therefore, test websites were randomly selected from lists of top U.S. hospitals in the first study and New Zealand hotels in the second. Since all of our participants were of European origin it was assumed that these choices would minimize the possibility of prior familiarity with the specific test websites and thus influence the "first impression" effect. None of the participants reported previous experience with any of the test websites. In all cases participants volunteered to take part in the evaluation studies and did not receive any compensation. In all studies screenshots were used instead of actual websites, as our goal was to study the impact of various visual design aspects on participants' first impressions. Although, studies have shown [3] that perceptions about constructs such as visual appeal are relatively stable over time evaluations after actual website use could be biased by non-visual design related aspects (e.g. content).

2.1 Procedure

In each evaluation study participants first viewed screenshots of all the websites in a random order and then rated them according to their overall preference on a linear, unmarked scale (from 0 to 100) with the verbal anchors "least preferred" and "most preferred" at the two ends. In a subsequent evaluation phase participants were asked to rate the websites again on the constructs *visual appeal*, *credibility*, *perceived* *usability* and *novelty*. Since participants had to rate multiple websites in each study only a limited number of evaluation constructs had been included in the studies in order to avoid participant fatigue. However, the identification of important design characteristics required a better profiling of the test websites that could be provided by these four constructs alone. For this reason all websites were rated on various descriptive attribute by an expert panel.

Nine experts (visual designers, HCI practitioners, and web developers) identified a list of visual design aspects which could possibly influence user preferences. The goal was to find a comprehensive list of descriptive attributes, such as *symmetry* or *complexity*, which could reflect variations in website designs. After a literature review and panel discussions, a preliminary list of attributes was tested on a set of generic websites in order to eliminate unsuitable attributes. The criteria for elimination were: limited discrimination ability and disagreement between assessors about meaning. In a subsequent session our experts rated the actual test websites on the final 15 attributes. Before finalizing the descriptive dataset, the attributes that did not discriminate significantly between our actual websites were identified through mixed model ANOVA's (websites as fixed and experts as random factors) and were excluded from further analysis.

In addition to participant and expert ratings we also used 16 objective measures that could be grouped into three categories: a) *text related metrics* (e.g. number of words, number of visible links), b) *area related metrics* (e.g. percentage of website used to display images or text), and c) *color related metrics* (e.g. average brightness, saturation). The text related measurements were taken with the help of optical character recognition (OCR) software and were double checked manually. This technique was used instead of html parsing since contemporary websites use graphics or flash instead of plain text very frequently. Websites fragmentation to specific areas (e.g. navigation, images) was done manually with the help of graphic editing software. For the color related measurements, a color recognition program has been written that parsed the website screenshots and calculated the metrics.

3 Analysis

In order to identify which design characteristics were most influential in preference formation, a common approach would be to perform multiple regression with *preference* as the depended variable and the various attribute ratings as the predictors (e.g. [2]). However, since a purely exploratory approach was followed in regard to the selection of predictors the number of independent variables (30-35) in most studies was larger than the number of observations. Using all predictors at once would over-fit the regression model. In addition most of the predictor variables are highly correlated to each other, which can lead to multicollinearity problems. Principal Component (PCR) or Partial Least Square (PLS) regression models in which a large amount of predictors are transformed into view orthogonal uncorrelated components are better suited for these circumstances. A data analysis method based on PCR called Internal Preference Mapping (IPM) has been used in our study in order to identify important design characteristics in each case.

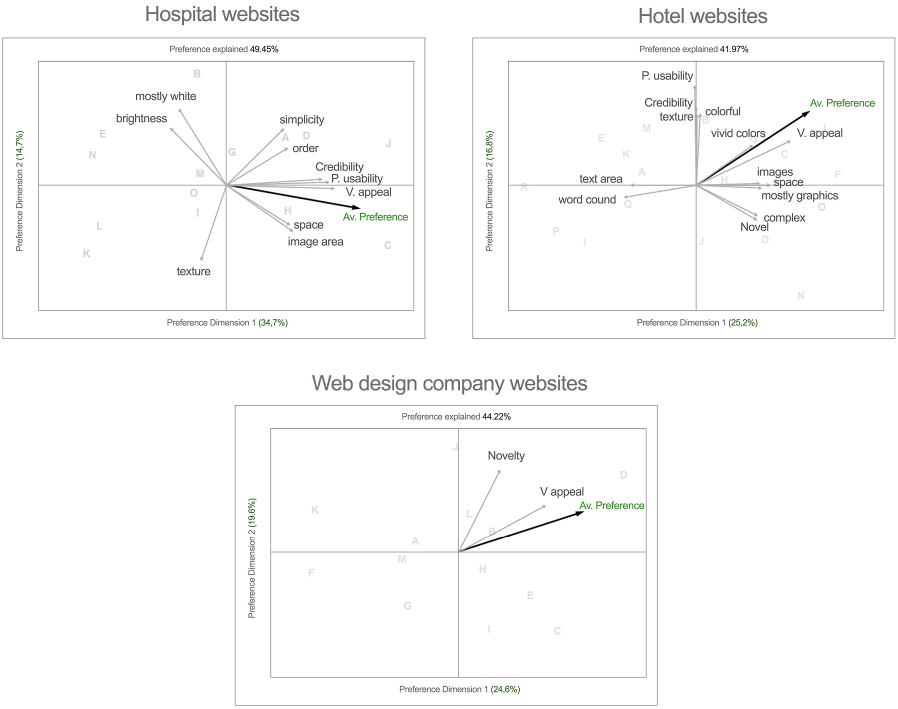


Fig. 1. Preference maps with attribute projection for each of the three studies. Light vectors represent attributes. Attributes that begin with a *capital letter* are participant construct ratings. The bold vector indicates the average preference direction of the participant sample.

Preference *mapping* is referred to as a group of multivariate statistical techniques aimed at gaining deeper understanding of participants' preferences toward stimuli [10]. The method is a data summarization and visualization technique that creates low dimensional maps depicting stimuli and individual participant preferences simultaneously. This is usually accomplished by conducting Principal Component Analysis (PCA) on a data matrix consisting of stimuli in rows and participant preferences in columns. Since in PCA the first components account for the maximum possible amount of variance two or three dimensional spaces are usually sufficient to capture the majority of the underlying preference structure. Based on Euclidean distances conclusion can be drawn regarding website similarities as well as individual participant preferences towards them.Interpretation of the resulting dimensions as well as identification of important drivers of preferences can be accomplished by projecting additional website attributes into the preference map. Additional data about websites can be projected in the preference space by using average attribute scores as dependent and website factor scores as independent variables in a regression model. The regression coefficients represent the strength of the relations between the additional attributes and the preference dimensions. Attributes that have no relationship with any of the preference dimension cannot be used for preference interpretation and should therefore be removed from further analysis. Thus, during website data projection insignificant attributes that cannot explain participant's preferences can be identified and discarded.

Participant preferences ratings were submitted to a PM analysis for each study individually. Figure 1 shows the three resulting preference maps. In these maps website designs are represented by capital letters while attributes are depicted as light vectors. The vectors indicate the general direction in which the intensity of each attribute increases. Attributes with vectors that point to a similar direction are positively correlated while attributes with vectors pointing to opposing directions are negatively correlated. Websites that lie in the general direction to which an attribute vector points have high intensities of that attribute while objects in the opposite directions have none or low intensities of the same attribute. The bold vector indicates the average preference direction of the participant sample.

From the initial list of 35 attributes only 9, 13 and 2 could be successfully projected in the 2-dimensional preference spaces of the healthcare, tourism, and web design business studies respectively. Although, analysis has been conducted in higher dimensional spaces the results presented here are primarily focused on the first two dimensions which represent the most important components in participant preferences.

In the healthcare study the first preference dimension represented 34.7% of the total preference variance and was highly correlated with the constructs of visual appeal, perceived usability and credibility. The second dimension which explained 14.7% of preference variance correlated with descriptive attributes such as simplicity, order, brightness on one site and large image area, white space and texture on the other.

In the *tourism* study the first dimension that captured 25.2% of preference variance differentiated among websites on the left site of the map which were *mostly graphics* based, had more *white space*, and *larger image areas* while designs on the right site were mostly *text based*. The later was confirmed by trained panel data (*mostly graphics* attribute) and by objective measures (*text area*, *word count*). The second dimension was primarily correlated with *perceived usability* and *credibility* as well as to the descriptive attributes *texture*, *colorful* and *dimensional*. *Novelty* was positively correlated with the first and negatively with the second preference dimension.

In the *web design business* study only two constructs could be projected in the 2-dimensional preference space. Interpretation of the preference dimensions solely based on these constructs indicate that the most important component is more related to *visual appeal* and the second to *novelty*. It is noteworthy that none of the descriptive attributes as well as none of the constructs *credibility* and *perceived usability* were useful in interpretation of the first two preference components in this study. The attributes *order, credibility* and *perceived usability* were actually found to be highly correlated with the third dimension that represented only 13.6 % of preference variance.

Considering the average preference direction of the participant sample the most important predictor of website preference was *visual appeal* in all studies (as in [2]). However, in the healthcare study *visual appeal* was highly correlated with *perceived usability* and *credibility*. In addition, visual characteristics that were common among preferred websites in this category were *simplicity*, *order large image areas* and *white space*. In the tourism study visual appeal was not correlated with *credibility* or *perceived* *usability*. Preferred websites in this category had *vivid colors*, many *images*, plenty of *with space*, and a small *text area*. In the web design business study only *novelty* could be recognized as an important design factor apart from *visual appeal*.

4 Discussion

Preference mapping analysis revealed that diverse design characteristics can be important drivers of preferences for designs that originate from different website domains. Although, the most important preference predictor in all studies was *visual appeal,* what constitutes an appealing website differed considerably among the three studied cases. In the healthcare study the constructs *perceived usability* and *credibility* were found to be equally important as *visual appeal*. Generally preferred and appealing websites in this study were *simple*, *ordered*, *spacious* designs with large *image areas*. In the tourism study participants showed an aversion towards websites that were primarily text based. Designs that used *mostly graphics, vivid colors* and large *images* of hotel rooms were the most preferred and *appealing* websites. Perceptions of *usability* and *credibility* were exclusively related to the second dimension and were therefore less important drivers of preference than in the healthcare study. Finally in the web design business study only *visual appeal* and *novelty* could be found to be important drivers of preferences in the two dimensional space. Websites in this study varied on more design factors and could be characterized as unusual or extreme compared to designs in the others studies. This was generally expected since visual design in this domain serves as a first showcase of the company's ability to produce cutting edge design. Creativity and unconventionality cannot be appropriately captured by descriptive attributes and therefore none of the expert panel or objective measures were useful in preference interpretation in this study. The constructs of *perceived usability* and *credibility* that were found to be important drivers of preference in the other two website domains were less important in this one.

These results demonstrate that different design aspects play a determining role in preference creation towards websites within a specific domain. Therefore, misleading conclusions can be drawn by relying on a fixed set of evaluative or descriptive attributes for evaluation purposes of websites in general. Use of generic questionnaires for website design evaluation could lead to consideration of less relevant attributes into the evaluation process while design aspects that are central to a specific website domain could be ignored. For example, *novelty* was found to be one of the most important drivers of preference in the third (web business), fairly important in the second (tourism), and not important at all in the first (healthcare) study. Furthermore, *symmetry* which is an item in the classical aesthetic dimension in the questionnaire of Lavie and Tractinsky [4] was not found to be an important preference attribute in any of the studies presented in this paper. In addition, use of *images* that has been found to be an important driver of preference in two out of the three studies is a design aspect that is ignored by all aforementioned multiple item questionnaires [4][5][6].

5 Conclusion

This paper reports results from three different studies in which the influence of website domain on users' aesthetic preferences has been investigated. The results suggest that diverse design aspects can influence participants overall impressions of website designs in different domains. A different set of descriptive and evaluative constructs could be identified as important drivers of preference in three distinct studies involving healthcare, tourism and web design business websites. *Visual appeal* was the most important predictor of participant preferences towards websites in all case studies. However, appealing websites had different visual characteristics in each domain. To conclude, the results in this paper demonstrate the influence of website domain in shaping users' aesthetic preferences. This finding emphasizes the need for flexible evaluation methods that do not ignore the visual design aspects that are important in specific website domains.

Acknowledgements. This work has been partly funded by the European Territorial Cooperation Operational Programme "Greece-Italy 2007-2013" under the project Intersocial.

References

1. Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. Interacting with Computers 13(2), 127–145 (2000)
2. Schenkman, B.N., Jönsson, F.U.: Aesthetics and preferences of web pages. Behaviour & Information Technology 19(5), 367–377 (2000)
3. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, P.: An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Transactions on Computer-Human Interaction 18(1) (2011)
4. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer Studies 60(3), 269–298 (2004)
5. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: A questionnaire to measure perceived hedonic and pragmatic quality. Mensch & Computer. Interaktion in Bewegung (2003).
6. Moshagen, M., Thielsch, M.T.: Facets of visual aesthetics. Int. Journal of Human-Computer Studies 68(10), 689–709 (2010)
7. Roth, S.P., Schmutz, P., Pauwels, S.L., Bargas-Avila, J.A., Opwis, K.: Mental models for web objects: Where do users expect to find the most frequent objects in online shops, news portals, and company web pages? Interacting with Computers 22(2), 140–152 (2010)
8. Norman, D.: Some observations on mentai models. Mental Models 7, 7–14 (1983)
9. Tuch, A.N., Presslaber, E.E., Stöcklin, M., Opwis, K., Bargas-Avila, J.A.: The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. Int. J. of Human-Computer Studies 70(11), 794–811 (2012)
10. Meullenet, J., Xiong, R., Findlay, C., Knovel: Multivariate and probabilistic analyses of sensory science problems. Wiley Online Library (2007)

WATTSBurning: Design and Evaluation of an Innovative Eco-Feedback System

Filipe Quintal, Lucas Pereira, Nuno Nunes, Valentina Nisi, and Mary Barreto

Madeira Interactive Technologies Institute, University of Madeira, Caminho da Penteada, Funchal, 9020-105, Madeira, Portugal
{filipe.quintal,lucas.pereira,mary.barreto}@m-iti.org, njn@uma.pt, valentina.nisi@gmail.com

Abstract. This paper reports a 15 weeks study of artistic eco-feedback deployed in six houses with an innovative sensing infrastructure and visualization strategy. The paper builds on previous work that showed a significant decrease in user awareness after a short period with a relapse in consumption. In this study we aimed to investigate if new forms of feedback could overcome this issue, maintaining the users awareness for longer periods of time. The study presented here aims at understanding if people are more aware of their energy consumption after the installation of a new, art inspired eco-feedback. The research question was then: does artistic eco-feedback provide an increased awareness over normal informative feedback? And does that awareness last longer? To answer this questions participants were interviewed and their consumption patterns analyzed. The main contribution of the paper is to advance our knowledge about the effectiveness of eco-feedback and provide guidelines for implementation of novel eco-feedback visualizations that overcome the relapse behavior pattern.

Keywords: Sustainability, Aesthetics, Eco-feedback, User Interfaces, Prototyping.

1 Introduction

Individual household consumption accounts for a significant part of the total worldwide energy consumption. For example domestic electricity is responsible for two thirds of the electricity used in the United States, 36% of the greenhouse gasses, and 12% of the fresh water consumption [1]. In the European Union final energy consumption for households is about 31% of the total energy consumed, only second to transportation, which accounts for 36%. Between 1985 and 1998, the actual amount of energy consumed per household remained nearly constant, but the growing number of households increased energy use by 4%. This effect will have a dramatic global impact as developing countries contribute to the increase of households with energy access. Also while household electrical appliances are becoming more efficient, there are more of them and they are being used more often. Reports from the EU show that consumption by all-electrical appliances and lighting represents about 55% of the electricity used by households. These appliances include the six large consumers of

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 453–470, 2013.

electricity (refrigerators, freezers, washing machines, dish-washers, TVs and dryers), and many other small appliances. [2].

Studies have shown that individual activities can control up to 50% of residential electricity consumption, depending on the physical structure of the building and what set of appliances users can control [3]. Moreover, most people are in fact concerned about the consequences of their actions. However, they are also unaware of the impact of their daily activities and more importantly how they can change their behavior to reduce consumption. This gap between users concerns and their actual knowledge of energy consumption habits, motivated companies and researchers to develop technologies that present users with information about their consumption. This type of technology is commonly called eco-feedback and is defined as technology that provides feedback on individual or group behaviors with a goal of reducing environmental impact [4].

The work presented here is part of a broader sustainability interdisciplinary research project which involves the deployment of a combination of sensing and eco-feedback technologies to motivate and trigger people to think, act, reflect and consume sustainably.

2 Related Work

The advancement and availability of sensing systems for environmentally related activities (e.g., human activity inference [4]) and interactive displays to feedback this data (e.g., mobile phones) provides a rich space of prospects for new types of eco-feedback solutions [5]. Currently there is an increasing number of commercial applications that provide real time energy monitoring. These solutions range from low-cost single-outlet (Kill-a-Watt[1] and Watts Up[2]) to medium cost whole house power consumption (CurrentCost[3], TED[4], Efergy[5], Owl[6], etc.) to higher cost ambient feedback solutions (Wattson[7], Energy Orb[8]). Through these solutions, feedback is often presented as raw energy use (e.g., Watts), personal cost (e.g., money), or environmental impact (e.g., CO2 emissions). Furthermore, eco-feedback is also an increasingly important research arena, confirmed by the growing number of articles presented at top international venues like CHI, INTERACT, DIS and Ubicomp, which nowadays devote specific sessions to sustainability. As a consequence, literature is abundant in design strategies and guidelines to implement such systems [6, 7, 8, 9].

---

[1] http://www.p3international.com/products/special/p4400/p4400-ce.html
[2] https://www.wattsupmeters.com/secure/products.php
[3] http://www.currentcost.com
[4] http://www.theenergydetective.com
[5] http://www.efergy.com
[6] http://www.theowl.com
[7] http://www.diykyoto.com/uk/
[8] http://inhabitat.com/the-energy-orb-monitor-your-electricity-bill/

One of the main challenges of designing eco-feedback systems is to present to users, how individuals or group activities can impact the environment. Commercially available feedback systems, have tried to present this information as kilograms of CO2 emissions. Pierce et al. [10] surveyed several publications exploring the impact of eco-feedback technologies in energy consumption and on consumers' behaviors. The authors found out that the use of eco-feedback technology resulted on savings between five and twelve percent of daily energy consumption. They also concluded that when savings didn't happen, the eco-feedback was displayed too infrequently (monthly) and hence was disconnected from the consumption behavior. Egan [11] confirmed that receiving timely feedback is key to motivating behavior change. Moreover, Fischer reports that eco-feedback is more efficient when given frequently, clearly presented, using computerized tools and allowing historic or normative comparisons [12].

Normally eco-feedback is associated with behavior change. There is evidence that the single goal of saving money (less electricity you consume the less you pay), is not enough to motivate conspicuous behavior change – between 5 and 15% reduction on average [13]. Consolvo et al. [14] implemented a system that promotes a more physically active lifestyle. Here the authors propose eight qualities that a system should have in order to be well accepted by users: Abstract & Reflective, Unobtrusive, Public, Aesthetic, Positive, Controllable, Trending / Historical and Comprehensive. These studies have encouraged research into non-traditional feedback systems, turning to digital art and disruptive design for help. In the digital art domain, efforts were made to raise awareness regarding their behaviors in relation to sustainability.

For example, digital artist Tiffany Holmes, visualizes energy consumption through an art installation situated in a public space [15]. Holmes uses digital art to display hidden data of real time usage of key resources (such as electric appliances) and providing an aesthetically striking visualization in the public space of the building hall. Dwellers were able to relate to the visualization as a community, negative feedback was avoided and people living in the building were empowered to act upon their consumption by getting to know data that wouldn't be available to them otherwise. Another example of digital art employed to raise awareness about energy consumption is the Helsinki based project presented in [16]. The project consists of a public installation, in which a green cloud proportional to the city energy consumption is laser projected onto the smoke generated by the chimney of a coal energy power plant in the city. During the seven days of the installation, the green projected cloud would grow and shrink in direct proportion to the city energy consumption. In the final day of the installation, residents were asked to unplug the devices to reduce consumption, in order to dramatically increase the size of the green cloud. This resulted in a reduction of the peak demand in 800 kVA (approximately the same power generated by a windmill running for one hour).

Despite these developments the effectiveness eco-feedback is known to have problems. Peschiera, reports that that after a certain period of usage of the feedback devices, users consumption relapse to values prior to the study [17]. Holmes reported this phenomenon in [15] where it was possible to see that users gradually returning to their previous behaviors if feedback was less frequent or no longer present. Our own research [18] confirmed this: *"We would check our consumption more often initially until we got a rough idea or perception of what our consumption was but after that it* *would less frequent ",* furthermore we observed that this lack of attention starts to happen after the fourth week of usage.

2.1 Previous Work

Prior to the development and test of the WATTSBurning system, the research team was responsible for 2 long-term eco-feedback deployments. Those deployments involved 30 houses and apartments located in an urban area in southern Europe. To better understand our starting ground, we briefly discuss the results from those deployments that lead us to the implementation of the WATTSBurning system. A detailed explanation can be found in the corresponding reference.

First Deployment.
In the first deployment the research team designed an eco-feedback interface based on an evaluation of commercially available systems. The system presented real-rime and historical consumption using bar charts and involved three modes of operation: idle, attention and detail. These modes were triggered by a camera (using face and motion detection algorithms), which were also used to log user activity. The initial system was deployed in 21 houses for a period of 9 weeks. The data showed a 9% average aggregated decrease of consumption. A deeper look into the data disclosed that families that used the eco-feedback system more often had higher decreased consumption. Nevertheless there was a steep decrease in the users' interested in the eco-feedback, particularly after the first 4 weeks (users interest was measured by the amount of access to the different application features) [18].

Second Deployment.
Based on the findings of the first study, the feedback system and sensing infrastructure was completely redesigned and revised. The new eco-feedback interface was developed with the help of a designer following a set of "guidelines" extracted from the research literature. We conducted a 52 weeks long-term study using the revised system deployed in 13 houses starting from the first sample [19].

Results from the second study revealed that initially people interacted with the system more than with the previous one. However, the frequency of the interaction started to decrease after two weeks, and again the steepest decrease happened after four weeks confirming our initial findings [18]. Detailed analysis of interaction data revealed that after four weeks, some users even stopped checking the system altogether. A few others kept using it but less frequently (only once or twice a day). During this second deployment users' consumption remained virtually unchanged from the first deployment, suggesting that savings come from the initial understanding of consumption patterns.

3 Design Rationale

The experience gained from the studies described above suggested that a different approach was required to overcome the decreased interest of users in energy data. Using inspiration from artist we brainstormed ideas and decided to change strategy and test a new paradigm of eco-feedback. We postulated a more inspiring and emotional visualization towards a less information driven eco-feedback system. After brainstorm sessions with artist designers and engineers it was decided to implement the feedback based on mapping the energy consumption of the household with elements of the local landscape, with the goal of leveraging the emotional connection of families with the local natural patrimonies, which is home of a UNESCO heritage forest. It is important to note that most of the inhabitants already feel a strong connection with the local forest. This was visible after the tragic mudslides in 2010 and the forest fires in 2011. A thorough description of this process can be found in [20].

3.1 Pilot

To test the concepts described above a novel interface was developed for the eco-feedback system. It consisted of a video based animation of a well-known forest site. Based on the consumption level several elements of that landscape would change. The electricity real time consumption was mapped as the movement of the clouds in the background, and by adding and removing animals in the landscape. More consumption meant that the clouds would pass by faster, and more appliances turned on or off meant more animals in the forest. This was a neutral feedback since more appliances being turned on or off does not necessary means more consumption, and the movement of the clouds does not have any direct negative meaning. Figure 1 shows these how these two feedback modes where displayed in the application.

This new eco-feedback version was tested in eight households during one month. Consumption and interaction data was saved and four of the eight families were interviewed. Quantitative data revealed no significant differences in the energy consumption of the families neither prior or during the study. However, families using this new version of the system had an increased number of interactions when compared with families that used a "traditional" eco-feedback device during the same period. Nevertheless there was still a decrease in the interaction along the period of the study, consistent with the response-relapse behavioral pattern reported above. However, this was not a linear decrease as there were several peaks in interaction during the month.

Qualitative analysis of the interviews revealed that neutral feedback was not well understood by users. Users that were exposed to previous versions of the feedback also missed out on the quantitative data provided previously:

"No I didn't relate it to my consumption levels. I wasn't even near to realize that (laughs). I was looking for data, logical data about it (...) I could see some extra elements, I thought you were decorating the landscape but I didn't understand why were they being placed there" Family 4 mother.

"This one is simpler, there's just the image and the consumption on the right. In the other one I could see the consumption in terms of the whole day and this was more elucidative. I find this one more interesting but we need some kind of heads ups about

how can we see the consumption throughout the whole month, the whole week" Familiy 3 husband.

On the other hand after the artistically inspired mappings of the energy consumption to the landscape were explained to the users (during the final interview) they found this new feedback paradigm interesting, and suggested merging the landscape visualization with the quantitative data display:

"I think both are valid. Maybe there could be a symbiosis between the two. This one is more pleasant the other one is just data. What matters to me is to have the data, it's probably the most important for me. However, I like the way this one is presented, I feel it's more interactive than the other one" Family 3 husband.

Three of the four families interviewed suggested that the concrete data provided by the first two versions should be merged with the aesthetic pleasure of the natural landscape shared this view.

Fig. 1. Screenshot of the application used in the pilot. At left there's the main view of the forest, the consumption is mapped as the movement of the clouds in the background. At right there is the landscape with elements added based on the appliances used.

3.2 Refinements of the Feedback System

The experience from running the pilot of the WATTSBurning, made it clear that users found the feedback interesting, but they still wanted the concrete data about the consumption. Additionally the mapping between the forest and consumption was not fully understood and some users needed further explanation of how it was represented.

In the WATTSBurning system we also wanted to address some limitations found in the two previous deployments. In the initial deployments, the feedback was given via the display of a small netbook installed behind the main door of the house (where the main fuse box is located in most of the houses) [21], this type of stationary feedback is clearly limited since it is not accessible to all the household members at the time of decision (e.g. turning off a high consumption appliance). Also the fact the system was connected to the main fuse box made some families worry about the safety of the device, and sometimes would not allow children to use it. Furthermore some additional requirements emerge form the state of the art. The feedback should be accessible to all family members since the family dynamics and communication have an impact in the decision [22]. The system should allow the comparison between different periods (hour, day, weeks and months) so that people can explore and better understand their consumption patterns. The system should also provide simple tips about energy conservation and best practices promoting sustainable behavior change. The hardware itself should also be aesthetically pleasant, as Petersen [23], argues that pragmatist aesthetics is a promising approach for designing interactive systems as it promotes aesthetics of use, rather than aesthetics of appearance. Aesthetics play an increasingly important role in interaction design, in particular when designing for homes and everyday lives rather than for the workplace.

Collectively our prior experience and these requirements meant that the new version had to undergo major updates both on the software and in the hardware side. The eco-feedback device should ideally be portable so it can be accessed anywhere in the house (or outside). The sensing framework should provide data remotely to the eco-feedback device so that the sensing is removed from the house, addressing users safety concerns. The visual eco-feedback should be aesthetically pleasant while still providing concrete consumption information.

3.3 System Design

In order to provide accurate and meaningful eco-feedback, the system needs to measure the energy and resources consumption effectively. Measuring energy/resources consumption is in itself a challenging research problem. Researchers are striving to measure energy consumption in more cost-effective, accurate and less intrusive ways. One of the most promising research approaches is non-intrusive load monitoring (NILM), which reads data from a single point and tries to monitor and desegregate the consumption per appliance. The main assumption of most of the exiting NILM approaches is that every change in the total load consumption of a household happens as a response to an appliance changing its state. Therefore specific appliances can be isolated and their individual consumption calculated using complex signal processing and statistical learning techniques [18]. Low cost and non-intrusiveness are the main advantages of the NIML approach thus it was chosen as the building block of the sensing infrastructure that supports this research [21].

3.4 Implementation

In this section we explain how the requirements described in the sections above were implemented. Firstly we explain how the sensing framework evolved to the current version. Then we explain the implementation of the front end eco-feedback visualization of the system.

Sensing Platform/Framework.
Our sensing and eco-feedback platform is based on a custom made NILM system [18,21]. Our research was target at real world deployment of low-cost eco-feedback systems that are capable of sensing and disaggregating energy data in households. Our technical requirements were to find a low-cost solution that could run different NILM algorithms and eco-feedback visualizations while also collecting relevant user data. Originally our system was based on a netbook that acted as both as the sensing infrastructure (reading current and voltage from the microphone jack) and the visualization

(via the built-in screen) and human sensing (via the built-in camera) system. Our initial prototype provided an easy way to deploy the system and evolve the visualizations during the first deployments. However, it was limited in terms of location (next to the mains at the entrance) and required installation inside the homes and unreliable Internet access. A second version of our system was implemented using a more capable DAQ (Data Acquisition Board) at the entrance of apartment buildings, hence measuring the energy consumption of multiple houses from one single location. This new version was less intrusive and enabled data to be pushed via web-services to different visualization platforms inside the home (web, tablets, etc.).

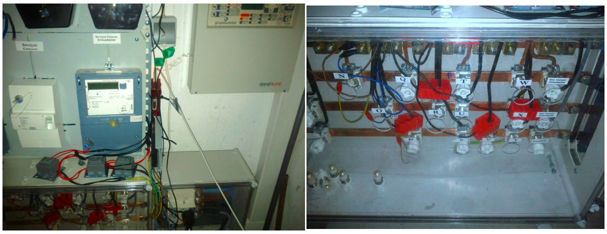


Fig. 2. Shows the system as it was deployed in one of the buildings of our study. At left there is the DAQ Board. At right there is the current clamp sensing each house.

Figure 2 shows the last implementation of our multi-apartment NILM sensing system. On the left hand-side one can see the three voltage sensors and the DAQ board, on the right it is possible to see current clamps installed on the current conductors of the apartments monitored. All of these signals are acquired and processed by a single computer that also runs the NILM algorithms and provides the data for eco-feedback devices via restful web services. The system also stores all the data in a database and provides access for consolidation in a datawarehouse.

3.5 WATTSBurning Eco Feedback Visualization

Following the requirements described before we implemented a new hardware and software platform but also modified the eco-feedback component, which now could be deployed on different portable devices. The new visualization addressed our experience in pilot (section 3.1) producing an aesthetically pleasant landscape that mapped the household consumption to elements of the natural forest. We also provided a way for users to access a second layer of more detailed consumption information.

The eco-feedback visualization was implemented on a 7'' android tablet using the android native SDK (Figure 3). The tablet specifications allowed us to implement rich visualization, which can be accessed in different places of the house, or even outside. We believe that the device itself (a tablet) was still viewed as a novelty, and that helped our application to fit in the household as well as encouraging interaction.

The application receives real-time consumption and historical data from the sensing framework. The real time data is received using sockets and a custom made communication protocol. Historical data is gathered by using the web-services described in the previous sections. The historical data is also stored in the tablet so the users can still check their past consumption without an Internet connection. The transitions between the different modes of operation are stored locally enabling the analysis of the usage patterns by the research team. For example if a user picks up the tablet, presses the back button to go to the home screen and selects the daily consumption, the application will store 3 interactions with 3 different ids reffering to the 3 different views accessed by the user.

Fig. 3. System installed in one of the households

Our eco-feedback system involves two main modes of operation. When it is not used for two minutes it goes into the *Energy Awareness* mode that shows the consumption mapped as a digital illustration of the local endemic forest. Once the user interacts with the tablet, by pressing the back soft key, the system goes to *Detailed Consumption* mode and shows daily, weekly and monthly information about the home energy use. It is important to point that *Energy Awareness* was the default mode of the system, the tablet never went to an idle mode like it happens by default in some android applications.

Fig. 4. Different landscape possibilities used during the think aloud process

Energy Awareness Mode.
To select the landscape and the metaphors used in this mode we performed a think aloud with two families (these families were later recruited as participants for this study). In these sessions the family members were presented with several paper prototypes displaying the consumption in different landscapes and represented with different items in the forest as seen in Figure 4. We decided to use a landscape of a

well know forest site that was easily recognizable by the local community. The digitally modified pictures of the forest represented the comparison between the real time consumption and an average baseline consumption level. After the think aloud session we rejected some concepts like displaying the historical consumption as items in the landscape (e.g. moss growing on a tree) or the real time consumption mapped in the movment of an animals, these concepts were not clearly understood by the users.

In total there were five levels of consumption represented in the forest (as shown in Figure 5). These five levels represent when an household consumption is slightly above/bellow, well above/bellow or belongs to the average baseline. The baseline was composed of an average of the consumption on that period, for example the real time consumption on a Monday at 12:20 was compared against an average of all the Mondays during the period between 12:00 and 13:00. Aditional ilustrations of the forest are used to ensure a smooth transistion between the states, however the animation only stops in the five levels aforementioned.

Fig. 5. Different views of the landscape according to the consumption. Ranging from low consumption (Image 1) to high consumption (image 5).

Detailed Consumption Mode.
This mode is triggered when the user presses the tablet back button. As a consequence the system presents a tabbed menu with four options: "Home", "Day", "Week" and "Month". The "Home" tab shows a summary of the overall consumption as well as the current real-time consumption (Figure 6 Left). The summary contains aggregated consumption of the current day/week and month, and comparisons between homologous periods. Also in this tab the user is presented with a "tip of the day" with general sustainable actions. The "Day" "Week" and "Month" tabs (Figure 6 Right) present a chart displaying the consumption over that period and the total aggregated consumption. It also informs the user of where the peak consumption happened and how it compares to the average of that period (for example in the "Week" tab the system shows how the consumption in the current week compares with an average of the previous weeks). By default the information presented here refers to the current day/week/month but the user can select preceding periods.

Fig. 6. Left: Home screen of the system. Right: Tab with the consumption of the current week

3.6 Collected Data

Our sensing framework samples the current and voltage waveforms with at a high sampling rate (about 3.2 kHz) but due to database size constraints only average power consumption is stored in a database (in 30 seconds sampling intervals). Additionally the application stores locally every transition between the different modes of operation of the eco-feedback system which is then periodically uploaded to one of our servers, hence allowing us to keep track of how the system is being used without having to wait for the end of the deployment.

The data analysed corresponded to two weeks of baseline consumption data collected when no eco-feedback was available and 15 weeks of consumption data after the eco-feedback was installed, as well as all the user interactions with the feedback during that time.

4 Evaluation

To test our system we recruited users from an apartment complex in an urban area in southern Europe. All the apartments were relatively new and they had similar appliances (provided by the construction company), consequently all of the families were relatively new to their houses. The sample is composed of six households: five couples with one or more children, and one young couple with no children. Two of the families had to leave the study earlier, but in both cases this was related to the fact that the families had to leave their apartments and not because they wanted to abandon the eco-feedback study. The families were told to use the tablet as they wanted (for example for browsing the web). But only the interactions in the WATTSBurning system were recorded, as described in the previous section.

4.1 Method

In order to assess the changes in consumption we collected baseline consumption data before the eco-feedback system was installed. This data was collected between the 13th and the 27th of August. After that each family received a tablet with the WATTSBurning application installed and given a short explanation of how to access the consumption data. After 22 days we visited the families and performed a short semi-structured interview. The families continued to use the system until the end of the year. Here we analyze the consumption and interaction data until the 17th of December for a total of 17 weeks. We wanted to avoid the Christmas period, since the results in that period would most probably be biased.

4.2 Qualitative Assessment

When the families were interviewed after the first three weeks, we wanted to understand how the system was received, and if the *Energy Awareness* mode was being clearly understood and creating an increased awareness about energy consumption. It was also important to verify if any behavior change was triggered by the presence of the eco-feedback system. Five of the six families were interviewed, they will be referred as F1, F2, F3, F4, F5. We asked all the family members to be present in the interviews, in order to gather every member' opinion about the system. However, for F2 and F3 only the husbands were present during the interview.

The system was well received by all the families, and none of them had any major problems using it.

"I think it was simple, even the wireless connection was stable" F4

All the families agreed on the fact that the system increased their awareness about electricity consumption.

"It's raising my awareness, I don't think I've changed my consumption patterns yet, but I'm more aware now". F1

"… if we have more devices turned on we can see right away there is an increase" F2

"It provides us with immediate feedback, such as daily and weekly consumption, and we can see what we do and how we behave in our daily routines to reduce our energy consumption" F3

"I got more alert, so necessarily I will try to consume less" F5

Another observation transversal to all of the families was that every family learned something about the consumption of certain devices.

" In the weekend for example, we were using the oven and I noticed it consumed a lot, then I turned on other devices out of curiosity" F1

"…the electric stove for example, I learned from the system that it consumed a lot" F2

"Especially the oven, it increases to 2000/2500W… I see there that it goes to 200 ºC but I wasn't expecting so many watts… I was even surprised with the TV's" F3.

"I've learned that the oven consumes a lot, also the fridge" F5

It was also important to verify if the energy aware mode of the application was well understood. All families found that it was easy to relate their current consumption with the animation of the forest drying up and eventually being set on fire. It was also mentioned that the picture showing the state of the forest in the feedback interface worked as an alert of what was going on in the house, and it made them more aware of their consumption at specific moments.

"Yes but…. For example when the dryer is working I see the forest catches fire, but I was expecting that, but I knew I couldn't keep that consumption for a long time, maybe I got more conscious" F1

"It gets dryer until it catches fires, when I have a lot of things turned on (…) also when there were only a few things on it showed rainbows and butterflies" F2

"(…) there's an association, the bigger the consumption the more destruction is visible in the forest, when we see it goes from green to yellow to red,… it's scary," F3.

All the families opted to place the system in a central place in the house, where it could be visible to all the family members. However in the five families that were interviewed it was mostly the adult male and young children who used it more often. "*… I was curious in the beginning to see the forest on fire, but it was mostly my kid"*F2

*"It was mostly me and my son, my wife and my daughter didn't pay much attention to it, it was me and my younger son"*F3,

"It was mostly me, she (girlfriend) wouldn't use the system".

Although the system was usually located in a central visible place in the house, it was common for participants to move it around, mostly because some of them wanted to check the consumption of a device in real time.

"It was mostly there, but it was in the kitchen for some time" F2

*"It was there close to the sofa, because it had better reception for the internet (…) but I would also take it to the balcony while sitting there"*F3

"I used it in different places in the house, I thought of it as one of those portable weather forecast displays that can be placed anywhere". F5

4.3 Quantitative Assessment

From the qualitative assessment users reported a noticeable increase in their energy awareness. However, it is important to confirm if the reported awareness translates into an actual behavior change, confirmed by the quantitative data acquired from our sensing platform. In the following sections we analyze and discuss this issue.

Energy Consumption

In order to compare energy consumption we use the week as the standard period of time. This is the unit that best spans the routines of a family impacting their energy consumption. For instance, some families organize the major cleaning in a single day of the week, others the ironing. The week also comprises the working days and the weekend that usually correspond to very different routines. Therefore we compare the consumption data in the 15 weeks of eco-feedback deployment with the average of the two weeks of baseline data.

If we consider the 6 households as a whole, after 15 weeks the average weekly consumption dropped on average 2% from the initial baseline (n =6, SD=7.99%). However, the standard deviation suggests that the savings were not constant across the houses. As a consequence we analyzed each house individually. Three families (F1, F2, F5) reduced their consumption by 5%, 7% and 13% and their consumption was below the baseline for long periods – 12 weeks for F1, 11 weeks for F2 and 5 weeks for F5. The other three families consumed on average more than the baseline , respectively 10% (F3), 1.5% (F4) and 2.5% (F6). These 3 houses were below the baseline average during 3, 6 and 6 weeks respectively. Figure 7 shows how each family's consumption changed during the study.

Fig. 7. Consumption in the 6 households during the study. The baseline period is highlighted in red. The horizontal axis represents the week in the study, and the vertical represents the average consumption in kWh.

Interactions with the System

The analysis of the logged user interactions (the saved interactions are defined in section 3.6) showed that families used the system in completely different ways. Two of families (F3,F6) used the system more on the Mondays ,F1 and F4 used the system more on Wednesdays, F2 used the system mostly on Fridays and F1 on Tuesdays. This diversity was also noticeable when looking at how families used the system throughout the day. However, in al the families the view that was most triggered was the summary, followed by the animation view.

During the 15 weeks of the study a total of 1577 interactions with our system were logged among all of the households. That value represents an average of 266 interactions per house and an average of 16.3 interactions (N=1597, SD=29) per day and 4.2 interactions per house and per day (N=6, SD=6.06). Our system was mostly used in the afternoon (between 13:00 and 19:00), almost a third of the interactions happened during this period. It was also noticeable that all the families had a lot of interactions on the first four to five weeks of the study. After that the number of interactions dropped significantly (by more than two a thirds). Again the pattern was not similar in all houses, which justified the high standard deviation. Table 1 shows a summary of the number of interaction with the system, in the first four weeks of the study and rest of the period and in total. We choose to isolate the first four weeks because our previous research suggests that that after this period there is a steep drop in the usage of the feedback devices [18,19].

To better understand the usage patters we analyze each household individually. The number of interactions on F1 only decreased by less than half, they kept using the system with an average of 14 interactions per day after the initial four-week period. Similarly with F2 the decrease in interaction was not considerable, but it was noticeable that they stopped using the system on a daily basis. Families 3, 4 and 5 had a lot of interactions in the first week, after that the interaction values dropped to less than one interaction per day (0.6 ,0.4 and 0.2 per day respectively). F6 simply didn't interact directly with the system, even though this family agreed to participate in the study, they never showed much interest regarding the system nor they were ever available for interviewing.

We also analyzed the correlation of the user interactions with the energy consumption. The three households that on average didn't reduce their consumption are among the households that interacted less with the system (F4, F5 and F6). This finding is also consistent with our previous research.

Table 1. Summary of the average of interactions with the system by day, in different periods of the study

Family	*Average interaction by day*		
	First 4 weeks	Rest of the study	Total
F1	23.2	13.9	16.2
F2	13.3	1.2	4.2
F3	5.3	0.4	1.6
F4	2.1	0.6	0.9
F5	3.8	0.2	1.6
F6	0.6	0	0.4

It is important to note that the system only logged direct interaction with the interface. It would be very hard to count the number of time the users looked at the tablet with the *Energy Aware* mode on. Furthermore users confirmed that a significant amount of interactions were done with the tablet in the *Energy Aware* mode.

"It worked as an alert for me (...) it was easier to see from a distance (…) when we started to consume more it would get darker and catch fire" F5.

5 Discussion

The presence of eco-feedback increased users knowledge about the devices that they had at home. In fact all of the participants learned something about a particular appliance, this indicates a rise in the awareness about energy consumption, despite the level of usage or the change in overall consumption. However, the decreasing interest for eco-feedback after several weeks is an important factor leading to the relapse effect. Our novel eco-feedback system tried to overcome this issue by trying to map energy consumption to elements of the natural landscape. From the interviews with users we can conclude that the mapping the consumption with images of the local forest landscape was clear and well understood by participants. Although users didn't mention an emotional connection with the illustrated forest landscape, the energy awareness mode, where the forest dries up as more energy is consumed, did work as an alert for when consumption was higher than normal. Some users even found the forest on fire for a long time a disturbing factor that would motivate them to investigate which appliances were responsible for the higher level of consumption. Also, displaying the eco-feedback in a portable device allowed users to move the device through the household and explore the consumption of different appliances.

In terms of usage the system had a lot of interactions during the first four to five weeks, after that period the number of interactions decreased. However, unlike in our prior research the reduction was not linear and here were several peaks until the count finally settled at a low value. We believe this was a consequence of placement of the eco-feedback but also the presence of the energy awareness mode, which reminded people of the long-term consequences of their daily actions. The results in terms of user-interaction are an improvement over previous studies with classic quantitative forms of eco-feedback, since most families kept using the system after the four/five week period although less frequently and with different patterns. This difference in how families use the system is inline with the "one size does not fit all" [24] argument for eco-feedback systems. If we relate the quantitative interaction data with the consumption information it is noticeable that the houses that saved more energy are the ones that used the system more.

6 Conclusions and Future Work

In this paper we describe the studies and refinements leading to the design, prototyping and testing of WATTSBurning an eco-feedback device designed to foster awareness of energy consumption in households. Our research aimed at overcoming the know problem of people relapsing to previous behavior after several weeks of exposure to eco-feedback. Through an iterative design, testing and refinement process we improved our eco-feedback system introducing a new artistic metaphor that combines energy consumption levels to artistic representations of the local forest landscape. Our goal was to verify if this novel metaphor would improve on significant reduction of interaction with eco-feedback after four weeks of deployment. After initial prototyping that removed some ambiguity in the mappings of consumption to natural elements, the WATTSBurning system was successful in improving the levels of user attention and usage. Most families kept using the system even after four to five weeks, although to a lesser extent. The placement of the eco-feedback device and the presence of an energy awareness mode showing the landscape changing was an important motivator to retain user attention and awareness over time. These findings provide a good motivation to explore new forms of eco-feedback that go beyond traditional quantitative information. Clearly without more research eco-feedback technology will not confirm the promising results coming out of short-term three-week studies as the ones published in many HCI venues. The households in our study did manage to reduce their consumption, but they showed very different consumption and eco-feedback usage patterns. Overall households that used the system more saved more energy but there is still a lot to learn about the long-term consequences of eco-feedback. We believe our attempts with non-conventional forms of eco-feedback and, in particular, when exploring emotional and aesthetical aspects is a promising path to explore further.

Our deployment is still ongoing, and the plan is to further investigate the qualitative data. Relevant results might be achieved when comparing across houses or even between different periods in the same house. Furthermore our sensing framework is storing all the power events extracted from the NILM, which can also be useful to research into users consumption and behavior patterns. Even thought the feedback system was already removed the sensing framework is still gathering consumption data and it will be interesting to observe if there is any change in the consumption after the feedback is removed.

Acknowledgments. We would like to acknowledge designers Joana and Sofia Gomes for their contribution building the illustrations of the forest used the in WATTSBurning application. We would also like to thank all the families that took part of this study. The research presented here was funded by the SINAIS project (CMU – Portugal) CMU-PT/HuMach/0004/2008.

References

1. Wilson, A., Yost, P.: Buildings and the environment: the numbers. Environmental Building News 10 (2001)
2. European Environment Agency, `http://www.eea.europa.eu/data-and-maps/figures/final-energy-consumption-by-sector-5` (retrieved on January 2013)
3. Schipper, L.: Linking lifestyle and energy used: a matter of time? Annual Review of Energy 14, 273–320 (1989)
4. Froehlich, J., Findlater, L., Landay, J.: The Design of Eco-Feedback Technology. In: CHI 2010, Atlanta. ACM (2010)
5. Mankoff, J., et al.: Leveraging Social Networks To Motivate Individuals to Reduce their Ecological Footprints. In: Proceedings of the 40th Annual Hawaii International Conference on System Sciences 2007, p. 87. IEEE Computer Society (2007)
6. Mills, E.: Review and comparison of web- and disk-based tools for residential energy analysis. Lawrence Berkeley National Laboratory, LBNL (2002)
7. Yolande, A.: Designing eco-feedback systems for everyday life. In: 2011 Annual Conference on Human Factors in Computing Systems, pp. 2135–2144. ACM (2011)
8. Mills, E.: Review and comparison of web- and disk-based tools for residential energy analysis. Lawrence Berkeley National Laboratory, LBNL (2002)
9. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design Requirements for Technologies that Encourage Physical Activity. In: CHI 2006, Montreal, pp. 457–466 (2006)
10. Pierce, J., Odom, W., Blevis, E.: Energy Aware Dweling: a critical survey for interaction design for eco-visualizations. In: OZCHI 2008. ACM Press, New York (2008)
11. Egan, C.: How customers interpret and use comparative displays of their home energy use. In: Proc. European Council for an Energy-Efficient Economy, Panel III, 19 (1999)
12. Dunlap, R.E.: Lay Perceptions of Global Risk. International Sociology 13(4), 473–498 (1998)
13. Fischer, C.: Feedback on household electricity consumption: a tool for saving energy? Energy Efficiency 1, 79–104 (2008)
14. Consolvo, S., McDonald, D.W., Landay, J.A.: Theory-driven design strategies for technologies that support behaviour change in everyday life. In: CHI 2009, New York, NY, USA, pp. 405–414 (2009)

15. Holmes, T.G.: Eco-visualization: combining art and technology to reduce energy consumption. In: Proc. of the 6th ACM SIGCHI Conference on Creativity & Cognition, pp. 153–162. ACM Press (2007)
16. Nuage Vert Nuage Vert Project (2008), `http://www.pixelache.ac/nuage-blog/` (retrieved July 3, 2008)
17. Peschiera, G., Taylor, J.E., Siegel, J.A.: Response– relapse patterns of building ccupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data. Energy and Buildings 42, 1329–1336 (2010)
18. Nunes, N., Pereira, L., Quintal, F., Berges, M.: Deploying and evaluating the effectiveness of energy eco-feedback through a low-cost NILM solution. In: Proc. The 6th International Conference on Persuasive Technology (2010)
19. Pereira, L., Quintal, F., Barreto, M., Nunes, N.: Understanding the Limitations of Eco-feedback: A One Year Long-term Study. To appear at 1st SouthCHI (2013)
20. Nisi, V., Nicoletti, D., Nisi, R., Nunes, N.: Beyond eco-feedback: using art and emotional attachment to express energy consumption. In: Proceedings of C&C, pp. 381–382. ACM (2011)
21. Pereira, L., Quintal, F., Nunes, N.J., Berges, M.: The Design of a Hardware-software Platform for Long-term Energy Eco-feedback Research. In: Proceedings of The ACM SIGCHI Symposium on Engineering Interactive Computing Systems (2012)
22. Barreto, M., Karapanos, E., Nunes, N.: Social Translucence as a theoretical framework for sustainable HCI. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 195–203. Springer, Heidelberg (2011)
23. Petersen, M.G., et al.: Aesthetic interaction: a pragmatist's aesthetics of interactive systems. In: Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 269–276. ACM, Cambridge (2004)
24. He, H.A., Greenberg, S., Huang, M.E.: One size does not fit all: applying the Transtheoretical Model to Energy Feedback Technology Design. In: CHI 2010: Sense and Sustainability, pp. 927–936. ACM Press (2010)

Finding-NEVO: Toward Radical Design in HCI

Sharon Lynn Chu, Francis Quek, Yao Wang, and Rex Hartson

Center for Human-Computer Interaction,
Virginia Tech
{sharilyn,quek,yxw5167}@vt.edu, rex.hartson@gmail.com

Abstract. We address the methodology of design-oriented research in HCI, whereby researchers design and implement technology to test concepts. The task is to produce a testable prototype (that we call *NEVO, Non-Embarrassing Version One*) that faithfully embodies the concept. We probed leading HCI researchers and CHI authors about the challenge of *Finding NEVO*. We found uncertainty on how to design prototypes that allow for both design and scientific contributions. We propose the *Finding-NEVO* model that articulates a process yielding prototypes that are faithful to the rationale and idea being studied. We conclude by discussing our theoretical and methodological contributions.

Keywords: Radical design, design method, innovation, HCI.

1 Introduction

HCI is an interdisciplinary field with a foot each in the doors of science and design [1-3], causing an inherent tension since "[s]cience and design have different principal objectives" [3]. While design values innovation and is generative in nature [4], science builds on prior advances in verifiable steps. Approaches like *theory-based design (TbD)*, *design-oriented research* (*DoR*), and *research through design* (*RtD*) have been proposed to resolve this conflict. In *DoR* [2], design and technology implementations serve to test and validate the research concepts. Similarly, *TbD* [3] and *RtD* [5] see the value of design in the light of artifacts embodying some form of knowledge. Conversely the design of commercial products focuses on design practice, or solving problems and real-world obstacles, that Fallman calls *research-oriented design* (*RoD*) [2]. Our research looks at process in the context of the DoR tradition in HCI.

To properly answer a research question, DoR prototypes have to embody the ideas to be tested. Yet, there is little concrete guidance on how to design such research prototypes. Our goal is to gain an understanding of prominent design processes in HCI through interviews with HCI researchers and domain experts, and to suggest a plausible process model for the design of prototypes in research.

2 Design of Research Prototypes

We employed a general model (Figure 1), as a probe for interviews, of how research prototypes are designed. This model is similar to Gaver's [4] idea of designs occupying points in the design space or creating a design space around themselves. The HCI

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 471–478, 2013.

researcher advances a concept from some *conceptual rationale* (label A), and produces a *seed idea* (B). The *conceptual rationale* may be a theoretical construct, a new technology, a research intuition, or a user need, and the *seed idea* is an instantiation of how the theoretical construct may be embodied in a system. This *seed idea* is situated in the *space of possible designs* of the domain in question (C), which may be populated by a set of *existing* designs (D). If the *conceptual rationale* is novel, and the *design space* is not densely populated, the new *seed idea* would likely be far from *existing designs*. We call such a design a *potentially radical design* (E) because of this difference from existing solutions. The *seed idea* does not fully specify a testable system. Many *pathways* of design choices or decisions have to be made to develop a system (F). Some of these design choices flow directly from the *seed idea*, and some may be necessary to realize the system but do not find guidance from the *seed idea*. All of these design choices have potential impacts on the testing of the *conceptual rationale* of the research. The term ***NEVO*** (*Non-Embarrassing Version One*) designates the first testable prototype that faithfully embodies a concept.

There is an inherent tension, that we seek to address, when realizing the *design pathways* (F). This tension can be characterized by different types of 'creeps' (G) that threaten fidelity to the original conceptual rationale. Prior experience with existing systems by designers, developers, and study participants introduces forces to move the design to resemble existing solutions. We call this *experience creep*. Also, our tendency to resort to the most convenient way for implementation and design using tools and frameworks at hand can introduce *convenience creep*.

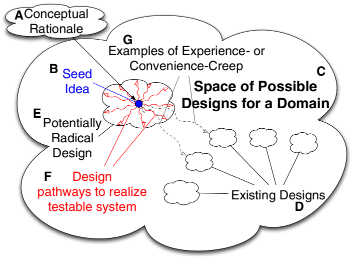


Fig. 1. Going from concept to testable prototype

3 Contextualizing Our Work

Design of Research Prototypes: There are few well-defined, systematic process methods to develop good *research* prototypes. We briefly describe two frameworks that are particularly relevant to the problem of finding *NEVO*. Carroll's & Kellogg's [3] *theory-based design (TbD)* proposes that successful HCI designs embody psychological claims in contexts of use, and advances that claims coalescing together in the implementation of a system projects "a model world", "a believable illusion" to the user. The careful recording of this 'design rationale' can inform the design of future artifacts. Keyson and Alonso's [6] *Empirical Research through Design method* embeds interaction design hypotheses into working prototypes to contribute to design knowledge by creating experimental variability. The method lists several broad guidelines (e.g., "prototype variability has to be carefully defined so as not to confound the research question at hand"), and specifies the use of design iteration and techniques, but presents no clear process model.

Process for Radical Design: Work on radical innovations tends to come overwhelmingly from the management sciences. Many either emphasize the importance of the individual with passion and vision, the "human side" (e.g. teamwork, networks, roles [7]) and organizational mechanisms (e.g. rewards, management policies), or advocate the approach of concept refinement through the involvement of users (e.g. [8, 9]). Management science is concerned with the potential of a product to be adopted by users, thus defining radical products "in the sense that they imply changes in consumers' everyday lives" [9]. In contrast, our concept of radicalness in HCI research is that it is valuable because it is an extension to knowledge in and of itself.

Closely aligned with Figure 1, a 'hill-climbing' model was proposed by Norman [10]. He characterizes the design space for a particular domain as being occupied by multi-dimensional 'design hills'. A particular design sits on the slope of a hill, and ascending the hill constitutes a design improvement. Human-centered design (HCD) approaches allow designers to climb a particular hill by an iterative incremental process of design improvements driven by lessons learned from user studies. Norman states that while HCD enables a design to ascend its hill, it cannot move the design to another hill with a higher peak. Such a jump to a different hill would constitute a radical design shift. HCD is thus "only suited for incremental innovation", to improve existing products according to the user's contextual needs. Norman & Verganti [10] argue that potentially radical *seed ideas* have to be driven by technology or meaning change. For radical innovation, they suggest that HCD must admit the "simultaneous development of multiple ideas and prototypes" (essentially many *seeds*).

4 Study Methodology and Data Analysis

We carried out hour-long semi-structured interviews with two sets of HCI people: 1. *Meta-Interviewees* (MIs) and 2. *System Builders* (SBs). The MIs are leading researchers from prominent HCI research labs throughout the world who have done substantial conceptual or theoretical work in design methodology. The SBs, researchers who develop technology using the DoR approach, were selected by combing through CHI papers published in the last decade. The selection, done by three members of our team, identified papers that 1. Adopt an approach where a system is built to investigate a research question; 2. Have a recognizable *seed idea* and an explicit or implied design process; and, 3. Present some form of user testing. The selection process yielded 97 papers across HCI domains (e.g. accessibility, web search, ubiquitous computing, embodiment, social interactions, 3D interaction).

An email request for interview was sent to the first authors of all the papers. If we did not receive a reply, a request was sent to the second authors of the papers. Three local researchers (2 SBs and 1 MI) were used as pilot interviewees. Twenty-two HCI researchers, 20 from university research labs and two from industry were interviewed as SBs. Eight leading HCI researchers were interviewed as MIs. Five of the eight MIs also did the SB interview. Our interviewees were from the US, Canada, the UK, Japan, Singapore, South Korea, and Australia. Eighteen participants were interviewed through video teleconferencing, two by telephone, and two in face-to-face sessions. Oral consent for audio recording was first obtained from each interviewee.

The SB interview guide comprised 32 questions (multiple-choice survey and rating, and open-ended questions) on four main themes: 1. *Description of the research undertaken specifically in the CHI paper*; 2. *Methods used to design the system*; 3.

Testing of the system; 4. *Design in HCI in general.* The MIs' interview was divided into two sections: 1. The first six questions probed on research of the MI in general, and on her perspective on design in HCI. 2. We explained the purpose of our study using Figure 1. Nine questions guided the discussion and prompted the interviewees about design processes to move from an idea to a research prototype and the role of user testing in radical design. We asked the permission of some MIs to quote their responses non-anonymously, and they were given the opportunity to review the paper.

All interviews were transcribed and coded for themes of importance in several iterative passes. The research process used by each interviewee was identified. A preliminary process model was constructed through a first high-level analysis. We refined our model in subsequent iterations through deeper-level analyses and discussion of the insights gleaned. Survey data collected in the interviews were collated on a spreadsheet and summarized. These helped us to get a sense of how each component of our model was manifested in each project. We present a few relevant study findings below in a synthesized account of the qualitative and survey data.

5 Findings

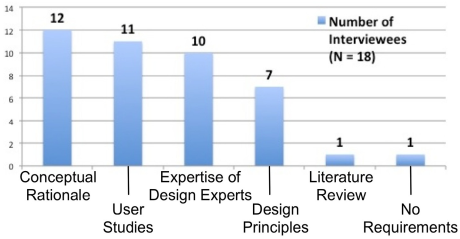


Fig. 2. Basis for design decisions

The basis used by most of our SB interviewees to make decisions on what to include in a design was their own *intuition* and *experience* with their original *conceptual rationale*, or by generating a list of design requirements based on *studies of users*. Figure 2 shows the different categories of how researchers made design choices.

However, none of our SBs could provide a reliable method of how to interpret study results (e.g. unexpected study results could be because the original design idea is unsuitable for the particular user group or the system was improperly implemented). All interviewees who provided an answer said that an *iterative design* process was needed, and some said that using *low-fidelity prototypes* for user studies could enable the researcher to make an informed guess "based on the reaction we got from people". Elements of *intuition and experience* were also put forward by some MIs as key to making judgments about study results.

All MIs agreed that there is currently no process to guide researchers on how to embody a *seed idea* in a *design*, resulting in many ineffective designs. The MIs' perspectives can be summarized into six 'high-level strategies', all pointing to the uncertainty inherent in the ad-hoc way in which HCI research is conducted: **1.** One needs to try "existing solutions first so that one doesn't innovate unnecessarily. When you fail with existing solutions, that's when you innovate and fly." (Scott Klemmer, Stanford U.) **2.** One should "stick" to one's idea and persist: "You simply have to face your own ideas and go forth and most of the time you will fail. But the few times you succeed make up for the failures." (Norman, Norman Nielsen Group) **3.** One should generate many designs and test often: "The principle is to try to do this (minimal prototyping) with little effort as possible. So you can try a lot of designs" (Jacob, Tufts U.). **4.** One should make use of intuition, skills and experience. But for Norman, "this is where the huge experiences and skills and intuitions of the designers come in. Is that a method? No." David Frohlich (U. of Surrey) contrasted "inspiration for design" and "requirements for design". In the former, one may "generate ten times more designs" than the latter which employs UCD to produce just one or two from the requirements. Frohlich states that intuition is the only way to bridge these two approaches. **5.** One should be pragmatic and be flexible to adopt any method that makes sense without limiting oneself to a "fixed methodology". **6.** One should persist based on faith. Norman suggested that the only evidence that the researcher needs to move forward with her idea is to see (from test results) whether at least "one or two people really believe deeply in what you are doing".

6 Proposed Model

Through iterative modeling of current methods we extracted in our interview data, we derived a model (Figure 3) of how design decisions (Label F in Figure 1) may be made to better arrive at NEVO. As in Figure 1, the radical design process begins with a *conceptual rationale* in the form of an integrated *Research Aim, Motivation or Rationale* from which a *seed idea* (A) is generated. A set of *idea-defining characteristics* is produced (B) to guide the design process to further specify the original concept. Exhaustive articulation of these *characteristics*, however, is not always possible. More importantly, over-specification of *idea-defining characteristics* may even have the undesirable side effect of over-constraining creativity in the design process. The model addresses this by recognizing the need for an informal *Team Understanding and Consensus* (B). Together, the *Idea-Defining Characteristics* and the *Team Understanding and Consensus* embody the *seed idea,* which then guides the *Gatekeeper* process. The *Gatekeeper* process (D) is used to vet individual *design ideas* in advance of development. These *design*

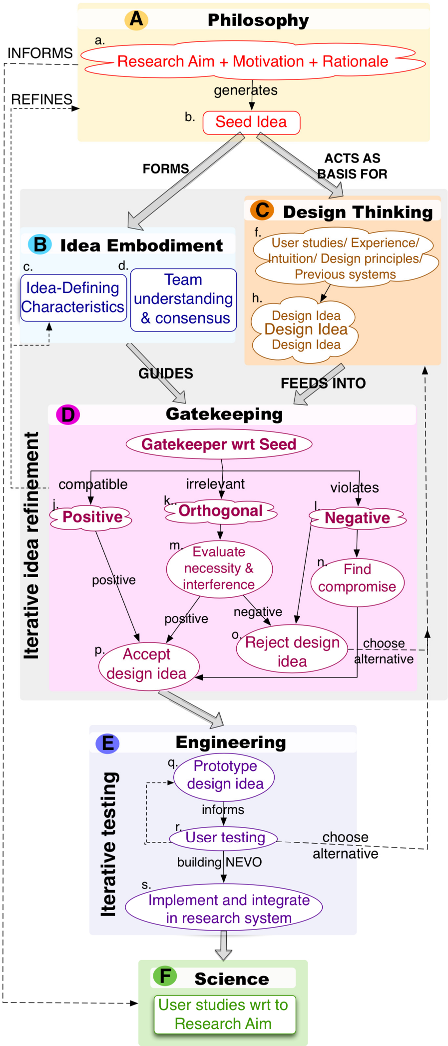


Fig. 3. Finding-NEVO model

ideas (C) may be the product of methods such as brainstorming and ideation, with respect to a particular design issue (which may be identified through methods like ethnographic studies, design principles, experience or simple intuition).

Our key contribution is in the insertion of a formalized 'evaluation' at block D into the accretion process of the *seed idea* into a full design. We do not prescribe any hard and fast way to implement the *gatekeeper*. This could be consensus among a group of stakeholders, the judgment of a single individual, or discussion in the research team. Each *design idea* may be determined by the *Gatekeeper* process to be: *positive* (it is in line with, supports or even extends the specified *seed idea*) or *negative* (it violates the premises of the *seed idea*). In the event of a negative judgment, either an *acceptable compromise* is found (e.g., owing to implementation and other pragmatic reasons such as time limitations, a modified version of the design idea may be judged the best alternative), or if the violation is so severe as to contradict the original research question and no compromise can rectify this, then the design idea is rejected.

Additionally, the *gatekeeper* may decide that the *design idea* is orthogonal to the *seed idea*. Here, a second choice needs to be made. In DoR, an unnecessary design component can introduce confounds into the study results. If the design idea is a necessary aspect of the system, the process proceeds to development, if not, it is rejected. The *prototyping process* (E) is typically done together with *iterative user testing*. As discussed earlier, this is an incremental process, that in Norman's parlance, climbs the design hill on which the *seed idea* is planted [10]. The result of this process is a system (*NEVO*) that can be used to answer the original research questions in formal user studies (F).

7 Discussion

Our model is distinct from previously proposed models in a number of ways. *TbD* argues for the use of psychological theories to inform the design of artifacts; the main aim is to produce new artifacts. In contrast, the purpose of *Finding-NEVO* is to reveal knowledge and understanding; it uses design in the service of research. Moreover, while the focus of *TbD* is on repeatability (i.e., being able to analyze, understand and maybe reproduce artifacts), ours is on validity (i.e., how to stay true to one's idea when testing it through the use of artifacts). *TbD* then can be considered to be a *design* method (others being ideation, contextual design, etc.) that can be used in Box C in our model to generate design ideas.

Much of our model is grounded in similar assertions as those of Norman's 'hill-climbing' model. However, we believe that even if the *seed idea* is potentially radical (through meaning change, technology push or random tinkering) and sits at the base of a higher 'hill', it may or may not develop to the top of the 'hill' using incremental HCD methods because of creeps (*convenience, experience, feature and user input creeps* described before). More importantly, for these same reasons, the design may jump back to some other 'hill' of a more mature but ultimately limited design trajectory. Our approach ensures that a potentially radical *seed idea* is allowed to blossom. To be clear, our model is not inconsistent with Norman & Verganti's position. We simply suggest an additional *gatekeeper* process that may make radical designs more likely without necessarily posing HCD as irrelevant to radical innovation.

Interestingly, the five MIs who were also interviewed as SBs gave differing perspectives on HCD depending on their roles. As MIs, they were candid concerning the inadequacies of HCD in bringing about radical designs needed for *DoR*, but as SBs, they employed and justified their approaches within an HCD framework. We posit two possible reasons. First, the field is in need of new models and methodological warrants to break out of the HCD paradigm. Second, as SBs/researchers they needed to publish their work, and employing HCD may make their work more acceptable to reviewers. Both interpretations suggest that the field of HCI needs to re-evaluate its dependence on the HCD paradigm. This paper contributes by adding to this discourse.

Beside the contributions above, this paper enriches our understanding of HCI as a scientific domain, and contributes to research and design methodologies in HCI:

Contributing to HCI as a Field: HCI is a relatively young field that is still being defined. Grudin asked the question "Is HCI homeless?" [11]. Much discussion has occurred on how to come to terms with the interdisciplinary nature of HCI (e.g. [12]). Our model provides a way for different disciplines to be integrated in one HCI research project. The *conceptual rationale* and *seed idea* (Boxes A and B) can come from any domain (e.g. social sciences/humanities). Methods from the creative disciplines (e.g. art/design) are employed in Box C, *design thinking*, to generate ideas for a prototype. If technical implementation is desired, engineering knowledge is required at Box E for prototype development, testing and integration. And finally research methods from the sciences are needed to carry the project through at Box F. HCI has also found it challenging to bridge many 'gaps' [13]. As we mentioned in the introduction, the conflict between the needs of science and those of design is very evident in HCI, and yet HCI is touted as being "the science of design" [14]. We believe this paper facilitates a better reconciliation of the two.

Contributing to HCI Methodology: *Finding-NEVO* is a formalized process model based on current methodological practices of researchers and refined to enable HCI research to be done more transparently. The inclusion of a gatekeeping procedure not only serves to infuse systematicity into the research process, but also provides a roadmap for new researchers to the field to know how to proceed. Even among our interviewees, many found it difficult to articulate clearly the steps that they took in their research project. Some stated that "it was messy". Moreover, our approach can provide a common mental model of research and common terminology among project team members with differing backgrounds and perspectives. Broadly, the core contributions of *Finding-NEVO* is not only in the methodological systematicity it allows, but also in the common basis of understanding that it can provide to the field.

Gaver [4] lamented that the *RtD* approach may be seen as unscientific since the criterion of falsifiability cannot be applied: the "synthetic nature of design is incompatible with the controlled experiments useful for theory testing". The danger that threatens the scientific validity of the approach is precisely that design is generative [4]. However, setting up a constant gatekeeper in the system development process enables one to stay 'on track' with regards to the initial *conceptual rationale.* I.e., two researchers starting out with the same *rationale* may not end up with the same systems, but following our model, there is greater chance for the systems to be truthful embodiments of the original idea, leading to similar results in final scientific user tests.

8 Conclusion

The *Finding-NEVO* model for *DoR* articulates a methodology to select design ideas that yield prototypes that are faithful to a *conceptual rationale* and *seed idea*. To the degree that the *rationale* and *seed* are novel, the method is likely to produce radical designs that may impact design for the marketplace. Our model is constructed through an analysis of current practice to understand both the applicability and failings of current methodology. In doing so, we contribute both theoretically and methodologically to HCI. Future work includes the validation of the *Finding-NEVO* model across various HCI projects through longitudinal adoption.

References

1. Fallman, D.: Design-oriented Human-Computer Interaction. In: CHI, pp. 225–232 (2003)
2. Fallman, D.: Why Research-Oriented Design Isn't Design-Oriented Research: On the Tensions Between Design and Research in an Implicit Design Discipline. Knowledge, Technology & Policy 20, 193–200 (2007)
3. Carroll, J.M., Kellogg, W.A.: Artifact as Theory-Nexus: Hermeneutics meets Theory-Based Design. In: CHI (1989)
4. Gaver, W.: What Should We Expect From Research Through Design? In: CHI, pp. 937–946 (2012)
5. Zimmerman, J., Stolterman, E., Forlizzi, J.: An Analysis and Critique of Research through Design: Towards a Formalization of a Research Approach. In: DIS (2010)
6. Keyson, D.V., Bruns, M.: Empirical research through design. In: In'l Assoc. of Soc. of Design Res. Conf., IASDR 2009 (2009)
7. O'Connor, G.C., McDermott, C.M.: The Human Side of Radical Innovation. Journal of Engineering and Technology Management 21, 11–30 (2004)
8. Lettl, C., Herstatt, C., Gemuenden, H.G.: Users' Contributions to Radical Innovation: Evidence from Four Cases in the Field of Medical Equipment Technology. R&D Management 36(3), 251–272 (2006)
9. Heiskanen, E., et al.: User involvement in radical innovation: are consumers conservative? European Journal of Innovation Management 10(4), 489–509 (2007)
10. Norman, D., Verganti, R.: Incremental and Radical Innovation: Design Research Versus Technology and Meaning Change. Design Issues (2012)
11. Grudin, J.: Is HCI Homeless? In Search of Inter-Disciplinary Status. In: Interactions, pp. 54–59 (2006)
12. Carroll, J.M.: HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science. Morgan Kaufmann, San Francisco (2003)
13. Lazar, J., Feng, J., Hochheiser, H.: Inherent Conflicts in HCI. In: Research Methods in Human-Computer Interaction, pp. 9–15. John Wiley & Sons (2010)
14. Carroll, J.M.: Human-computer Interaction: Psychology as a Science of Design. Int. J. of Human-Computer Studies 46(4), 501–522 (1997)

Method Card Design Dimensions: A Survey of Card-Based Design Tools

Christiane Wölfel[1, 2] and Timothy Merritt[3]

[1] University of Munich, Amalienstraße 17, 80333 Munich, Germany
[2] Aarhus University, Aabogade 34, 8200 Aarhus, Denmark
woelfel@cip.ifi.lmu.de
[3] Aarhus School of Architecture, Nørreport 20, 8000 Aarhus, Denmark
Timothy.Merritt@aarch.dk

Abstract. There are many examples of cards used to assist or provide structure to the design process, yet there has not been a thorough articulation of the strengths and weaknesses of the various examples. We review eighteen card-based design tools in order to understand how they might benefit designers. The card-based tools are explained in terms of five design dimensions including the intended purpose and scope of use, duration of use, methodology, customization, and formal/material qualities. Our analysis suggests three design patterns or archetypes for existing card-based design method tools and highlights unexplored areas in the design space. The paper concludes with recommendations for the future development of card-based methods for the field of interaction design.

Keywords: method cards, creativity cards, design methods, design tools.

1 Introduction

'Design methods are like toothbrushes.
Everyone uses them, but no one likes to use someone else's.' [13]

Physical cards have been popular design tools, perhaps because they are simple, tangible and easy to manipulate. Aside from the well-known Card Sorting method [22], cards have been used widely by designers to make the design process visible and less abstract [3, 10] and serve as communication tools between members of the design team and users [9]. There are many examples of unique method card systems, many have similar features and formal qualities, yet it is not easy to get an overview of the available card systems in order to decide which to use, and when. As suggested in the opening quote attributed to John Zimmerman, designers often develop their own methods or appropriate widely known methods to best suit their needs, yet there is a tendency to use methods that are familiar instead of venturing out. Even though these methods are shared with the research community, it is difficult and time consuming for designers to review all available tools to understand their strengths and weaknesses. Therefore, this paper aims at providing an overview of some well-known sets of method cards including features and limitations so that the interaction designer can

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 479–486, 2013.

quickly review and make informed choices when selecting card-based method tools or to serve as a source of inspiration as they develop and appropriate a suite of design methods of their own.

We begin by analyzing different card systems according to five design dimensions, which suggest possible archetypes or patterns for these tools. We then discuss limitations and future work on card-based design tools that could benefit interaction design researchers.

2 Analysis of Card-Based Design Tools

There are many examples of cards being used to assist or provide structure to the design. We selected and examined eighteen card systems through direct use and reviewed the accompanying literature to gain an understanding of the authors' intentions for use. As first steps toward describing the various card-based systems, we identified five design dimensions. While this is not an exhaustive survey, it begins to provide a sketch of the card-based method landscape and highlights key differences among the attributes of many card-based tools. This approach has been used to articulate the design space of various examples of interactive media [20, 26], and is used in the present paper to encourage the design community to engage and develop the field of card-based tools. When examined according to these dimensions, three broad groups of card systems emerge suggesting possible archetypes or design patterns [28]. A secondary contribution is that our review assists the review and selection of cards for design researchers and practitioners.

2.1 Design Dimensions

As stated previously, our aims in defining the five dimensions—and graduations within the dimensions—are in revealing key differences across the examples including: 1) Intended *Purpose & Scope*, 2) *Duration* of use and placement in design process, 3) *System* or Methodology of use, 4) *Customization*, and 5) *Formal Qualities*. These attributes describe claims from the literature of the authors, the formal characteristics, and the tools in use. While these may seem closely related, it is an initial step in developing a framework for discussing the design attributes in card-based tools. Graduations within these dimensions were chosen to differentiate the examples—future work is needed, however to validate and develop these further.

Intended Purpose and Scope. Based on research literature or from booklets and inserts included in the card packages, the respective authors have made claims as to where their tools fit within the design process (ideation, inspiration, engaging non-designers, etc.). In this category we can ask: where in the design process are the cards used and how should they be used? Do they have a specific purpose and do they focus on a particular context? We identified three graduations of intended purpose & scope, ranging from very general to context specific.

General/Repository card systems provide inspiration and challenge designers to take another point of view. An example here are the Oblique Cards [7, 8], which can be engaged with at any time in any context to increase lateral thinking and stimulate design problem solving in general. These types of cards aim for open-ended inspiration with little or no guidance on their use. These cards mainly function as repositories for design methods, capturing well-known methods from important literature [15] and offload the task of remembering the many design methods.

We also found various examples of cards, which focus on *participatory design*. They seek to develop sensitivity and empathy [24] for the context, and engage designers and users in the process. Some cards are designed for a better communication between users and designers, examples here are the Questionable Concept Cards [1], which encourage criticism and debate or the Inspiration Cards [2] that require collaborative work between designers and domain experts using the cards.

There are also *context specific/ agenda-driven* examples. This includes those cards focused on a particular context or design agenda as the Sound Design Deck [4], which facilitates sound design in games or the Design Play Cards [23], which focus on designing for sustainability.

Duration of Use/ When in Process. It is important to acknowledge the time investment that the various systems require – and to know when in the process they are used. This dimension includes key differences in the length of time ranging from one time use to sustained use of the system throughout the design process. Another aspect is the placement in the design process – whether the cards should be used in the very beginning, after initial field studies or prior to mockup sessions and prototyping. Four groups were identified, which range from anywhere/anytime to at a specific point in time.

The Oblique Cards are an example of cards that can be used *anywhere/any time* in the process. They can be useful in the very first phase of idea generation, but also when facing problems during the design, being stuck or looking for alternatives. Cards presenting a collection of methods as the IDEO Method Cards [16], are often positioned to be used *as needed*. As they provide a lot of different methods, some of those will fit in an early design stage, whereas others are for evaluation and testing. Other cards should be used at the *beginning of the process* as they provide input for further concept development; for example, PictureCARDs [12] are used after an initial field study and provide the basis for the card creation.

The last aspect of time is that cards are used at a *specific point*, for example in a workshop. The Sound Design Deck is used in this way, when applying the introduced methodology. But even though most of the work with the cards is done in a short session (~2h), one should still refer back to the cards later in the design process.

Methodology of Use. Some of the cards can be used very freely, whereas others provide a methodology how to use the cards. Some of the approaches are playful and game-like; some have rules or discreet steps that should be followed. This can be helpful to get started using the cards but might at the same time be restrictive. We identified three groups in this category: no methodology, suggestion for use and specific instructions.

Cards with *no system* are used ad-hoc with no suggested structured process provided by the authors. Cards of this type include IDEO [16], SUTD [14], and Oblique [7]. Most of the cards offer at least a basic *suggestion for use*. The DSKD Cards [17] come with a small brochure, which has some examples how the cards can be used. The authors of the PictureCARDs describe how they were using the cards, but there are no hard and fast, specific rules. The last category describes cards in which *specific instructions* are given. The authors of the Sound Design Cards introduced a specific method of how to use the cards, including a workspace with four regions in which cards can be moved, thus facilitating idea generation and keeping track of the design work at the same time. Inspiration Cards [2] also provide specific instructions, noting where the cards should be arranged on a poster to formulate a design idea.

Customization. Although we acknowledge that any technology tool will be adapted and appropriated into the user's life, in this dimension, we describe the degree to which the tool provides for customization as part of its use. The first group we identified in this category is *no customization*. When we examine the SUTD, Oblique, etc cards, they are intended to be static and unchanged. Cards offering *trivial customization*, do not allow the user to add or modify content, but only to structure or group the cards. This is the case with the IDEO iPhone app that in most respects replicates the paper cards [16] allowing the user to make groups and add cards to the groups. The Sound Design Deck provides for *optional customization*, whereas users can create their own cards and add them to a wiki. This is intended and welcome by the authors, as they aim to create a pattern language for sound design. The last group of cards *requires customization* in order to be utilized. Examples here are the Inspiration Cards, the Ideation Deck [6] or Questionable Concept Cards. The cards have to be created beforehand and are therefore applicable in the specific project, which helps the designer to get a better understanding of the project domain.

Formal Qualities. While the focus of this paper is on "cards", there are differences in the physical properties (2-sides, paper, size, shape), connections to virtual systems (stand alone or connected to objects in the room or in the virtual world), and appearances (images, diagrams, words, color schemes, etc.) Other formal qualities include issues such as the fact that some card systems have only one copy of a card vs. multiple copies, etc. We do not provide for all possible configurations, however, we provide graduations according to the use of media. The simplest type of cards have *only text or only images*, while most of the cards combine *text and image* or illustration, like the Inspiration Cards or the PLEX Cards [3]. The authors of PLEX Cards present their evaluation of the cards, and describe feedback regarding the images. This feedback highlights the importance of choosing suitable images for cards—they claim that the image should be abstract enough to allow an open interpretation, but at the same time detailed enough so that the user can relate to and interpret it. There are various card systems where the content is divided into different *categories*, as with the IDEO

Method cards or the SUTD Cards, which provides thematic structure in the cards and suggests how the cards relate to each other. Finally, there are some cards, which have a *virtual component*, as in the Sound Design Deck, which connects the physical cards to the online wiki providing additional information and example videos.

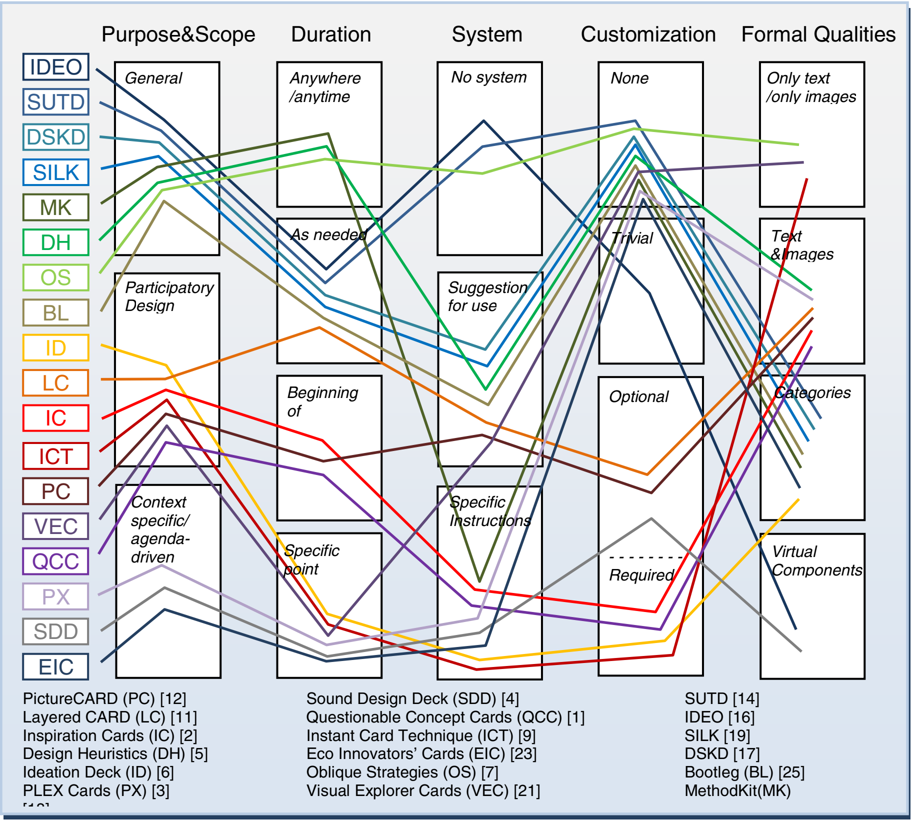


Fig. 1. Classification of method cards

2.2 Toward Patterns of Design

By plotting the card systems as shown in Figure 1, there are three archetypes or patterns of design [26,28] that emerge: *general purpose/repository cards, customizable cards,* and *context specific cards*. Further analysis and refinement of these groupings will be taken up in future work, however we can provide initial comments. Cards that are classified as "general purpose/repository" cards offer either a method repository or aim to stimulate inspiration and lateral thinking. They can be used during the whole process and only have general or no instruction for use. In most cases the deck is split up into different categories yet used without customization. "Customizable" cards are inherently customizable to some degree. Interestingly, we can see that most of these belong to the *participatory design* group and have specific instructions on how to use them to engage with end-users and non-designers. They are mainly used at a *specific point* in the design process and are composed of text and images. The third group includes “context specific” cards that are developed primarily to focus on a specific design agenda or context. They are primarily used at a specific point in time with specific instructions. In the next section we discuss how this exercise of analysis helps the design research community make sense of card-based tools.

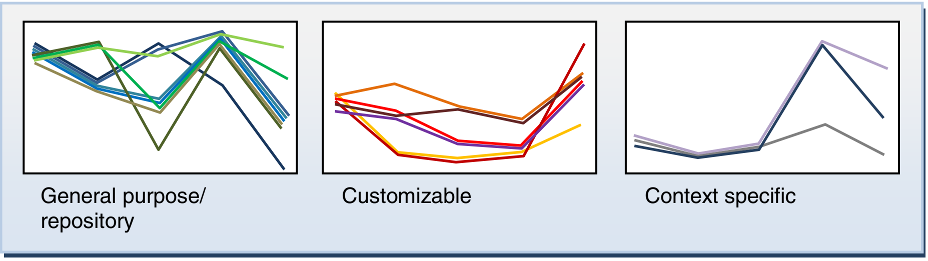


Fig. 2. Method card design patterns

3 Discussion

The classification exercise and resulting design patterns raise important issues for designers about when the cards can be used in the design process and what one should consider when choosing a card deck. We propose future work enabling customization and extending physical cards into digital media or augmented representations. To conclude, limitations of the present paper are provided.

3.1 Using and Choosing Design Method Cards

As noted earlier, designers can choose to use method cards in their design process either by using existing systems or by creating a new set of cards—essentially, should the designer use an existing toothbrush [13] or create a new one? This decision depends on several factors including time, knowledge of team members, needs of the design team, expected results and many other possible factors. Through the process of reviewing and analyzing the card systems and accompanying literature it would seem beneficial for designers to become familiar with at least one example of each of the three card archetypes we identified before choosing or creating a card-based tool. After reviewing these, the designer should evaluate the time and resource investment that can be made and to stimulate ideas about possible formats for a card system if they determine that they will build a tool to be used on an ongoing basis. If a new card system is created, perhaps the most critical attribute is the degree of customization that is expected—whether the cards are living and evolving, or serve more as a lasting resource of commonly needed methods. In the next section we engage with the issue of customization of card systems and the use of technology as ways to extend traditional card-based tools.

3.2 Customization and Use of New Technology

It was surprising that only few examples were found that offer customization directly, or in some form of digital component to augment the tool. There is much interest in the development of robust technologies including QR codes, augmented reality, e-ink, etc. that bridge the gap between the digital and physical world with the aim of engaging people into complicated work processes [27] or to capture lasting impressions of work. Showing similarities between individual cards within different card systems could go much further to reduce the time spent in the design process searching for appropriate methods. Various digital technologies could facilitate this by showing connections virtually or presenting additional information augmented onto the cards on the table or wall. Designers could add their own notes to cards and share those with their team members, which was a future development suggested also in [2], yet robust examples were not found in any of the card system except in [4], however, it is not clear that the digital components are used for anything more than a repository. This does however go beyond the digital representation approach offered in the iPhone App for the IDEO Method Cards. There is no connection to the physical world except that all of the digital cards are copies of the physical cards—there is no customization other than adding cards to groups. In a sense, the digital app loses functionality—the designer can not look at more than one card at a time, can not write on the card itself, and there is no new information despite the content being more than ten years old. In our future work we intend to explore meaningful ways of connecting physical card-based tools with interactive digital elements.

3.3 Limitations

While the present paper begins to sketch out the design space for design method cards, there are limitations that we would like to acknowledge. Our list of method cards does not represent all method card systems, yet eighteen examples is adequate to begin to understand the design space. We also did not provide in-depth reviews of each of the card systems, excluded the number of cards in each deck, etc. These limitations do not diminish the contribution here, but rather signal important future work.

References

1. Vines, J., Blythe, M., Lindsay, S., Dunphy, P., Monk, A., Olivier, P.: Questionable concepts: critique as resource for designing with eighty somethings. In: Proc. CHI 2012, pp. 1169–1178. ACM, NY (2012)
2. Halskov, K., Dalsgaard, P.: Inspiration card workshops. In: Proc. DIS 2006, pp. 2–11. ACM, NY (2006)
3. Lucero, A., Arrasvuori, J.: PLEX cards: a source of inspiration when designing for playfulness. In: Proc. Fun and Games 2010, pp. 28–37. ACM, NY (2010)
4. Alves, V., Roque, L.: A deck for sound design in games: enhancements based on a design exercise. In: Proc. ACE 2011. ACM, NY (2011)

5. Daly, S.R., et al.: Assessing Design Heuristics for Idea Generation in an Introductory Engineering Course. International Journal of Engineering Education 28(2), 463 (2012)
6. Golembewski, M., Selby, M.: Ideation decks: a card-based design ideation tool. In: Proc. DIS 2010, pp. 89–92. ACM, NY (2010)
7. Eno, B., Schmidt, P.: Oblique strategies. Opal, London (1978)
8. Burns, A.: Oblique Strategies for Ambient Journalism. M/C Journal (May 2010), http://www.journal.media-culture.org.au/index.php/mcjournal/article/view/230
9. Beck, E., Obrist, M., Bernhaupt, R., Tscheligi, M.: Instant card technique: how and why to apply in user-centered design. In: Proc. PDC 2008, pp. 162–165. ACM, NY (2008)
10. Lafrenière, D., Dayton, T., Muller, M.: Variations of a theme: card-based techniques for participatory analysis and design. In: Proc. CHI 1999, pp. 151–152. ACM, NY (1999)
11. Muller, M.J.: Layered participatory analysis: new developments in the CARD technique. In: Proc. CHI 2001, pp. 90–97. ACM, NY (2001)
12. Tschudy, M., Dykstra-Erickson, E., Holloway, M.: PictureCARD: A Storytelling Tool for Task Analysis. In: Proc. PDC 1996, Cambridge, MA, USA, November 13-15 (1996)
13. Harrison, S., Tatar, D.: On Methods. Interactions 18(2), 10–11 (2011)
14. Singapore University of Technology Design: SUTD Method Cards (2012), http://design.sutd.edu.sg/wp-content/uploads/files/extraResources/MethodCards.pdf
15. Rogers, Y., Sharp, H., Preece, J.: Interaction Design: Beyond Human - Computer Interaction, 3rd edn. Wiley (June 2011)
16. IDEO: IDEO Method Cards: 51 Ways to Inspire Design. IDEO, Palo Alto (2003)
17. Kolding School of Design: DSKD Method Cards (2011), http://www.designskolenkolding.dk/index.php?id=2241
18. Möller, O.: MethodKit. Idea Society, Stockholm, Sweden (2013)
19. SILK (Social Innovation Lab Kent): Method Deck. Kent, UK (2007)
20. Cheok, A.D., Kok, R.T., Tan, C., Fernando, O.N.N., Merritt, T., Sen, J.Y.P.: Empathetic living media. In: Proc. DIS 2008, pp. 465–473. ACM, NY (2008)
21. Center for Creative Leadership: Visual Explorer Post Card Facilitator's Set. Greensboro, NC (2013), http://solutions.ccl.org/Visual_Explorer_Post_Card_Facilitator's_Set
22. Hudson, W.: Playing your cards right: getting the most from card sorting for navigation design. Interactions 12(5), 56–58 (2005)
23. Eco Innovators: Design Play Cards: Designing for Sustainability. Melbourne (2013)
24. Bowen, G.: Grounded theory and sensitizing concepts. Int. J. Qual. Meth. 5(3) (2006)
25. Stanford University Institute of Design: Bootcamp Bootleg (2010), http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf
26. Pousman, Z., Stasko, J.: A taxonomy of ambient information systems: four patterns of design. In: Proc. AVI 2006, pp. 67–74. ACM, New York (2006)
27. Zhao, S., Nakamura, K., Ishii, K., Igarashi, T.: Magic cards: a paper tag interface for implicit robot control. In: Proc. CHI 2009, pp. 173–182. ACM, NY (2009)
28. Alexander, C.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, UK (1977)

The Design and Usability Testing of DACADE – A Tool Supporting Systematic Data Collection and Analysis for Design Students

Madihah Sheikh Abdul Aziz, Gitte Lindgaard, and T.W. Allan Whitfield

Faculty of Design, Swinburne University of Technology, PA Building,
144 High Street Prahran, Victoria, 3181, Australia
{msheikhabdulaziz,glindgaard,awhitfield}@swin.edu.au

Abstract. Norman claims that designers are bereft of statistical knowledge to perform effectively [10], stating that designers must understand technology, business and psychology to support design decisions. For designers to acquire the necessary statistical skills, design curricula should incorporate statistical courses teaching systematic data collection and data analysis. This paper presents the development and formative usability tests of the prototypes of a software tool called DACADE intended to support design students collecting and analyzing data systematically early in the design phase. It uses a 2D map and a Napping® technique to support effective and efficient communication between designers and target audiences in the design decision process by providing visual data and descriptive statistics without needing statistical knowledge.

Keywords: Software Engineering (Usability Testing), Human Factors in Software Design (User Interfaces), user-centered design, human-centered design.

1 Introduction

With the spread of the Internet, large-scale data collection, analysis, and product/ service testing are feasible and cost-effective. Specialist companies provide statistical packages, such as SPSS to analyse test results as well as offering data entry services and even data analysis. This complements the trend of data mining in both the retail and the social media advertising domains. Inevitably, however, the analyses tend to be biased towards quantitative rather than qualitative outcomes; statistical procedures employed vary from simple Cluster Analysis to Multidimensional Scaling and Multiple Regression, even Structural Equation Modelling. Design company clients will increasingly receive market research reports incorporating such analyses.

In our study, we first interviewed 20 design students and 10 design lecturers at one University, followed by a survey of 51 universities with design courses in Australia, North America, Europe, United Kingdom, and Asia. Results revealed that none of the curricula included statistics, thus supporting Norman's claim that design students do not understand statistical concepts. It is therefore legitimate to question future designers' ability to understand statistical information as well as their ability to engage

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 487–494, 2013.

actively and meaningfully in this emerging world of statistical sophistication. Norman contends that they cannot [10]. The **Da**ta **C**ollection and **A**nalysis for **De**signers (DACADE) software tool described here aims to help future designers collect and analyze product-related data from prospective consumers.

1.1 Existing Visual Statistical Tools

Research shows that designers prefer visual information [6]. Following Kälviäinen et al., [8], this study thus focuses on tools allowing visual data collection and analysis. Several visual research tools that automate data collection and analysis are available in the literature, including Computer Aided Kansei Engineering (CAKE) with XML [2], the Web-based 2 Dimensional (2D) analytical tool [9] allowing respondents to position images of products on a 2D map using Semantic Differentials (SD), 3), and others. Tools based on SDs scales require designers to pre-assign sets of bipolar adjectives to product images (e.g. Bad – Good, Ugly – Pretty). Kälviäinen et al., [8] proposed that product images with pre-assigned sets of bipolar adjectives could constrain the respondents' possible suggestions and that respondents should be free to evaluate products as they see fit [8]. Some of the available tools assume that designers know statistics: none of these have been tailor-made for future designers.

1.2 Napping® and 2D Map

DACADE includes two techniques for collecting and analyzing opinion data. One allows evaluators to position product images on a blank screen in any way they like. Products positioned close to others are perceived to be similar; those positioned far apart are seen to be different. Once all images have been placed, evaluators enter adjectives to represent the products/groupings meaningfully. Based on these evaluator-selected terms, designers can then select suitable bipolar adjective pairs for further study. This technique is called projective mapping or Napping®. It is used widely in sensory analyses in the food industry [11]. Based on the sets of original adjective-pairs, other respondents' product perceptions can be tested in later design stages.

Involving a 2D map that enables the direct reading of consumers' perception, the second technique relies on designer-selected adjective-sets from which the designer generates a perceptual map of consumer perception. Perceptual maps are popular in marketing for studying the perception of products ranging from consumer products (e.g. toothpaste, cars) to activities (vacation spots, movies) [12]. DACADE uses perceptual maps because they are easy to interpret, and they provide visual output as designers prefer. Marketing researchers and practitioners tend to rely on statistical software packages such as SPSS to run highly advanced statistical analyses that may be too complex for statistically naïve design students. Instead, DACADE uses nonparametric Guttman-Lingoes Series MDS [4, 5] that rely on visual and spatial information for generating perceptual maps.

2 Initial Design and First Formative DACADE Usability Test

DACADE was designed using User-Centred Design (UCD) [13] and the ISO-9241/11 [7] definition of usability, i.e. effectiveness, efficiency, and user satisfaction. Five graduate design students evaluated the initial evaluation low-fidelity paper prototype in individual sessions, audio recorded with permission. A separate prototype was produced for each of four tasks, namely (1) designing a new study, (2) editing an existing study, (3) collecting data (i.e. running the study) using both abovementioned techniques, and (4) conducting simple descriptive statistical analyses. The analysis component calculates the median, mode, and mean of the research data as well as providing a visual output of the distribution of objects tested. Written task instructions and relevant task scenarios were given; answers were written on a blank sheet of paper. On each screen, participants placed a pen on the object they wished to select. Task requirements were generated separately for each task and given to participants one at a time. All tasks were completed in the same order (1-4). The stimuli comprised some 10 pictures of cars borrowed from Effendi [3]. Upon completing a task, participants filled in the System Usability Scale (SUS) [1] to assess user satisfaction. Notes were taken throughout the sessions; audio records were transcribed verbatim.

Some 31 comments were made of which 15 concerned usability of the remaining 16 issues; six clearly indicated that the students had not grasped the underlying statistical concepts concerning the central tendencies of a distribution. For example, one participant said: "*I think I like them, but I think it assumes that I already know what does this means or even what frequency or mean is, I want to do that, but maybe I don't even know what that is. It seems like a basic thing someone should know in statistics*". This feedback such as this revealed the need for a tutorial component for designers to understand simple quantitative data analysis. Additional comments concerned slight confusions. For example, the 'Edit' button was changed to 'Add Image'. One participant did not recognize check boxes drawn on one screen, possibly because of the rough drawing. Six suggestions were made, for example: "*I don't see why I have to press Cool and click Insert. Wasting my time. Should be done automatically, just click Cool*". All of these were incorporated into the second prototype.

3 Design and Test of the DACADE Tutorial

The tutorial designed next, covers the four DACADE components as well as a basic explanation and illustration of descriptive statistics. The tutorial was first tested with four new participants who did not complete the tasks in the formal tool. In order to illustrate the calculation of the mean, mode, and median, a picture of bananas varying in size was produced as shown in Fig. 1. Surprisingly, the students perceived this as having sexual connotations, which was both inappropriate and embarrassing for the researcher. The image was therefore changed to a set of books as shown in Fig. 2. This appeared to convey the intended meaning. Several exercises were included in the tutorial. One required participants to calculate the mean, mode, and median of a set of numbers differing from those in the illustration, as shown in Fig. 3. Unfortunately, the median had inadvertently been placed in the center of the distribution in Fig. 2. Rather than calculating the median of the unsorted data set as required, one participant simply counted the number of books from each end of the distribution, selecting the central number ("13") without considering the values of the numbers in the data set. Evidently, the concept was still not conveyed clearly enough. The image was therefore changed again to avoid the median being in the central position.

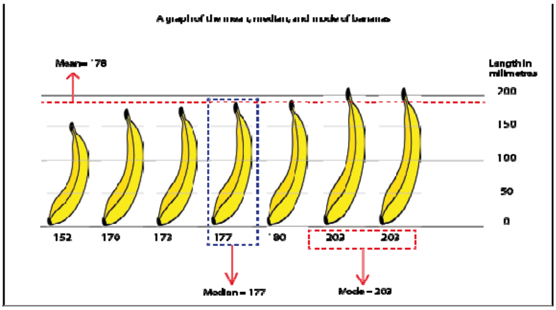


Fig. 1. The original illustration of the mean, median and mode

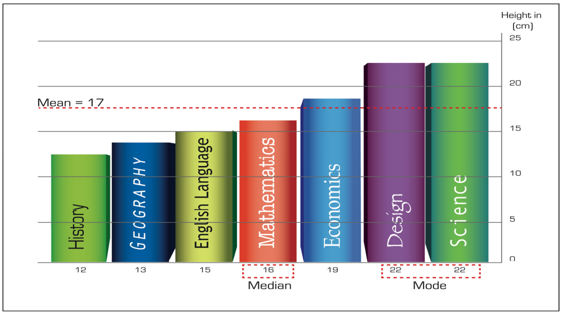


Fig. 2. The replacement illustration of the mean, median and mode

Find the mean, median and mode, for the following list of values:
13, 18, 13, 14, 13, 16, 14, 21, 13

Fig. 3. The data set shown in the tutorial exercise

4 Second Formative Usability (and comprehension) Test

As there were no additional usability issues or apparent comprehensibility issues, the next test, involving the revised tutorial and the revised DACADE tool, included a new sample of six participants together with the original instructions and tasks. Participants worked through the tutorial and exercises first. They were then given the cover story and the task instructions and asked to work with the tool. The results showed a distinct improvement to DACADE: only five (9.43%) of the 53 issues were usability-related. However, some 19 (35.85%) issues concerned a lack of comprehension. Another 17 (32.08%) were suggestions, eight (15.09%) were participants' uncategorized comments, and four (7.55%) were the researcher's observations.

Among the comprehension issues, one showed that participants had difficulties understanding the purpose and meaning of the 2D map, shown in Fig. 4. In both the tutorial and the tool, references were made to a fictitious study of water bottles. In one tutorial exercise, participants were asked to identify the most interesting and modern water bottle, the most traditional and boring bottle, and so forth. We explained that, for both sets of adjectives (boring-interesting and traditional-modern), 'more is better', with the most negative term placed on the bottom and to the left, with the most positive term placed at the top and to the right.

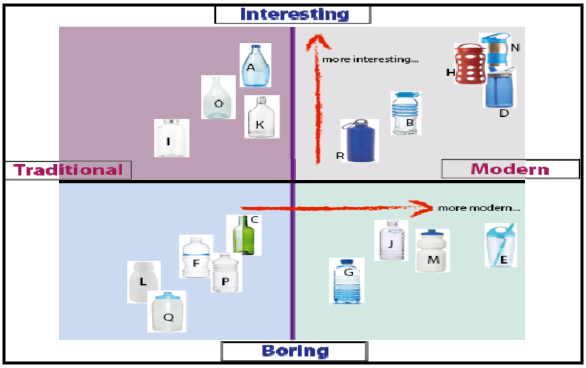


Fig. 4. A perceptual map showing the placement of eighteen water bottles

We also explained the meaning of clusters (mode), and so on. Still, when asked to interpret the contents of the map in Fig. 4, said to represent the placement of bottles by another participant, one participant, noting certain similarities between bottles, said: "*Its like more of a sample, I would say they should be in the same place, because they are all the same. Similar bottles should be in the same place, like now they are kind of everywhere. They are all the same*". Similarly, when asked to identify the best bottle, one participant said: "*The answer is R, because it is made of metal, so you can bring it on to outdoors or, go [*on an*] expedition, that's my judgment*". Another participant said: "*The best bottle, it depends on environmentally, or the needs, like a flask or something, for me I'll just go for B*". Another participant, who was unsure of the position on a 2D map, asked, "*Does it matter if it is located further up on the quadrant*?" This participant further commented that the perceptual map, "*…itself looks very statistical and technical; you need to make it more fun and interesting. I want to be able to enjoy the graphics. It really depends on how you design this later*" At this stage, we are at a loss as to how to convey the meaning of visually presented data more clearly, but this problem must be solved before redesigning the prototype.

Other misunderstandings were due the prototype being presented on paper. For example, as shown in Figure 5, two participants did not recognize that the meaning of a greyed-out button was the same as on a computer.

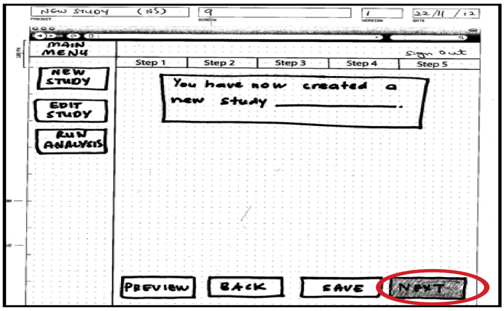


Fig. 5. A screen with a greyed-out button

Participants were also confused about adding a new map on a paper screen. That function enables designers to add one or more 2D maps to be tested in their study. However, instead of adding a new map, they just changed the adjectives on the existing map. This will hopefully be overcome once we use high-fidelity prototypes.

One participant wanted a definition of the term 'adjectives', "*The word adjective might confuse people. I think 90% of people don't read instruction manuals; there should be some rough instructions. You can add a link or a question mark symbol so that people can click on them to get some hints. Or roll over to get some tips"*. Tooltips will be provided in the implemented version of DACADE. Another suggestion was to present the tool online, to enable the designers to involve more people in their studies. While large samples make sense, one of the aims of DACADE is to educate design students and help them to interact and communicate with prospective consumers face-to-face. In future, DACADE might be extended once designers begin to use it, but for the moment, this suggestion will not be considered. Failing to understand the concepts of groups of participants and total number of participants, some participants suggested that DACADE should allow an unlimited number of participants, which it does, but also that the number of participants in a group should not be constrained. That is, they wanted the concept of participant-groups removed. This is infeasible because the tool is specifically designed to guide them in conducting systematic research. This includes the ability to compare between-group responses, e.g. seniors versus teenagers, or allowing the same number of males and females.

Participants made several good suggestions, for example, adding clear instruction to consumers to position images on a 2D map or a blank screen because consumers will not be reading the tutorial. For the task of creating a new study, it is clear that additional instructions are needed. However, in the light of the comment that *"...90% of people don't read instructions*", it is difficult to balance the amount of instructional text and diagrams. One possible solution is to animate the tutorial, thereby changing it to a 'show and tell' show, with voice-overs and animated diagrams only. This may be done later. Note, however, that if students are unfamiliar with terms such as 'adjectives', even a voice-over will not necessarily help them. Another suggestion was to enable simultaneous display of all images to be placed on the map instead of presenting them one by one. This will be done to give consumers a general idea of the entire range of product images to be positioned from the start of the task. In previous research [2,9], images were presented one at a time or in pairs.

One important observation was that participants looked bored while working through the tutorial. They complained that it was too long and had too much text, again suggesting that animation may be best for creating a playful, yet useful and comprehensive tutorial while retaining its brevity. Finally, it was clear that some participants failed to read the tasks thoroughly; they needed several reminders to refer to the task instructions on how to proceed.

5 Second Formative Usability (and comprehension) Test

Usability goals were set for both tests in terms of the number of questions, hints, and errors allowed during the usability tests. Table 1 shows the distributions of these for each of the four tasks and the results for both usability tests of the DACADE tool.

Table 1. The usability goals set for both tests according to tasks

Tasks	Test1				Test2			
	NS	ES	CD	RA	NS	ES	CD	RA
Questions	3	2	1	3	2	0	0	2
Hints	2	1	1	1	1	1	0	1
Errors	2	2	1	2	0	0	0	0
Results	P	P	P	F	P	P	P	P

**Note: NS=New Study, ES=Edit Study, CD=Collecting Data, RA=Running analyses, P=Pass, F=Fail*

The between-task variation in these numbers in Test 1 was based on our best guess of the difficulty associated with each task. Clearly, the task of 'Running one or more analyses' or RA was most difficult, as it was the only task that failed in Test 1. In Test 2, the goals were based on the outcome of Test 1. Note that the goals were more stringent in Test 2 than in Test 1 because we assumed that the tutorial added in Test 2 would enable participants to understand the descriptive statistics and interpret the perceptual maps. The fact that all four tasks passed in Test 2 could lead one to misinterpret the tool's apparent effectiveness. In cases in which user comprehension may be as important as, if not more important than, usability, comprehension should be added in the task protocol as a separate variable in the performance assessment.

6 Conclusion and Next Steps

This paper described the design of a visual statistical technique to help future designers understand consumers' product perceptions. We showed that it was relatively easy to identify and eliminate usability problems, but that finding effective ways to convey even simple statistical knowledge is fraught with unforeseen difficulties.

The issues from the second usability test are now being addressed, but rather than producing a slightly another prototype, DACADE will be implemented, and professional designers will be consulted on its look and feel. It is anticipated that the results of a user acceptance study will be available by the time INTERACT'2013 takes place.

Acknowledgements. This PhD research is funded by the Swinburne University of Technology Australia and International Islamic University Malaysia.

References

1. Brooke, J.: SUS-A quick and dirty usability scale. ©Digital Equipment Corporation (1986)
2. Chuang, Y., Chen, L.-L.: Computer Aided Kansei Engineering with XML Technology. In: 6th Asian Design International Conference, Japan Tsukuba, Japan Tsukuba, pp. 1–10 (2003)
3. Effendi, R.A.A.R.A.: The Product Effect: Do Designed Products Convey Their Characteristics To Their Owners? Degree of Doctor of Philosophy, Faculty of Design, Swinburne University of Technology, Melbourne (2011)
4. Guttman, L.: The basis for Scalogram Analysis. In: Stouffer, S.S. (ed.) Measurement and Prediction, Princeton University Press, Princeton (1950)
5. Guttman, L.: A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika 33, 469–507 (1968)
6. Md Hashim, A., Ahmad, R.A.A.B.R., Whitfield, T., Jackson, S.: The Multi Dimensional Scaling: An Interactive Method for Establishing Perceptions of the Appearance of Product, pp. 1–18 (2007)
7. ISO 9241-11.: Ergonomic Requirements for Office Work with Visual Display Terminals (VDT) s-Part II Guidance on Usability, ISO/IEC 9241-11 (E) (1998)
8. Kalviainen, M., Miller, H.: Visual Research: Means of producing shared meanings, Joining Forces, University of Art and Design Helsinki (2005)
9. Lin, J.S., Huang, S.Y.: Analyzing target user group's preferences and product form design specification through web-based 2-dimensional design decision tool. International Journal of Business Research and Management (IJBRM) 1, 14–32 (2010)
10. Norman, D.: Why Design Education Must Change (2010), `http://www.core77.com/blog/columns/why_design_education_must_change_17993.asp`
11. Pagès, J.: Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the Loire Valley. Food Quality and Preference 16, 642–649 (2005)
12. Rice, M.: An Introduction to Brand/Perceptual Mapping, Tracking Consumer Perception, pp. 1–4 (2001)
13. Rogers, Y., Sharp, H., Preece, J.: Interaction Design beyond human-computer interaction, 3rd edn. John Wiley & Sons Ltd. (2011)

The Effect of Physicality on Low Fidelity Interactive Prototyping for Design Practice

Joanna Hare, Steve Gill, Gareth Loudon, and Alan Lewis

Cardiff Metropolitan University, Cardiff, UK
jhare@pdronline.co.uk,
{sjgill,gloudon}@cardiffmet.ac.uk,
profalan.lewis@wales.ac.uk

Abstract. In this paper we propose the concept of 'active' and 'passive' physicality as mental models to help in understanding the role of low fidelity prototypes in the design process for computer embedded products. We define ‘active physicality’ as how the prototype and its software react to users and ‘passive physicality’ as how the prototype looks and feels offline. User trials of four different types of ‘low fidelity’ prototypes were undertaken using an existing product as the datum. Each prototype was analysed in terms of active and passive physicality and user responses were collated and compared qualitatively and quantitatively. The results suggest that prototypes that balance both active and passive physicality produce data closer to the final device than those that are strong in one at the expense of the other.

Keywords: Physicality, interactive prototypes, computer embedded products, design, product design, iterative product development, information appliances.

1 Introduction

This paper builds on previous research on physicality and low fidelity interactive prototypes. Virzi et al. [1] found that there was little difference in usability data for high and low fidelity models of standard two dimensional graphical interfaces and an interactive voice response system. Yet a number of researchers [2] [3] felt that the concept of low verses high fidelity is not quite enough to convey the whole manner of situations that prototypes are constructed for. McCurdy et al. [3] argued for a mixed approach that allowed various aspects of a prototype to be built at different fidelity levels according to the design component being prototyped. They go on to suggest that there are five ‘dimensions’ or fidelity aspects that can be defined as somewhere between high and low within the same prototype, namely, aesthetics, depth of functionality, breadth of functionality, richness of data and richness of interactivity. So far this concept of mixed fidelity has been trialled with software but not physical prototypes. Despite several authors conducting studies on prototypes of computer embedded devices the physical properties of both the model and interaction have been largely ignored.

In 2008 we demonstrated that in order to trial an interactive device with users an interactive prototype must be constructed [4]. The same study went on to lower both

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 495–510, 2013.

the level of physical fidelity of the model and the visual fidelity of interface until usability data started to significantly differ from the results of the final device. It was proposed that subtle differences in physicality, in this case removing the tactile feedback of buttons, affected the results suggesting that considerations of physicality are more important than the level of fidelity. This poses the question of how we 'consider' physicality.

However, a study published in 2009, demonstrated that some effects of physicality on user trials were only apparent through in-depth analysis because the effects were often subtle and the picture sometimes confusing [5]. This study seeks to clarify the position physicality occupies in user interactions.

The 2009 study sought to uncover the resulting differences in physicality based on low, medium and high(er) fidelity prototypes. In this study physicality was considered to fall under two areas: the physicality of the device (e.g. form, finish, weight) and the physicality of the interaction (the feel of the buttons and wheel in this case). But this method only allows the prototypes to be described and not directly compared which is essential when using physicality to determine the differences between the prototypes on trial. The physicality of the device and interaction was an appropriate way to describe the prototypes and, with subsequent analysis, this has been adapted to form the concept of 'passive' and 'active' physicality where:

Passive Physicality is how the prototype looks and feels when turned off, for example the weight, finish and button locations.

Active Physicality is how the prototype reacts to the users, typically the reaction of the interface (software), the feel of the buttons when operated (or sliders, dials, screen etc.)

To explain these terms a useful starting point is that of Dix et al. [6] who regard the physical device removed from its context and 'separated' from its digital operation in order to consider the mapping of the device 'unplugged'. This is the basis of 'passive' physicality; the judgments that can be made about the device without switching it on. Do you grasp a cup by its handle or by the body? Decisions are made about the comfort of the cup's handle by its appearance and the perceived weight of the contents of the cup [7]. Passive physicality also has its roots in Gibson's description of affordances [8] which suggest ways of interaction. Affordances are not simply a property of the object; they are the way a specific user relates to that object. When Norman [9] applied Gibson's idea to design; he divided the idea of affordances into those of real and perceived affordances. Whilst real affordances tell the user what they could actually do with the device, meaningful or not, perceived affordances tell the user 'what actions can be performed on an object and, to some extent, how to do them'. Yet passive physicality is more than affordances, it includes the physical properties of the device, its weight, finish and locations of the interactions.

Active physicality is concerned with the interactive portion of the device; what happens when the device is being used. It is still the physical that is of concern but in relation to the device's purpose and ease of use; how buttons operate the interface and how those buttons (or any interactions) feel when operated.

The exact drivers behind active and passive physicality might differ depending on the product being prototyped but the essence of active and passive physicality will remain.

This study proposes that a prototype can be considered by its level of active and passive physicality. For example, a prototype that is driven by the technology of the experience rather than the proposed size of the design would have a high level of active physicality but low passive physicality.

By attempting to understand physicality and using this to drive the physicality of low fidelity prototypes we aim to draw out just how physicality can be used by the designer to create efficient low fidelity prototypes. The efficiency of a prototype is of great importance; an efficient prototype can supply reliable data for a fraction of the cost of a high fidelity prototype enabling an iterative process. The early stages of the typical user-centred design process are highly iterative in order to react to and inform the developing project. User trials are a key tool to gathering data needed to inform the project, techniques include rapid ethnography [10], usability evaluation [11] and task centered walkthroughs all of which can be supported by interactive prototypes, and these prototypes need to be fast, low-cost and stage appropriate. This paper presents an early stage study on four low fidelity prototypes of the same device.

2 Methodology

2.1 The Prototypes

An existing product was chosen to provide a datum against which the retrospectively developed prototypes could be measured. The choice to retro-prototype an existing device as a method was taken after considerable thought. The alternative would have been the development of a new device. Both methods have been used in prototype evaluation studies [4] [12]. Retro-prototyping was chosen because it has the benefit of access to a real, mass produced product, identified by the manufacturer as a worthwhile idea and having successfully undergone a product development process. The finished device can be used to compare the results from the user study in a manner that is all but impossible to recreate in a research study.

Fig. 1. The iRiver SPINN

The product chosen was the iRiver Spinn (Figure 1), a personal music player. The main features and interactions of the iRiver Spinn are shown in Figure 2.

Four low fidelity prototypes were constructed using techniques currently in use in industry. Each prototype was planned giving due consideration to active and passive physicality levels, with the intention of placing one in each of the quadrants shown in the graph in Figure 3.

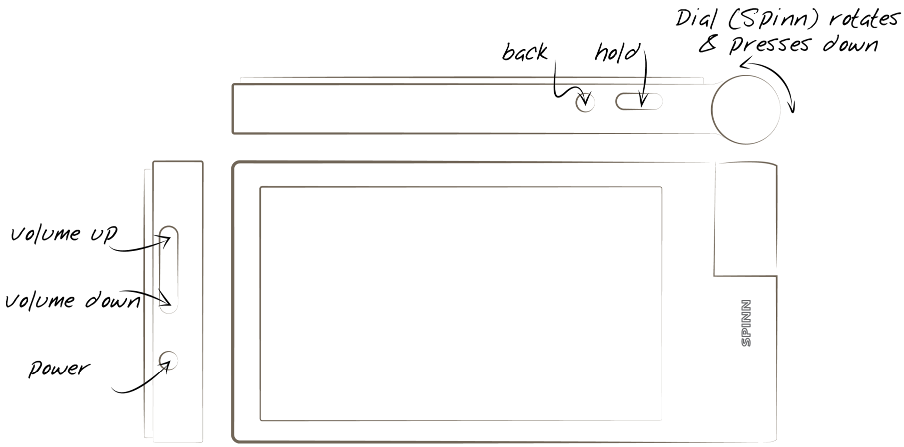


Fig. 2. The interactions of the iRiver Spinn

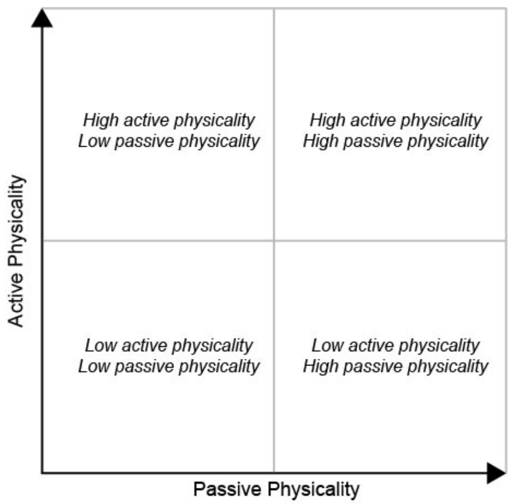


Fig. 3. Areas of physicality

The time taken to make each type of prototype is a critical issue. ‘Time is money’ and so we timed the building process and applied an hourly rate of £40 in order to cost each prototype. These are shown in Figure 4.

Fig. 4. The time taken to construct the prototypes

A single interface core was coded in Adobe Flash for all prototypes and adapted to the needs of each. Preparatory work ensured that this interface would be suitable for all prototypes and that the adaptation of the interface was possible for all. As is typical at this stage of the design process, only a limited selection of features were included in the software [11]. A single Computer Aided Design (CAD) model was created.

Prototype 1 (Figure 5; named ‘blue foam’) was constructed from model making foam board. Interaction was based on the Wizard of Oz technique [13], the Flash interface was operated remotely by the facilitator and viewed on the Tablet, the participant was asked to follow the ‘think out loud’ protocol [14], the facilitator could react to what the participant was saying and interacting with on the foam prototype.

The physical model for **Prototype 2** (Figure 6; named ‘IE4’) was constructed using rapid prototyping techniques (FDM). The CAD model was adjusted slightly to house the buttons and the dial which were integrated to make the prototype interactive. An IE4 [15][1] was used to connect the buttons to a laptop. The Flash interface, shown on a tablet, ‘listens’ for key presses from the IE4 and triggers changes in the interface when the participant interacts with the prototype.

[1] The IE4 is a wireless device which converts buttons presses into keyboard presses.

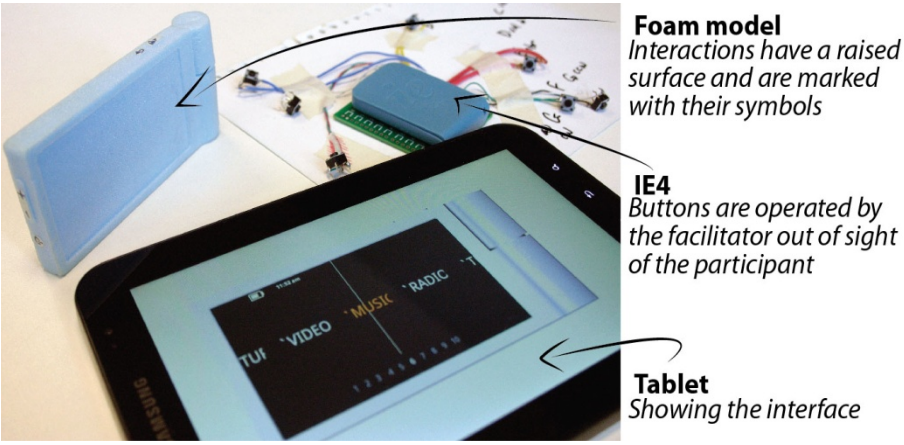


Fig. 5. Prototype 1: Foam prototype

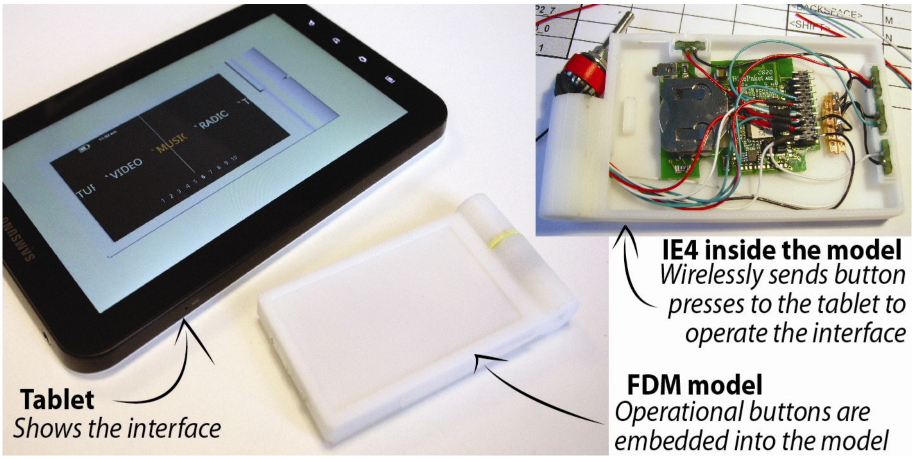


Fig. 6. Prototype 2: IE4

The physical model for **Prototype 3** (Figure 7; named 'appearance model') was intended to reflect the final device as accurately as possible. The form was rapid prototyped (using FDM) then finished to facsimile level. The Flash interface was operated by the participant on a touch screen tablet.

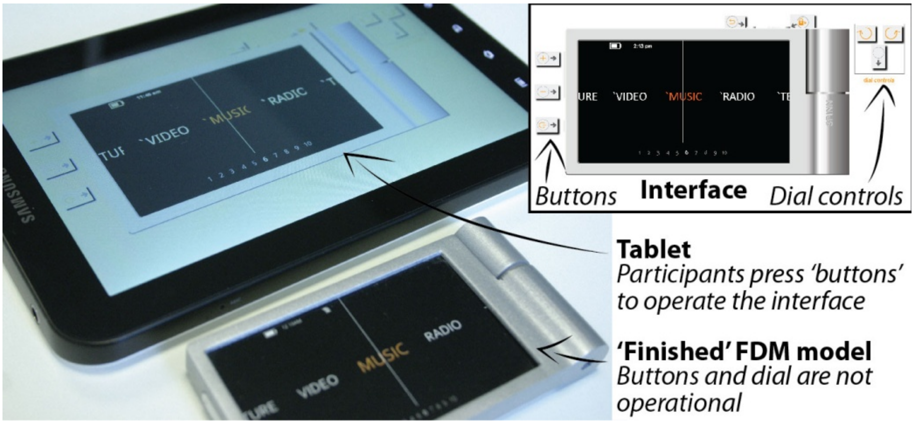


Fig. 7. Prototype 3: Appearance prototype

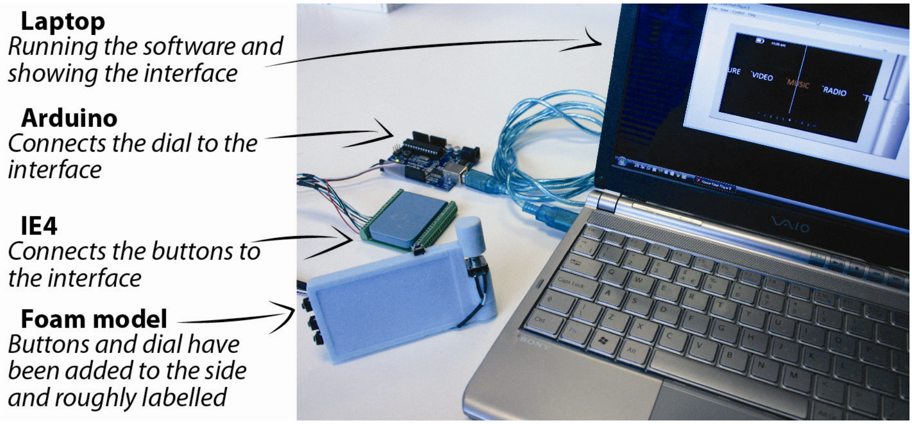


Fig. 8. Prototype 4: Arduino prototype

A rough foam model was constructed for **Prototype 4** (Figure 8; named 'Arduino') to accommodate the off-the-shelf buttons and dial. The dial was connected to an Arduino [16] which received the analogue signals and outputted them to the computer running the Flash interface. The buttons were connected to an IE4. Due to the extra code required for the Arduino, the interface was shown on a laptop rather than the touch screen tablet.

2.2 Assessing Physicality

Each of the prototypes was analyzed in terms of active and passive physicality. The main factors in the design that would determine the passive physicality levels of the prototype were determined to be: scale, form, finish and button location. For active physicality the main issues were: Spinn physical feedback, Spinn digital feedback, button physical feedback and button digital feedback. Initially a 'scoring' system was trialed but this was discarded, for when we call a prototype 'low' fidelity we do not assign that 'lowness' a value, as designers we intrinsically know when a prototype is low fidelity. It is only when conducting studies such as this that a prototype is considered lower or higher than another. Figure 9 shows the considerations for assessing each prototype.

Prototype	Passive physicality	Active physicality
Blue Foam	Low This prototype looks approximate and feels light, buttons are obviously cardboard and not working.	Low Buttons are obviously intangible and the participant is speaking through their expected interactions which are being interpreted by the facilitator who is operating the Flash based interface.
IE 4	Mid This prototype looks reasonable with no distracting wires. The prototype can be held comfortably yet it is very obviously an early stage prototype.	Mid Interactions mimic the design intent satisfactorily directly operating the interface which is a reasonable approximation of the design intent.
Appearance model	High The prototype looks and feels very similar to the final product.	Low The interactions are not obvious as the participant does not use the tangible prototype to operate the interface; instead the interface is operated on a touch screen breaking the link between the tangible product and its interface.
Arduino	Low The prototype has tacked on switches and wires are distractingly apparent in both the aesthetics and tangibility of this prototype.	High The prototype accurately mimics the way the final device feels when it is operated, both in the way the buttons work and the functionality of the interface.

Fig. 9. Assessing the levels of active and passive physicality of the prototypes

The Appearance and Arduino prototypes are high in one area of physicality at the expense of the other, whilst the Foam and IE4 prototypes 'balance' both active and passive physicality, as shown in Figure 10.

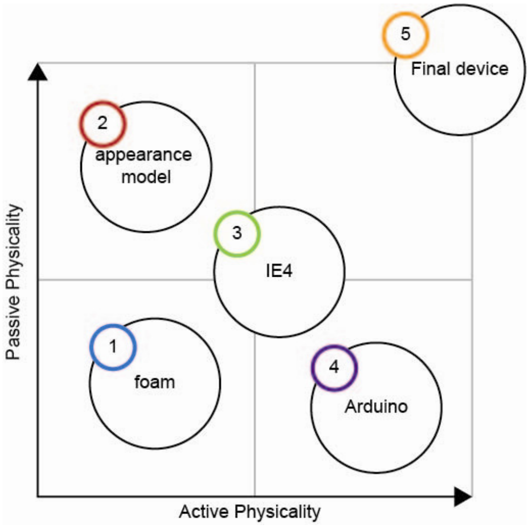


Fig. 10. The resulting physicality of each of the prototypes

2.3 The User Study

40 participants were recruited for the study (eight per prototype [17]), two did not turn up and three tests were rejected due to technical difficulties so the total number included in this analysis is 35.

16 of the participants were female and 19 were male. Participants were screened in accordance with the target market identified by iRiver to be between 23 and 45 years old; recruited participants fell predominately into the <28 (49%) or 29-33 (34%) age groups. All listened to music on a dedicated player or mobile phone and none had used the iRiver Spinn before.

Task-orientated trials, typical of usability trials, can be an effective way to demonstrate the product to a participant in a controlled manner and the participants were encouraged to 'think aloud' during the study to communicate their thought process [18]. Five tasks were chosen to introduce the participant sequentially to the device and no time constraint was imposed for the tasks. The tasks were:

Task 1: Turn the device on
Task 2: Find and play a specific track
Task 3: Adjust the volume of the track
Task 4: Stop the track and navigate to the first screen
Task 5: Turn the device off

Next, each participant was asked to scroll through the main menu titles and discuss what they expected within each menu. This user-led exploration ensured each participant had the same knowledge of the features of the device. After which a semi-structured interview sought to gain feedback about both the physical design and the users' interaction experience of the product. The explicit nature of the tasks and user-led exploration is one of the recommendations to reduce the evaluator effect on studies [19].

Finally, users were introduced to all the prototypes and asked to fill in a questionnaire ranking the quality of feel, appearance and quality of interaction for each of the prototypes. This enabled the participants to directly compare prototypes and offer an opinion about their construction.

Participants were bought into a controlled environment and the entire user trial was recorded on video. A facilitator ran the study with an observer monitoring the study via the video link. The observer was able to ensure continuity across the studies; this was deemed more suitable than introducing them as a second evaluator due to their level of experience with the prototypes and user testing methodologies. The Facilitator has conducted a number of similar studies before in a research and commercial context and is therefore able to reflect on techniques with colleagues of similar experience. Thus although the evaluator effect cannot be eliminated, it has been considered for this study [19].

3 Results of the User Trial

The analysis was performed by the facilitator. Discourse analysis provided a framework to analyse the video footage of the tasks, menu exploration and semi-structured interview. The strength of this approach is that it gives the ability to structure the conversational feedback typical of this type of study in a rigorous manner. The video footage was reviewed with event logging software and comments were assigned 'codes' based on the type of comment. 50 comment groups were recorded in total. In order to compare the prototypes comments made by just one participant were removed. These comments were then reviewed and collated to form high-level design recommendations typical of a report from user trials [20]. Further recommendations could be drawn from the data produced by the studies that would be used in a commercial context. For the purpose of this study only the comments that have emerged through the formal discourse analysis are included. It is important to note that the recommendations themselves are not important to this study and have therefore been simplified for this paper; it is the number of recommendations identified for each prototype in relation to the final device that is of importance in this context. The ten key comments that the design recommendations address are:

1. Help required from the facilitator
2. Difficulties in finding the required interaction
3. Tried other interactions
4. Pressed back to stop track playing
5. Tried turning dial to get to pause icon
6. Observation that it looks like a touch screen device
7. Like the 'Spinn' interaction
8. Long-winded interface
9. No unique selling point
10. Vertical menu navigation not obvious

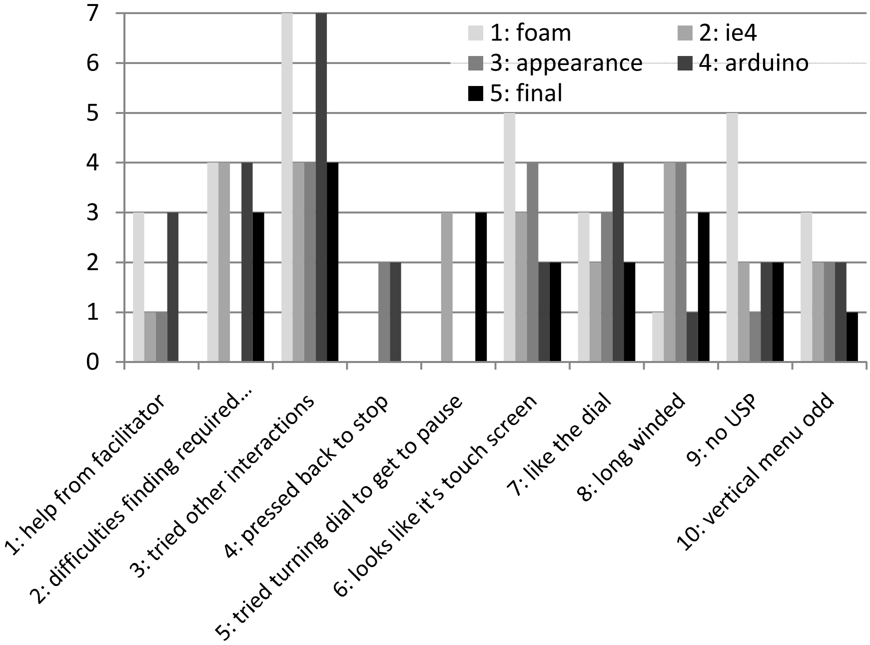


Fig. 11. The ten key comments addressed by the design recommendations

Figure 12 shows the results of the ranking exercise where each of the participants were introduced to all the prototypes and asked to give a rating where 6 is positive and 1 is negative. The participants were asked to rate three elements of the prototypes; the 'quality of feel' and 'appearance' which aimed to prompt the participant to consider the passive physicality elements and the 'quality of interaction' roughly equates to active physicality. Although these terms cannot be directly described as active and passive physicality, it goes some way to enable a comparison to the assessment of physicality shown in Figure 10. The data from the prototype the participant used for the study was not included to eliminate any bias from familiarity with the prototype. Figure 12 shows participants consider the foam prototype to have a low ranking but roughly equal for both elements which supports our assessment of the prototype to be low in both active and passive physicality. Likewise the appearance and Arduino are ranked in a similar way to our assessment. The IE4 gives interesting results with it being considered a higher quality of interaction than the Arduino and a more marked difference between active and passive physicality than anticipated. It could be that the visual aspects of physicality are undervalued in the current definition of passive physicality or that these questions are not adequate at obtaining participants views of active and passive physicality, this is beyond the scope of this paper but could be an interesting topic for further research. This exercise enabled participants to reflect on the prototypes themselves during the ranking exercise and the comments made were also captured, these will be brought into the discussion.

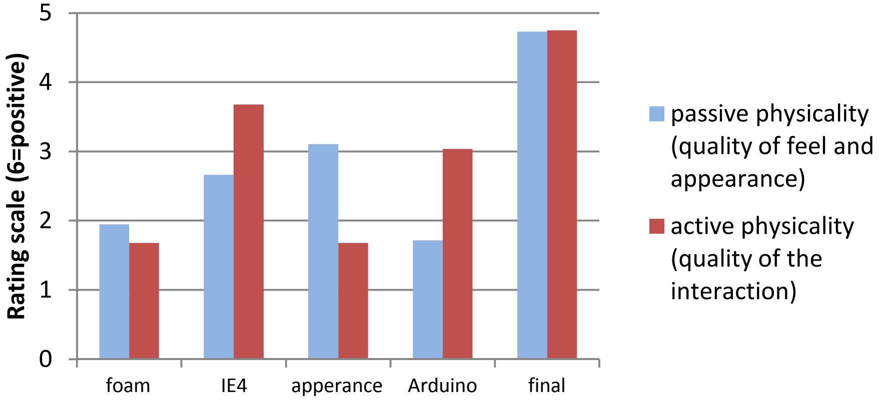


Fig. 12. Data from the ranking exercise; comparing the prototypes

4 Limitations of the Study

This study is recognized to have limitations that could be addressed in future work. The study has been designed, conducted and analyzed by one of the authors; therefore presumptions concerning active and passive physicality will inevitably influence the outcomes. Future work would seek to determine if the notion of active and passive physicality are applicable beyond this study. This is planned in a number of ways; firstly by re-evaluating studies conducted prior to the active and passive physicality notion, secondly by seeking discussion with those involved with interactive prototyping from an academic and commercial context, and finally by evaluating future studies conducted by colleagues.

5 Discussion

In Figures 11 and 12 the IE4 prototype appears to give feedback that is closest to the final iRiver device. These will be discussed along with other, more subtle, differences across the prototypes bringing in comments from the ranking exercise. Observations fall into two categories; recommendations about the design and obstructions caused by the prototype. Recommendations positively help identify how the design can be improved whilst obstructions are caused by features of the prototype that hinder participants in giving meaningful feedback.

5.1 Recommendations about the Design

Physicality of the Dial.
The IE4 prototype was the only prototype that highlighted participants trying to turn the dial to get to the pause function. The physicality of the dial itself could be the cause of this, for the IE4 each rotation has a distinct ‘click’ which causes a reaction in the interface. However the Arduino prototype did not produce this feedback and its dial had a similar physicality to the final device. This suggests that there must be something else about the prototype that causes the participant to miss feedback for this design recommendation. Several users made comments about the wires of the Arduino prototype being “very distracting” and looking “messier” than the other prototypes, this ‘messier’ appearance could possibly be the cause of this.

Information Architecture.
The feedback that the interface was longwinded was a common comment from participants of the trial with the final device. The IE4 and Appearance model were both good at drawing the same feedback. The Foam prototype was not able to elucidate this, possibly because the participant was not directly manipulating the prototype and therefore not creating the direct mental link between the physical and digital ‘I did not like the fact that I couldn't control the device (interface) from the model’. Meanwhile the Arduino prototype produced few comments about this possibly because the novelty of the prototype itself suppressed the participant’s potential frustration with the navigation of the interface “this thing (dial) works alright. I quite like the ability to click”. The IE4 seems to give a very direct feel between the interface and interaction, mimicking the final device well. The Appearance model forced the participant to have to continually press the scroll button to navigate the interface, highlighting the sheer number of button presses required to navigate the interface “Very tedious going through all the songs like this”.

5.2 Obstructions Caused by the Prototypes

Modeling Physical Interfaces on a Touch Screen.
The Appearance model used a touch screen for the interactive element of the prototype. This prototype gave participants the least difficulties in finding the interactions. Due to the need to represent all the buttons on a touch-screen this prototype clearly indicated where interactions were, even when they were on the side of the device. This made the interactions more obvious for those using this prototype than would otherwise have been. Paradoxically, the very usability of the touchscreen prototype devalues it given the issues users had with the real device.

Obstacles to the Participants Understanding the Prototype.
Figure 11 shows the Foam and Arduino prototypes forced participants to ask for the most help from the facilitator. The Foam model requires the participant to fully engage with the ‘speak aloud protocol’ because the buttons provide no active feedback. The participant therefore has to wait for the facilitator to operate the interface. In contrast, the Arduino prototype allows the participant to operate it independently, but it may be that the appearance of the wires that seem to the biggest barrier to acceptance.

It may also be that techniques which require the participant to understand the *way* in which the prototype works are not suited for this type of early stage trial.

5.3 Overview of the Four Prototypes

The IE4 Prototype.
The real-time nature and simplicity of this prototype seem to be the important factors in making this prototype the most effective of the prototypes. Participants were able to operate and receive immediate feedback from the interface without an overly complicated looking prototype or altering the scale and form of the model. “I felt very little difference in terms of the final version and white model (IE4) for the quality of interaction - white model (IE4) had a few blips but nothing that is stopping me using the device successfully.” “The addition of working buttons on the prototypes increases the quality of the feel, as the ways in which interaction occurs can be more readily envisioned.”

The Foam Prototype.
This prototype used the ‘speak out loud’ protocol for participants to engage with the interface. Results show that this prototype was less effective at enabling participants to build a mental model of the device resulting in reduced effectiveness of the comments received. “The colour, weight, size and cable connections play a big part of my initial interaction with a product, for this reason the blue foam compared to the final unit was clearly a visual aid as opposed to actual real product comparison.”

The Arduino Prototype.
Participants required more assistance using this prototype. This was a surprise from the most interactive of the prototypes. Participants seemed to be affected by the wires and appearance of this prototype. “The model with blue foam & wires looks messier than the blue foam model but it looks a little bit more functional than the model with blue foam alone.”

The Appearance Prototype.
This prototype used a touch screen to convey the interactions of the prototype. Participants did not identify as many usability errors and had the weakest performance in relation to the final device. This outcome supports Gill et al.’s study in which it was proposed that interactions are easier for a participant to identify on a screen [4]. “Although the silver model (appearance model) looked more like the final version, I did not like the fact that I couldn't control the device from the model and I didn't think having the model alone, without much interaction, was very worthwhile.”

6 Conclusion and Application

The four prototypes trialled in this study explored different aspects of active and passive physicality. The results show that both active and passive physicality are important considerations for early stage user feedback; but it is an even balance of these that produces the most effective prototypes, as seen in the IE4 and Foam prototypes.

Resources should not be used exclusively to ensure the prototype functions well in an electronics and interaction sense (active physicality) if it severely impacts the way the prototype looks or can be held by the user (passive physicality). Likewise, resources spent creating a prototype that looks very close to a final device are not effective if interactions are not well supported.

The IE4 and Foam prototype provided the most accurate data compared to the user experience of the real device. Both the IE4 (£760) and Foam prototype (£60) were of balanced physicality. The Arduino (£1,100) was very strong on active physicality to the detriment of passive physicality whilst the Appearance model (£1,160) was very high on passive physicality but low on active physicality. This suggests that it is those prototypes that are well balanced that are the most effective in this study. Since they are also cheaper they represent strong value for money.[2]

The prototype has long been accepted as a valuable approach to creating valuable and insightful design outputs. However, for interactive devices that have both a physical and digital form, visual fidelity alone is clearly not enough to fully conceive the complete prototype and ensure it will accurately fulfil its purpose. Whilst visual and dimensional fidelity is very much the staple of prototyping, physical fidelity clearly has a role in creating a well-targeted prototype. This study indicates that for interactive prototyping, 'physicality' needs to be an even combination of active and passive physicality.

7 Future Work

Future work needs to be conducted to determine if active and passive physicality can be usefully used in assessing prototypes beyond those used in this study. The outcome of this study indicates that a balanced prototype is the most effective. The prototypes used in previous studies [4] [5] should now be assessed in terms of physicality to determine for example if notions of active and passive physicality aid in determining why the data for the 'flat-face' prototype differed considerably from the final device. In addition prototypes used in studies by other authors could be categorized to see how they relate to our prototypes.

References

1. Virzi, R.A., Sokolov, J.L., Karis, D.: Usability problem identification using both low and high-fidelity prototypes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground, Vancouver, British Columbia, Canada (1996)
2. Lim, Y.-K., Stolterman, E., Tenenberg, J.: The Anatomy of Prototypes: Prototypes as Filters, Prototypes as Manifestations of Design Ideas. ACM Transactions on Computer-Human Interaction 15(2) (July 2008)
3. McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, B., Vera, A.: Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montreal, Quebec, Canada (2006)

[2] Costs are those shown in Figure 4 minus the software prototyping (shared by all prototypes).

4. Gill, S., Walker, D., Loudon, G., Dix, A., Woolley, A., Ramduny-Ellis, D., Hare, J.: Rapid Development of Tangible Interactive Appliances: Achieving the Fidelity / Time balance. International Journal of Arts and Technology, 309–331 (2008)
5. Hare, J., Gill, S., Loudon, G., Ramduny-Ellis, D., Dix, A.: Physical Fidelity: Exploring the Importance of Physicality on Physical-Digital Conceptual Prototyping. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 217–230. Springer, Heidelberg (2009)
6. Dix, A., Ghazali, M., Gill, S., Hare, J., Ramduny-Ellis, D.: Physigrams: Modelling Devices for Natural Interaction. Formal Aspects of Computing 21(6), 613–641 (2009)
7. Reeves, S.: Physicality, spatial configuration and computational objects. In: First International Workshop on Physicality (2006)
8. Gibson, J.: The theory of affordances. In: Perceiving, Acting, and Knowing, John Wiley & Sons Inc., Hoboken (1977)
9. Norman, D.: The Design of Everyday Things. Basic Books, New York (1988)
10. Norman, D.: The Invisible Computer: Why good products can fail, the Personal Computer is so complex, and information appliances are the answer. MIT Press, Cambridge (1998)
11. Nielsen, J.: Usability Engineering. Academic Press Inc. (1993)
12. Lim, Y., Pangam, A., Periyasami, S., Aneja, S.: Comparative Analysis of High- and Low-fidelity Prototypes for More Valid Usability Evaluations of Mobile Devices. In: Proceedings of NordiCHI 2006 Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles (2006)
13. Maudsley, Greenberg, Mander: Prototyping an intelligent agent through Wizard of Oz. In: Interchi 1993, pp. 277–284 (1993)
14. Gould, J., Lewis, C.: Designing for Usability-Key Principles and what Designers Think. In: CHI 1983, Boston, pp. 50–53 (1983)
15. Gill, S.: PAIPR Prototyping: Some Thoughts on the Role of the Prototype in the 21st Century. In: Prototype: Craft in the Future Tense. Berg (in press, 2013)
16. Burleson, W., Jensen, C.N., Raaschou, T., Frohold, S.: Sprock-it: a physically interactive play system. In: Proceedings of the 6th international Conference on Interaction Design, Aalborg, Denmark (2007)
17. Goodwin, K.: Designing for the Digital Age: How to create human-centered Products and Services. Wiley Publishing Inc. (2009)
18. Dumas, J., Redish, J.: A Practical Guide to Usability Testing (revised edition). Intellect, Exeter (1999)
19. Hertzum, M., Jacobsen, N.: The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interaction 15(1), 183–204 (2003)
20. Molich, R., Jefferies, R., Dumas, J.: Making Usability Recommendations Useful and Usable. Journal of Usability Studies, 162-179 (2007)
21. Buchenau, M., Suri, J.: Experience prototyping. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, New York, pp. 424–433 (2000)

CapTUI: Geometric Drawing with Tangibles on a Capacitive Multi-touch Display

Rachel Blagojevic and Beryl Plimmer

Department of Computer Science
University of Auckland
Private Bag 92119
Auckland, New Zealand
rachel.blagojevic@auckland.ac.nz, beryl@cs.auckland.ac.nz

Abstract. We present CapTUI, an innovative drawing tool that detects tangible drawing instruments on a capacitive multi-touch tablet. There are three core components to the system: the tangible hardware, the recognizer used to identify the tangibles and the drawing software that works in tandem with the tangibles to provide intelligent visual drawing guides. Our recognizable tangible drawing instruments are a ruler, protractor and set square. Users employ these familiar physical instruments to construct digital ink drawings on a tablet in an intuitive and engaging manner. The visual drawing guides enhance the experience by offering the user helpful cues and functionalities to assist them to draw more accurately. A user evaluation comparing CapTUI to an application with passive tools showed that users significantly preferred CapTUI and found that the visual guides provide greater accuracy when drawing.

Keywords: TUI, tangible, multi-touch, physical interaction, capacitive, drawing tools.

1 Introduction

CapTUI is an innovative drawing application for capacitive multi-touch tablets that detects tangible drawing instruments. To develop CapTUI we have designed and constructed tangibles, developed a recognizer to detect the tangibles on a capacitive touch display, and built drawing software that works in tandem with the tangibles to provide intelligent visual drawing guides.

A skill acquired very early in life is the ability to manipulate tangible objects. As a case-in-point, mathematics education has long used manipulatives for introducing young children to mathematical concepts as their intuitive understanding of the physical enables them to transition into the realm of the conceptual [1-3]. Using tangibles for computing operations takes advantage of this existing skillset: pen-based interaction is a testament to this. Tangibles lower the level of interaction abstraction allowing users to apply their natural tool-based skillset to the digital environment. The use of tangibles on touch displays has been shown to improve interaction with interface objects as they are easier to manipulate, acquire and control in comparison to virtual objects [4].

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 511–528, 2013.

Though the stylus has long been available as a drawing instrument for digital canvases, other instruments have not yet been explored. In addition to stylus input, designers often use passive tools such as rulers or French curves on tablets to guide drawing. Converting these passive tools to active tools seems a promising evolution for drawing based content creation. We have created three recognizable tangible drawing instruments: a ruler, protractor and set square. Through our application, users are able to apply these familiar physical instruments to construct digital ink drawings on multi-touch tablets. Transforming passive instruments into active tools, as we have done here, allows further intelligence to be built into the system. To our knowledge, recognizable tangible drawing tools (other than styli) have not been used on capacitive touch displays for creating electronic content.

Touch surfaces are a well-established technology. Small touch sensitive devices, e.g. phones and tablets, typically have capacitive displays, larger touch devices, such as tabletops, have camera-based detectors. Tangibles on tabletops have been explored for many years e.g. [5-8]. These systems typically use cameras to capture the position of the tangible; this requires specialist equipment and environments. The advantage of capacitive touch is that it is built into many readily available tablets that are portable and require no special environment. Using capacitive touch alters the recognition approach from image processing to gesture processing of the touch points. Touch points may be fingertips or conductive material [9] given that capacitive displays detect touch via the electrical properties of fingertips. As an extension of recent techniques [9] we have devised a novel approach which enables tangible drawing tools to work with capacitive displays (Fig. 1).

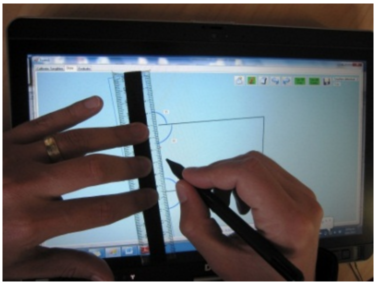


Fig. 1. CapTUI

To our knowledge tangible geometry drawing tools have not been used in this context before. There are two key contributions of this research: the design and construction of tangible drawing tools for capacitive touch surfaces, and visual drawing guides to enhance the geometric drawing experience. Together, these enhancements have the potential to provide benefits over using passive drawing tools on a screen. In a user evaluation, we explore this idea by comparing our prototype, CapTUI, to drawing with passive tangibles. Possible areas of application for our system include design fields such as architecture and mathematics education.

2 Related Work

Physical objects are beneficial for learning [10, 11]. For example, manipulatives have been used for years to assist in mathematics education [1-3]. Similarly tangibles are a particularly useful tool for children when interacting with technology because of the added physicality [4, 12, 13]. Other advantages of tangibles include increasing explorative behaviour [14], reducing conflicts in cooperation [15, 16], encouraging prolonged engagement [17, 18] and acting as a useful aid to problem solving in comparison to standard graphical user interfaces [19]. Tangibles have been used to enhance physics and mathematics learning [20, 21] while other e-learning tools link geometry with algebra in order to augment the learning experience [22, 23].

The use of tangibles on touch surfaces has been shown to improve interaction with interface objects. Tuddenham et al [4] compared the use of tangibles and virtual objects on multi-touch interfaces. They found that when using the tangible objects significantly fewer errors were made and significantly less time was taken to acquire and manipulate the objects.

Limited work has been done in the area of tangibles on capacitive touch screens. SmartSkin is an early system that used capacitive sensing with a mesh shaped antenna to detect hands and objects [24]. Conductive materials were attached to blocks in patterns to allow objects to be identified. Yu et al [9] present three tangible technologies for use on capacitive touch surfaces: spatial, frequency and hybrid. Their spatial tangibles employ patterns for object identification, from at least four touch points. Four touch points can be limiting if users have multiple tangibles on touch surfaces that are limited in the number of touch points they are able to detect. Frequency tangibles use a modulation circuit to generate touches of varying frequency. They only require one touch point; however they rely on a power source, are unable to detect object orientation, and are limited in tracking fast movements. The hybrid tangible combines the spatial and frequency technologies to address the aforementioned issues. Our observation is that the tangibles they have presented are quite large in size – particularly thickness, in comparison to our tangibles. They report that the size of the circuit board alone is 2x3x3cm. AppMATes[1] are another example of tangible objects, in the form of toy cars that can be used on the Apple iPad. The cars are uniquely identifiable using touch point patterns. More recently [25] stackable tangibles, sliders and dials for capacitive screens have been introduced. The stackable tangibles are able to sense changes in capacitance when blocks are placed on top of them – this in turn modifies the touch point pattern for identification. The dials and sliders are made of conductive zebra rubber and also use unique touch point patterns for identification.

Very few studies have been performed testing the usability of tangibles on capacitive displays. One study compared CapWidgets [26], tangible dials designed for mobile capacitive screens, to touch dials. The results showed that touch was significantly faster and rated higher for usefulness, satisfaction, ease of use, and learnability but there was no significant difference in precision. They identified aspects of the design of the tangibles and the software to be the main hindrance to the usability of their system. Jansen et al [27] designed tangible remote controllers for wall sized displays. The tangibles are manipulated on a capacitive tablet as a way of interacting with a wall sized display. The tangibles were stuck to the tablet to allow mobility and to reduce the need for visual attention. They used various conductive materials including conductive foam and ink to construct the tangible sliders and dials. Their user study compared tangible and touch interaction for a slider. Overall they found that using the tangible slider produced fewer errors, was faster to acquire and did not

[1] `http://www.appmatestoys.com/`

require as much visual attention as the virtual slider. However, they did comment that their tangible design "still does not measure up to commercial physical controls".

Geometric drawing applications are common. Most rely on mouse and keyboard input [22, 23]. Sketchpad [28] presents the earliest work using a light pen to construct line drawings. It includes the ability to apply geometric constraints to drawings. Balakrishnan et al [29] introduced the concept of drawing on a large scale with digital tape. Tangible rulers for camera-based systems have been used [4, 30] but their focus has been on investigation of object acquisition or positioning rather than how the tangible might facilitate drawing and precise measurement.

Although the use of tangibles and multi-touch has been investigated in the context of camera-based surface interaction, tangible drawing tools with capacitive screens is yet to be explored. The interaction with drawing tools differs from the dials, blocks and sliders from previous work in tangible user interfaces (TUI's) for capacitive screens. Of the few studies that have been performed it is evident that the tangible's design is the key challenge to success [26, 27]. The CapWidgets study [26] suggests that tangible design was a hindrance to the usability of their system, while [27] also allude to the difficulty of constructing good tangibles. The combination of tangible drawing tools with capacitive screens presents significant challenges in the hardware and the interaction design of the applications to ensure a good user experience.

3 Our Approach

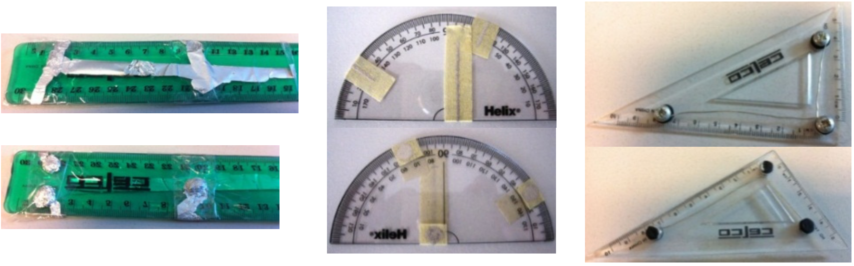
CapTUI combines recognizable tangible drawing tools with an intelligent drawing application. In this section the tangible hardware design is described, followed by the recognition approach used for these tools. Section 3.3 provides details on the drawing application and finally Section 3.4 describes CapTUI's visual drawing guides.

3.1 Tangible Hardware

Our goal is to create easily manipulated drawing instruments that are recognizable on a capacitive touch display. Capacitive screens are designed specifically to detect the electrical properties of fingertips. Hence we use conductive materials that act as an extension of the fingertips for the touch point connections between the tangibles and the touch display. These touch points are electrically connected to the user's body via his or her fingers as the user's body is the source of the electrical current.

To discriminate between the tangibles each requires a unique ID. The touch points on the bottom of each tangible are placed so as to form unique patterns. This ID allows differentiation between tangibles, and calculation of a tangible's orientation and location (see Section 3.2 for a full explanation).

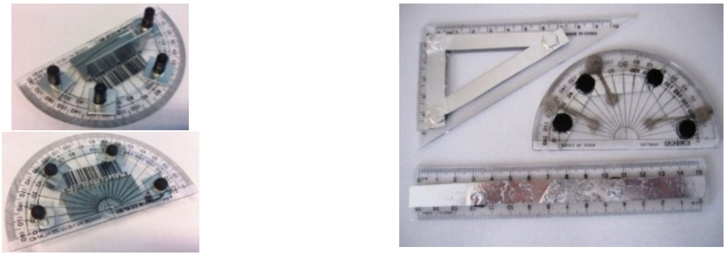
There were several iterations in our design process for constructing the tangible hardware. Off-the-shelf standard drawing tools were modified by adding conductive materials to allow detection on the capacitive touch screen. Fig. 2 shows some of our designs using various conductive materials such as: aluminum foil/tape, conductive rubber, conductive foam, steel nuts and bolts, copper wire, and conductive ink [31].



a) Aluminium foil and tape.

b) Conductive ink.

c) Steel nuts and bolts, copper wire and conductive rubber.



d) Metal washers, copper wire, conductive rubber and metal handle.

e) Conductive foam and aluminum tape.

Fig. 2. Tangible hardware designs

After preliminary testing of these prototypes, the following factors affecting the detection and use of the objects were identified.

- As a whole, the materials used must ensure a consistent circuit can be maintained, i.e. all circuit components are well connected so that the current from the user's fingertips has an unbroken path to the screen. We found when several materials are used, and therefore more connections required, a significant amount of pressure had to be applied before the touch points were detected.
- Tangible stability is important to maintain consistent contact between the screen and the tangible. This means that the touch points must be as flat as possible and positioned such that when the tangible is moved across the screen all points stay in contact with the surface. If the tangible is unstable, and a touch point loses contact with the screen then the tangible is no longer detected.
- To maintain tangible to body contact, there must be a clear point of contact for the user's fingertips when the tangible is in use. Our final prototype's top surface is covered with conductive material (Fig. 3).
- Usability of the tangible is poor if there is too little friction between the screen and the tangible. Using materials that add friction between the tangible and the screen, reduces unwanted movements when drawing, allowing users to draw along the tangible easily without having to concentrate on keeping it steady.
- Designs that require the conductive material to pass over the edge of the tangible block the user from drawing comfortably on the tangible edge (such as the prototypes shown in Fig. 2(a) and (b)).

Table 1. Tangible hardware designs vs design factors

	Consistent circuit	Stable	Clear point of body contact	Friction with screen	Comfortable drawing (material does not pass over edge)
a) Aluminium foil and tape			✓		
b) Conductive ink	✓	✓			
c) Steel nuts and bolts, copper wire and conductive rubber				✓	✓
d) Metal washers, copper wire, conductive rubber and metal handle			✓	✓	✓
e) Conductive foam and aluminium tape		✓	✓	✓	✓
Final design (conductive foam)	✓	✓	✓	✓	✓

Table 1 shows the performance of each tangible design in terms of the design factors listed above; where performance is based on expert judgement and informal testing. Design (a) was a proof of concept to show that recognition was possible; however it failed to meet most of the design criteria. Design (b) provided a consistent circuit and was a stable design, however its main drawback was that the conductive ink had to be drawn on tape to ensure the circuit was maintained over the edge of the tangible. An experiment using the ink without tape failed as the edge of the tangible was so thin it was difficult to maintain an unbroken circuit. Design (c) and (d) involved too many materials and therefore made it difficult to maintain a consistent circuit. The conductive rubber touch points, while providing a good amount of friction with the screen, were difficult to mould and therefore did not provide a stable base for the tangibles. Design (e) had conductive foam for the touch points which improved stability, as they are easier to cut into shape, and maintained friction with the screen. However it was difficult to maintain good circuits as there were still two conductive materials in use.

Our final design (Fig. 3) was constructed based on the identified design factors; these included a ruler, set square, and protractor. One conductive material was used, conductive foam. This reduces the number of connections that must be made to maintain a circuit from the user's fingers to the screen. The foam covers a large portion of the top surface of the tangible so that there is a clear point of contact for the user's fingers. The foam provides a good amount of friction so that the tangible does not make unwanted movements on the screen. It also provides a stable base for the tangible so that the touch points are consistently in contact with the screen. In addition the conductive foam is easier to work with than other materials as it can be cut into the shape required and does not have to pass over the edge of the tangible.

Fig. 3. Final tangible drawing tools (conductive foam)

3.2 Tangible Recognition

This section describes how each tangible can be identified and its position computed. The recognition approach is similar to that of SmartSkin [24]. Recognition of the tangible allows for further intelligence to be added to the tangible interaction as described in Section 3.4.

The tangible ID patterns are based on the distance between each pair of points. When constructing a pattern the distances must be distinct; for example, a pattern that forms an isosceles triangle is not allowed because it has two or more identical distances between points. Identical distances make it impossible to determine the correct orientation of the tangible, with the obvious exception of regular shapes (e.g. circle).

There are two main phases for the recognizer: learning and recognition. The learning phase allows users to register and calibrate the tangibles so that they are identified by the system. Several details about each tangible are stored: the location of the touch points, the distances between touch points (i.e. its unique ID), the outline points of the tangible and the type of tangible (ruler, protractor or set square) as specified by the user. This data is used later to identify the tangible and display its outline on the screen.

Tangibles only need to be registered with the system once. To register the tangible (Fig. 4), the user places it on the registration screen and moves the guidelines to delineate a bounding box. The user then specifies the type of tangible (ruler, set square, protractor) from a list.

During the recognition phase, touch input from the tangible currently in use is passed to the recognizer. The recognizer finds its ID by calculating the distances between each touch point detected and comparing these to the previously saved distances for each calibrated tangible. The recognition adds (or subtracts) a degree of error tolerance (*e*) when comparing the distances between the detected touch points and the saved set of distances. The error tolerance is required as capacitive hardware detects an area of touch but only reports the centre point position. This means the precise position of a touch point can be difficult to determine as we have no way of knowing which part of the tangible's touch point has been detected as the centre. We experimented with different sized touch points and found that a diameter of approximately 7mm is the smallest size that can still be detected on the touch screen for our tangibles.

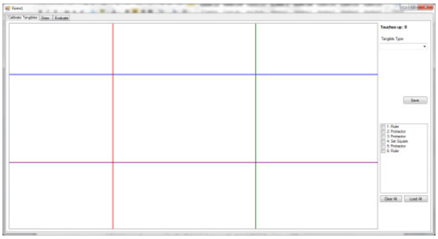


Fig. 4. CapTUI tangible calibration interface

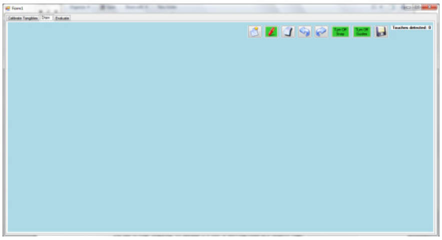


Fig. 5. CapTUI drawing interface

The tangible identification algorithm is as follows:

```
d: detected distances; s: saved distances; e: error tolerance
i, j: number of distances
For all d_i and s_j,
   if:  d_i - e <= s_j <= d_i + e;  then: d_i = s_j
```

If a match is found between all d_i and s_j then the tangible ID is identified as s. Our informal experiments showed that a 2mm tolerance level is sufficient for maintaining a good rate of recognition.

The advantage of including a level of tolerance for the patterns is that 100% accuracy can be achieved by the recognizer, resulting in better usability. There is a tradeoff here, as calculating the position and orientation of the tangible is less accurate. To minimize the effect of this tradeoff the user is provided with a visualization of the tangible on the screen; this is described further in Section 3.4.

3.3 Drawing Application

The tangible hardware and recognition algorithm were combined together in a drawing application (Fig. 5). Our application allows the tangible ruler, protractor and set square to be used to construct drawings on a capacitive multi-touch screen. It also includes basic functions such as save, erase, undo and redo.

To construct drawings using a tangible drawing tool the user slides a stylus along the edge of the tangible that is placed on the screen (Fig. 1). As the user draws, beautified lines (or curves) are rendered.

The application was developed using the .NET 4.0 framework on a Windows 7 Dell Latitude XT3 tablet, Intel core i7-2640M with 8GB RAM, built in stylus and four available touch points.

Ink Beautification

One of the main purposes of having geometric tools in drawing is to use the tangibles as a guide for precise drawing. However, due to the imprecise nature of contact points on the screen, the user's ink may not be rendered where they intend. Therefore we beautify the ink to try to produce what the user intends, in a similar way to many digital ink drawing programs such as [32]. Our system includes two forms of ink beautification: ink to tangible edge snapping and ink corner snapping.

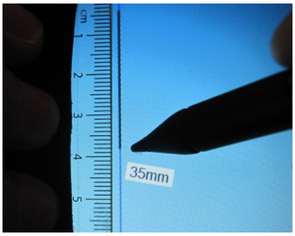


Fig. 6. Ink-to-edge snapping and length visualization

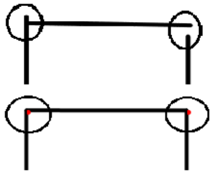


Fig. 7. Corner snapping

For ink-to-edge snapping, we assume if a user is drawing close to a tangible edge, they intend to draw right on the edge. Therefore ink points are dynamically translated so that they snap to the edge of a tangible's virtualized outline as they are drawn in real time. The result is a beautified straight/curved line drawn along the edge of the tangible (Fig. 6).

Corner snapping beautifies ink to render clean intersections between strokes that are close by. It is automatically enabled in two situations. The first connects the starting point of a stroke to the end of existing strokes when the user first starts drawing a stroke. The second connects the stroke end point to other strokes (at any position) once the stroke is completed. The beautification is applied to the newly added stroke and existing strokes; both strokes are extended or reduced such that they meet exactly at an intersection point (Fig. 7). A corner is snapped if the current stroke end points are close to another stroke – this requires a distance threshold to be set for determining 'closeness'. As this is heuristic-based, corner snapping can be reversed using the undo function. It can also be turned off if the user desires.

3.4 Visual Drawing Guides

Our early experiences suggested that drawing guides are essential for an enjoyable experience on current touch technologies [33]. The visual drawing guides we developed assist the user in constructing accurate geometric drawings. These guides are directly dependent on the unique identification of the tangibles; therefore making them only possible and relevant with recognizable tangibles.

Tangible Outline

When the tangibles are detected on the screen a visualization of the tangible outline is displayed (the blue line in Fig. 8). The visualization is rendered at the correct location and orientation, using the tangible recognition information, so as to match the tangible itself, though with a small offset such that the outline is visible.

Length Visualizations

When drawing straight lines using the ruler or set square, an adjustable line length function is enabled. As a stroke is drawn along the edge of a tangible, its length can be adjusted by moving the stylus up and down the edge. This function has two main advantages: when the user draws the line further than the intended length it can be easily corrected, and having a line dynamically adjustable in length provides a prototyping experience where the user can see how the line will look at various lengths.

To assist in solving occlusion issues caused by the tangible, stylus or hand on the display, the length visualization displays line length in millimetres as the user draws with a ruler or set square (Fig. 6). This visualization is supplementary to the markings on the tangible. The visualization works in tandem with the adjustable length function by providing real time stroke length feedback.

Without recognition of the tangible these visualizations would not be suitable functions as there would be no way of knowing that the user intends to draw a straight line as they are completing the action. When the ruler or set square is recognised we can be sure that straight lines are desired.

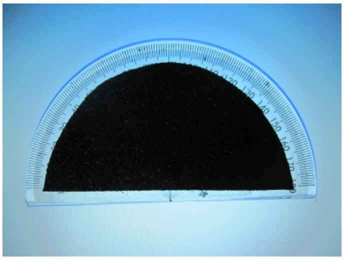


Fig. 8. Tangible outline

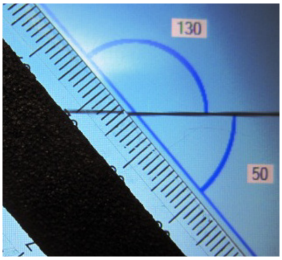


Fig. 9. Angle visualizations

Angle Visualization

The angle visualization displays the angle between existing strokes and the edges of a tangible ruler to allow a line to be drawn intersecting other lines at specific angles (Fig. 9). It is created by detecting intersections between existing strokes and the ruler edges, calculating the angle between these lines, and displaying the results. This enables the construction of various geometric shapes which consist of connected lines of specific angles, for example, regular polygonal shapes. We have also included an affordance which freezes the angle in place if it reaches multiples of five or ten degrees to assist the user in obtaining common angles. If the angle has been frozen, it is only unfrozen if the tangible is moved approximately 5 degrees away. This threshold was chosen after some informal testing. In essence, this visualization can replicate the protractor for measuring angles. However, it relies on the recognition of the ruler so that the angles between the tangible ruler and existing digital ink can be calculated in real time.

In a similar way to the length visualization, an angle visualization displays the angle in degrees as the user draws an arc with the protractor. This helps to solve occlusion issues and supplements the markings on the tangible itself by providing real time feedback. Again, this visualization relies on the recognition of the protractor to confirm that the user intends to measure angles.

4 Evaluation

The objective of this evaluation is to determine if CapTUI assists users to easily draw precise geometric drawings. In particular, we are interested in comparing drawing with CapTUI's recognizable tangibles to drawing with standard non-recognizable drawing tools on a screen. For this comparison we have developed another drawing application, referred to as Paint. Paint can be used to construct geometric drawings in much the same manner as CapTUI except that it does not recognize the drawing tools on the screen. To make the comparison as fair as possible it includes the same beautification and corner snapping functions as CapTUI. However, it does not include CapTUI's tangible outline, angle and length visualizations, as these functions are dependent on tangible recognition. Using this comparison we can determine if recognizable tangibles and visual guides, that are possible as a result of this recognition, assists users when constructing precise geometric drawings.

A within-subject design was used where each participant was given the same tasks to complete using CapTUI and Paint; half used CapTUI first and the other half used Paint first. Participants first completed a pre-questionnaire on their previous experience with touch, stylus and drawing applications. They were then given an introduction to the first system and time to familiarize themselves with the application. When ready, participants completed the tasks required using the first system and filled in a questionnaire on their experience. This process was repeated for the second system.

There were seven tasks designed for the evaluation; the first three tasks served as training tasks (Fig. 10). Each training task aimed at familiarizing the participant with a different drawing function or tangible. They were not told that they were training tasks: this was done to try to encourage them to complete the tasks with the same amount of effort they might apply for the real tasks and therefore get the most out of the training. The remaining four tasks comprised the evaluation tasks, providing data for our analysis (Fig. 11). The focus of task one was on connecting lines of specific length. Task two was aimed at evaluating drawing lines at a specific angle and use of the protractor to connect an arc. Task three focused on drawing angles of specific magnitude. In task four, angle and length measurements were required concurrently.

Quantitative metrics used to analyse the data collected from these tasks included the time taken, and average length and angle errors. Errors were counted as 1mm or 1 degree away from the required measurement. The average length and angle error was calculated for each task using the sum of the errors made / number of lengths or angles in the task. Participants also completed a questionnaire to record their experiences using each system. Questions were presented using a 5 point Likert scale, except for a small number of open ended questions. A Wilcoxon Signed Ranks Test was used to test for significant differences in the results, unless otherwise stated.

After a pilot study with two participants, twelve participants (eight males and four females) were recruited for the final study. They came from a range of backgrounds including computer science, law and health. Seven participants used touch interfaces frequently; the remainder had used such interfaces a few times or occasionally. The majority of participants had experience using stylus input before. Six participants had used a drawing program on a touch device before.

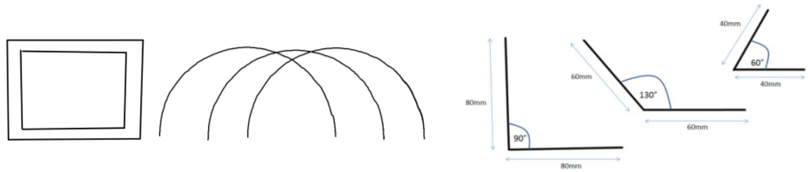
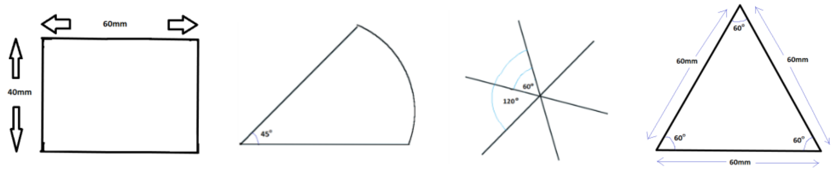


Fig. 10. Training tasks



Task 1 Task 2 Task 3 Task 4

Fig. 11. Evaluation tasks

4.1 Results

Participants were asked to give an overall rating for each system (using a 5 point Lik ert scale). Seven participants rated CapTUI higher than Paint, one rated Paint higher and four considered them equal. All participants rated CapTUI as good, very good or excellent (m = 3.75, s.d. = 0.62) (see Q1, Table 2). Ratings for Paint were more varied (m = 2.83, s.d. = 1.11). The overall ratings were found to be significantly different (z = -2.157, p = 0.031) where participants rated CapTUI significantly higher than Paint.

The results for the quantitative measures are as follows. A paired t-test showed that there was no statistically significant difference (t = 1.644, p = 0.107) in the time taken to complete the tasks between using Paint (m = 58.30, s.d.= 32.78) and CapTUI (m = 70.36, s.d. = 32.63), Fig. 12. This result is favourable for CapTUI as it shows that even with the time taken for detecting the drawing tools it did not take significantly longer to complete the tasks. There was no significant difference (z = -1.029, p = 0.304) in average length errors per task between using Paint (m = 1.79, s.d.= 2.67)

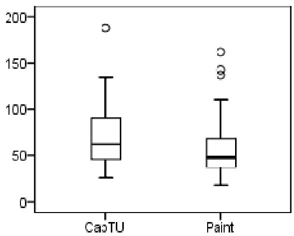


Fig. 12. Time taken to complete tasks (seconds)

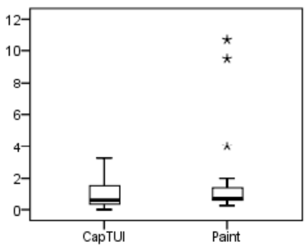


Fig. 13. Average length errors

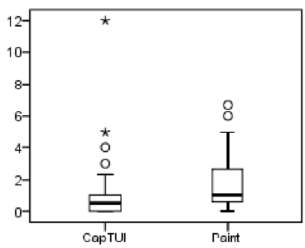


Fig. 14. Average angle errors

Table 2. Comparative Questionnaire Results (1st row: CapTUI, 2nd row Paint)

	Poor	**Fair**	**Good**	**Very Good**	**Excellent**
Q1. Overall rating			4	7	1
	1	4	4	2	1
	SD	**D**	**N**	**A**	**SA**
Q2. Drawing tools ease of use		1	5	5	1
	1	1	2	7	1
Q3. Drawing tools usefulness				8	4
	1		2	5	4
Q4. Ease of length measurement		1		4	7
	1	2	2	5	2
Q5. Ease of angle measurement		1	1	7	3
	2	3	1	4	2
Q6. Drawing accuracy		1	4	5	2
	1	5	1	4	1
Q7. Drawing tidiness			4	5	3
	1	6	1	3	1
Q8. Enjoyment				8	4
	1	2	2	7	

and CapTUI (m = 0.95, s.d. = 0.82), Fig. 13. This shows that both Paint and CapTUI were comparable when creating lines of accurate lengths. On the other hand, average angle errors per task were significantly less (z =-2.861, p = 0.004) when using CapTUI (m = 0.96, s.d. = 1.91) than when using Paint (m = 1.68, s.d.= 1.63), Fig. 14. This result confirms that angles drawn using CapTUI were significantly more accurate.

The following are the results obtained from questions answered using a 5 point Likert scale (see Table 2). Participants enjoyed (Q8) using CapTUI (m = 4.33, s.d.= 0.49) significantly more than using Paint (m = 3.25, s.d.=1.06), (z = -2.288, p = 0.010). All participants agreed that the drawing tools were useful (Q3) when completing the tasks for CapTUI (m = 4.33, s.d. = 0.49); however for Paint the results were more varied (m = 3.92, s.d. = 1.16). There was no significant difference found between these results (z= -0.877, p = 0.380). There was no significant difference found for the responses when asked if the drawing tools for each system were easy to use (Q2, z = -0.272, p = 0.785). This indicates that the recognizable drawing tools were just as easy to use as the non-recognizable tools. When asked which system is easier to use overall, five participants chose CapTUI, three chose Paint, and four considered them equal. The main reason for choosing CapTUI was that the visual guides made the system much easier to use than Paint ("the angle and length indicators were really helpful", "easier especially for the more technical drawing due to the ability to easily get exact lengths and fairly easily get precise angles"). The main reason for not choosing CapTUI was due to inconsistent tangible detection ("CapTUI was more accurate and tidy but I had a bit of problem with adjusting the tools").

All but one participant agreed that it was easy to measure line length (Q4) with CapTUI (m = 4.42, s.d. = 0.90) in comparison to seven participants for Paint (m = 3.41, s.d. = 1.24), however this difference was not found to be significant (z = -1.796, p = 0.072). Ten participants agreed that it was easy to measure angles (Q5) with CapTUI (m = 4.00, s.d. = 0.85), compared to only six participants for Paint (m = 3.08, s.d. = 1.44); this was also not statistically significant (z = -1.530, p = 0.126). It is possible that with a larger sample size statistically significant differences may be found for these factors given that the majority of participants preferred CapTUI for ease of line and angle measurement.

Eight participants agreed that their drawings were tidy (Q7) when drawn using CapTUI (m = 3.92, s.d. = 0.79), compared with four participants for Paint (m = 2.75, s.d. = 1.22). These differences were statistically significant (z = -2.124, p = 0.034). Only one participant did not believe their drawings were accurate (Q6) with CapTUI (m = 3.67, s.d. = 0.89), compared with half of participants with Paint (m = 2.92, s.d. = 1.24), however this difference was not statistically significant (z = -1.562, p = 0.118). When asked about the overall accuracy of each system, eleven participants considered CapTUI to be the more accurate of the two systems; this was attributed to the visual guide functionalities, in particular the angle indicators ("aids to angles were helpful in making accurate shapes", "accurate and faster to use given the angle/length indicators", "the onscreen angle calibration and tool recognition made it much easier to draw accurate shapes").

Several questions were asked specifically about CapTUI (using a 5 point Likert scale). All but one participant agreed that learning to use the tangibles was easy. Five participants disagreed that the recognition of the tangibles was good. This aspect appears to be CapTUI's primary weakness and will be discussed further in the next section. The majority of participants agreed that the angle visualizations were easy to understand. All participants agreed that the length visualization was easy to understand and that the length and angle visualizations were useful. Ten participants agreed that the visual guides helped them to draw more accurately.

When asked what the best thing about CapTUI was, eleven participants answered that it was the visual guides ("angle and length indicators extremely useful", "It is easy to use and accurate. Also the visual guides are helpful and make the drawings easy"). When asked what the hardest thing about CapTUI was, nine participants mentioned either tangible detection or difficulty with fine adjustments ("detecting the tools was sometimes not accurate", "adjusting the drawing tools was a bit hard").

Overall comments from participants were: "CapTUI is much easier to use and more accurate because of the visual guides", "very good for people who need accurate measurements", "made it easy to draw accurate lines and shapes. Would probably be very useful if integrated into a CAD package", "If the intention is to produce accurate geometric drawings I think CapTUI is superior – however still a little tricky to use in terms of accuracy and fine tuning", "With a little improvement to the localization of the tangibles it will become a very useful tool for precise drawings".

In summary, CapTUI's main strength is in the ability to produce precise geometric constructions using the visual guides. Evidence of CapTUI's accuracy was shown by its significantly lower average angle error. Participants rated CapTUI significantly higher than Paint overall, enjoyed using CapTUI significantly more and found it to produce significantly more tidy drawings than Paint.

5 Discussion

There is very little work with tangibles for capacitive screens that demand a high level of interaction with precise tracking. In other projects the various blocks, dials and sliders are rotated and tracked, however the CapTUI drawing tools are manipulated in a more fine grained fashion, where a single degree of rotation is significant to the user. The most common feedback received from participants during our study was that they found the visual guides in CapTUI to be extremely useful for drawing precisely; however, making fine grained movements with the tangibles was difficult.

Designing tangibles for ease of use on a capacitive screen is a difficult problem. Of the user studies which feature tangibles on capacitive screens [26, 27], both acknowledge tangible design as a challenge; our tangibles are no different. Although our final tangibles were the best of all our designs, there are two remaining issues: consistent detection and accurate positioning.

For consistent detection there must be unbroken body-to-tangible contact and tangible-to-screen contact. A partial solution to this problem is to power the tangibles with their own battery, similar to the work in [9]. This would eliminate the need for the body-to-tangible contact. Maintaining the tangible-to screen contact would require a stable tangible design to ensure all touch points are on the screen at all times. There are also hardware and operating system limitations to consider such as enabling simultaneous touch and stylus input. Although this has been demonstrated in recent research [34], the ability to use such technology is limited to very few devices. Most devices, including the tablet used in our study, are not able to detect touch and stylus simultaneously. This means that if the stylus is on or in range of the screen (as a stylus can still be detected when it is a short distance away from the screen) then the tangibles cannot be detected via touch. Participants were informed of this limitation; however it still caused some problems when drawing.

A powered tangible is unlikely to solve the issue of accurate positioning. The main limitation here is the hardware. Capacitive hardware is designed to detect a general area of touch from a finger rather than a precise point. Our tangible touch points mimic this area of touch, however in order to calculate an accurate position and orientation of a tangible what is needed is a more precise point of touch. Currently the area of the touch point that is detected is translated to a centre point, but there is no way to know which part of the tangible's touch point is detected as the centre.

To assist in precise drawing it is possible to make more affordances in the software, similar to freezing the angle visualization (Section 3.4). However there is a fine line between tuning affordances to be helpful rather than causing frustration as the users control of the interface with the tangibles lessens. If they are not tuned carefully users may find the tangibles to be unresponsive and hinder them in achieving their goals. For example, we tried using automatic snaps to the horizontal and vertical screen positions of the tangible outline if the tangible came 'close' to these positions. Informal tests showed this to cause frustration when users did not want the tangibles in these positions. For example when tracking the tangible to a new position, if it was horizontal or vertical on the way to its new location, the tangible would seem unresponsive for some time due to the automatic snapping. We also experimented with averaging across three touch point positions to find a single position; however this did not always result in accurate positioning and caused delay in the recognition. Other possibilities would be allowing the user to lock the tangible outline in place or have a degree of stickiness to the movement; however such functions take away from overall goal of controlling the interface with the tangible.

If the above recognition issues can be resolved we believe a tangible drawing system would have great potential. Compared to more complicated drawing programs, tangibles are familiar and intuitive to use, don't require training, don't require the user to look at the tangible markings, therefore providing more flexibility. Our study showed that the main contributor to providing such an environment is the use of visual guides; without these the interaction is cumbersome. They provide essential feedback, such as the tangible outline, and additional information such as length and angle indicators. Users found our visual guides to be helpful and easy to understand; confirming that they enhance the interaction for geometric drawing. In future work we plan to compare tangible drawing systems to pen and paper.

6 Conclusion

We have presented CapTUI, a tangible drawing application for geometric constructions on a capacitive touch screen. CapTUI is composed of tangible drawing tools, a recogniser to identify each tool, a drawing application and visual drawing guides to augment the drawing experience. Our user evaluation compared CapTUI to Paint, which used non-recognizable drawing tools. Our results show that participants prefer CapTUI significantly more overall. Participants especially found the visual drawing guides to be helpful to drawing more precisely. The main challenges for tangibles on capacitive screens remain in good tangible design and consistent recognition.

Acknowledgements. This project is funded by the Royal Society of New Zealand Rutherford Foundation.

References

1. Marzola, E.S.: Using Manipulatives in Math Instruction. Journal of Reading, Writing, and Learning Disabilities International 3, 9–20 (1987)
2. Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K., Silverman, B.: Digital manipulatives: new toys to think with. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 281–287. ACM Press/Addison-Wesley Publishing Co. (1998)

3. Uttal, D.H., Scudder, K.V., DeLoache, J.S.: Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. Journal of Applied Developmental Psychology 18, 37–54 (1997)
4. Tuddenham, P., Kirk, D., Izadi, S.: Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In: Proceedings of the International Conference on Human Factors in Computing Systems, pp. 2223–2232. ACM, New York (2010)
5. Ullmer, B., Ishii, H.: The metaDESK: models and prototypes for tangible user interfaces. In: Proceedings of User Interface Software and Technology, pp. 223–232. ACM, New York (1997)
6. Baudisch, P., Becker, T., Rudeck, F.: Lumino: tangible blocks for tabletop computers based on glass fiber bundles. In: Proceedings of the International Conference on Human Factors in Computing Systems, pp. 1165–1174 (2010)
7. Wellner, P.: The DigitalDesk calculator: tangible manipulation on a desk top display. In: Proceedings of User Interface Software and Technology, pp. 27–33. ACM, New York (1991)
8. Fitzmaurice, G.W., Ishii, H., Buxton, W.: Bricks: Laying the Foundations for Graspable User Interfaces. In: Proceedings of CHI 1995, pp. 442–449. ACM Press/Addison-Wesley Publishing Co., New York (1995)
9. Yu, N.-H., Chan, L.-W., Lau, S.Y., Tsai, S.-S., Hsiao, I.-C., Tsai, D.-J., Hsiao, F.-I., Cheng, L.-P., Chen, M., Huang, P., Hung, Y.-P.: TUIC: enabling tangible interaction on capacitive multi-touch displays. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 2995–3004. ACM, New York (2011)
10. Inhelder, B., Piaget, J.: The early growth of logic in the child. Harper & Row, New York (1964)
11. Piaget, J.: Play, dreams, and imitation in childhood. Norton, New York (1962)
12. Fails, J.A., Druin, A., Guha, M.L., Chipman, G., Simms, S., Churaman, W.: Child's play: a comparison of desktop and physical interactive environments. In: Proceedings of Interaction Design and Children, pp. 48–55. ACM, New York (2005)
13. Sitdhisanguan, K., Dechaboon, A., Chotikakamthorn, N., Out, P.: Comparative study of WIMP and tangible user interfaces in training shape matching skill for autistic children. In: TENCON 2007, pp. 1–4 (2007)
14. Chipman, G., Fails, J.A., Druin, A., Guha, M.L.: Paper vs. tablet computers: a comparative study using Tangible Flags. In: Proceedings of the International Conference on Interaction Design and Children, pp. 29–36. ACM, New York (2011)
15. Olson, I.C., Leong, Z.A., Wilensky, U., Horn, M.S.: It's just a toolbar!: using tangibles to help children manage conflict around a multi-touch tabletop. In: Proceedings of Tangible, Embedded, and Embodied Interaction, pp. 29–36. ACM, New York (2010)
16. Schneider, B., Jermann, P., Zufferey, G., Dillenbourg, P.: Benefits of a Tangible Interface for Collaborative Learning and Interaction. IEEE Trans. Learn. Technol. 4(3), 222–232 (2011)
17. Druin, A., Montemayor, J., Hendler, J., McAlister, B., Boltman, A., Fiterman, E., Plaisant, A., Kruskal, A., Olsen, H., Revett, I., Schwenn, T.P., Sumida, L., Wagner, R.: Designing PETS: A personal electronic teller of stories. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 326–329. ACM, New York (1999)
18. Karime, A., Hossain, M.A., Gueaieb, W., Saddik, A.E.: Magic stick: a tangible interface for the edutainment of young children. In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 1338–1341. IEEE Press, Piscataway (2009)

19. Xie, L., Antle, A.N., Motamedi, N.: Are tangibles more fun?: comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In: Proceedings of Tangible and Embedded Interaction, pp. 191–198. ACM, New York (2008)
20. Price, S., Sheridan, J.G., Pontual Falcão, T., Roussos, G.: Towards a Framework for Investigating Tangible Environments for Learning. International Journal of Arts and Technology Special Issue on Tangible and Embedded Interaction 1(3/4), 351–368 (2008)
21. Scarlatos, L.L.: Tangible Math. Special Issue on Computer Game-Based Learning. International Journal of Interactive Technology and Smart Education 3(4), 293–309 (2006)
22. Hohenwarter, M., Jones, K.: Ways of linking geometry and algebra: the case of Geogebra. Proceedings of the British Society for Research into Learning Mathematics 27(3), 126–131 (2007)
23. Straesser, R.: Cabri-géomètre: Does Dynamic Geometry Software (DGS) Change Geometry and its Teaching and Learning? International Journal of Computers for Mathematical Learning 6, 319–333 (2002)
24. Rekimoto, J.: SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 113–120. ACM (2002)
25. Chan, L., Müller, S., Roudaut, A., Baudisch, P.: CapStones and ZebraWidgets: sensing stacks of building blocks, dials and sliders on capacitive touch screens. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 2189–2192. ACM, New York (2012)
26. Kratz, S., Westermann, T., Rohs, M., Essl, G.: CapWidgets: tangible widgets versus multi-touch controls on mobile devices. In: Extended Abstracts on Human Factors in Computing Systems, pp. 1351–1356. ACM (2011)
27. Jansen, Y., Dragicevic, P., Fekete, J.-D.: Tangible remote controllers for wall-size displays. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 2865–2874. ACM (2012)
28. Sutherland, I.E.: Sketchpad: A man-machine graphical communication system. In: Spring Joint Computer Conference: American Federation Information Processing Societies, pp. 329–346 (1963)
29. Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., Buxton, W.: Digital tape drawing. In: Proceedings of User Interface Software and Technology, pp. 161–169 (1999)
30. Couture, N., Riviere, G., Reuter, P.: GeoTUI: a tangible user interface for geoscience. In: Proceedings of Tangible and Embedded Interaction, pp. 89–96. ACM (2008)
31. Wiethoff, A., Schneider, H., Rohs, M., Butz, A., Greenberg, S.: Sketch-a-TUI: low cost prototyping of tangible interactions using cardboard and conductive ink. In: Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction, pp. 309–312. ACM (2012)
32. Hse, H., Newton, A.R.: Recognition and Beautification of Multi-Stroke Symbols in Digital Ink. In: AAAI Fall Symposium Series, pp. 78–84 (2004)
33. Blagojevic, R., Chen, X., Tan, R., Sheehan, R., Plimmer, B.: Using tangible drawing tools on a capacitive multi-touch display. In: Proc. BCS-HCI 2012, pp. 315–320. British Computer Society (2012)
34. Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H., Buxton, B.: Pen + touch = new tools. In: Proceedings of the ACM Symposium on User Interface Software and Technology, pp. 27–36 (2010)

Evocative Computing – Creating Meaningful Lasting Experiences in Connecting with the Past

Janet van der Linden[1] , Yvonne Rogers[2], Tim Coughlan[3], Anne Adams[1], Caroline Wilson[4], Pablo Haya[5], Estefanía Martín[6], and Trevor Collins[1]

[1] Open University, Milton Keynes, UK
{janet.vanderlinden,anne.adams,trevor.collins}@open.ac.uk
[2] University College London, London, UK
y.rogers@ucl.ac.uk
[3] University of Nottingham, Nottingham, UK
tim.coughlan@nottingham.ac.uk
[4] Friends of Mill Road Cemetery, Cambridge, UK
friendsofmillcemetery@gmail.com
[5] Universidad Autónoma de Madrid, Madrid, Spain
pablo.haya@uam.es
[6] Universidad Rey Juan Carlos, Madrid, Spain
estefania.martin@urjc.es

Abstract. We present an approach – evocative computing – that demonstrates how 'at hand' technologies can be 'picked up' and used by people to create meaningful and lasting experiences, through connecting and interacting with the past. The approach is instantiated here through a suite of interactive technologies configured for an indoor-outdoor setting that enables groups to explore, discover and research the history and background of a public cemetery. We report on a two-part study where different groups visited the cemetery and interacted with the digital tools and resources. During their activities serendipitous uses of the technology led to connections being made between personal memories and ongoing activities. Furthermore, these experiences were found to be long-lasting; a follow-up study, one year later, showed them to be highly memorable, and in some cases leading participants to take up new directions in their work. We discuss the value of evocative computing for enriching user experiences and engagement with heritage practices.

Keywords: pervasive computing, user experience, heritage practice, memories, evocative computing.

1 Introduction

The lives of people from different times and places are preserved through cultural heritage sites, and as stories about them are passed on through generations, they provide us with a sense of humankind's history. We may, for example, consider our own eating habits and food preparation methods when we visit the kitchens of an old castle, and see the layout, the pots and pans and the recipes that were used all those years

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 529–546, 2013.

ago. Ubiquitous computing technologies are increasingly being used to augment such artifacts to help us delve deeper when exploring historical sites and museums. Museums often provide interactive exhibitions, digital information points and mobile guides, making it possible for people to hear or read about relevant information relating to the artifacts that are close by. There are however some concerns about the introduction of such technological tools, not only about their design and usability, but also about how they impact on the visitor experience as a whole. The social qualities of visits to such places can be impaired when using isolating single user tools, and it remains a challenge as to how to create experiences that are engaging and that draw visitors in – so as to enhance the visitor experience and to enable people to relate this to their own personal histories [9].

A number of researchers have designed visitor experiences with the aim of creating more engaging, participative approaches for people to explore history and the arts. These include setting up a game-playing environment in a museum [26]; providing tools where people can record their memories in outdoor settings so that they can be listened to by other visitors [21] and setting up special rooms and interactive objects for reflection and sharing opinions [9]. However, cultural heritage is not just found in museums but also in our physical environment and in the stories we tell around them; the buildings we live in, the monuments, installations and statues erected in memory of people who have died, and the objects we use as we go about our daily life.

Our research is concerned with taking this one step further: how can we design and use technologies to support the processes of reflecting and sharing of memories for historical sites, situated outdoors (such as cemeteries, roman ruins, castles)? In particular, we are interested in how a historical place can be brought alive, and histories revealed through discovering and learning more about the lives of the people who lived or are buried there.

User involvement can make the visit itself more engaging, but what happens after the visit? How can we go beyond creating environments where people can record their own perspectives, opinions and memories, to creating environments that enable people to make new associations between their own lives and the present visitor experience in meaningful and enriching ways? Can these both be felt in the moment as well as reflected upon later? If so, how can we design technologies to engender such moments, where new insights, emotional reactions, flashes of understanding and connections are made [3] – without knowing in advance what form they might take? Our approach: *evocative computing,* is an attempt to enable serendipitous enrichment and reflection in life: providing opportunities for personal memories to surface in ongoing activities that will result in them being forged and cemented in deep and long lasting ways.

To this end, we developed an assortment of distributed digital technologies - designed and configured for people to be able to readily pick up and put down, use momentarily or 'stumble upon' in the context of their activities. These were deployed in the setting of an old cemetery by a number of different community groups uncovering the stories of the people that lay buried there. Multiple entry points were provided [cf 17, 22] so that individuals, in the context of a group activity, could make connections between their personal memories and in-the-moment experiences. Through these, an assortment of digital images, records, maps, sounds were 'at hand', that could be accessed, annotated, added to, and searched for, at different times and in different places. The resources, devices or displays were not owned by or given responsibility by any one person but were meant to be shared and used casually, and to be used to delve deeper into the details of a story.

We describe the technology set-up that was initially piloted in a small cemetery, and then used in events at a large cemetery. Then, we present a user study of it being used by various community groups. A follow-up interview study, a year later, was conducted to determine how they had engaged with it and what they remembered from using it. The findings are discussed in terms of what people recall about their experiences, how they reflect upon them, and what aspects of the technology they particularly remember.

2 Background

The use of computing technologies in museums has largely supported the solitary visitor, rather than the museum visit as a social occasion with people discussing and sharing their thoughts as they go around the exhibits. A popular aid is the mobile guide; visitors are given small handheld devices that are listened to like a mobile phone, sometimes with little screens. Whilst providing additional information for the individual experience they do not lend themselves to sharing, in the way written labels and other exhibit displays, like posters do. The audio guides typically provide background information about a painting or other artefact that is being viewed. They tend to be curated and hence do not support multiple perspectives, including that of the visitor [16].

Social Interactions. The social aspect of visiting exhibitions has been explored in response to the realization that most visits to museums and other cultural places are in families or small groups [26]. Woodruffe [27] designed an audio guide that could be shared, through an 'eavesdropping' mechanism, meaning that people could both listen to information, but also have conversations with each other and overhear other people's conversations. Social interaction is also the focus of the Kurio museum guide system, in which visitors were assigned the role of time travelers as part of a game, working as a team to collect various forms of information from the museum environment in order to repair their time traveller's map [26]. Kurio, a hybrid system, comprised a set of tangible computing devices, a PDA, and a tabletop display. Participants enjoyed working collaboratively and also enjoyed 'doing things' by pointing, gesturing, reading and listening with the various devices. Mcloughlin [21] developed a system to augment an open-air folk museum which aimed to recreate life from late 19th century through a series of dwellings that people could walk up to and look inside. Visitors could meet the 'inhabitants' of these dwellings, who explained how they went about their household chores or carry out other jobs. An empirical study showed that when exploring the park, visitors tended to reminisce about their childhood and about memories of the past from parents and grandparents. The researchers, therefore, developed a system that centered around this theme of memories: visitors were able to listen to memories of historical figures and record their own in response, and through the collection of tangible tokens could listen to memories that had been left behind by other visitors [21]. Another angle on interaction between visitors was developed by Ciolfi in

'Retracing the Past' [9]. Here, a deliberate effort was made to keep upfront information about the objects to a minimum, in order to encourage discovery and exploration by visitors. As part of this exhibition, both a 'Study Room' and a 'Room of Opinion' were set up; visitors engaged through a number of interactive objects, including an 'Interactive Desk', to study the provenance of the objects, and an 'Interactive Radio' to listen to experiences of other visitors and record their own. All these approaches involve the active engagement of visitors, where they have the possibility to share memories, conversations, personal thoughts and ideas. This was found to lead to high levels of engagement with the artefacts on display.

The above are examples of technologies that have been specifically embedded in a cultural setting to encourage reflection and exploration, and involve artefacts or exhibits that are in some way special but not normally part of the visitor's day-to-day life. Here, we are concerned with how technology can be set up to encourage experiencing history in evocative ways that will be long lasting.

Lasting Memories. Turkle [25] writes about the process of reflecting with objects that are very much part of our normal life, but that are somehow 'evocative' to people. Such evocative objects often relate to 'thresholds', both widely experienced historical events (e.g., war), and important transitions and developments in our own lives. They often command a sense of provenance, being evocative because they are understood in relation to particular places and events. Through these characteristics, our experience of an evocative object often subverts the sense of distance between us, and another time and place. This creates further emotional and reflective responses that are important for personal development.

Another way of characterizing the relationship between the user experience and technology is in terms of enchantment. This refers to "a sense of something not yet understood in a way that leaves us feeling disrupted yet alive, attentive, and curious" [8]. Bennet [6] suggests it refers to how people get 'caught up and carried away', heightening their perception and attention, but often through a process of disorientation. McCarthy et al. [20] propose that new technologies can be designed to enchant, and enable people "to wonder and to the wonder of life". Anecdotal examples include encountering the Apple G4 Powerbook for the first time. They argue that the more depth there is when making discoveries the longer the enchantment will last. Ross et al. [23] discuss how enchantment can be viewed in terms of meaningful mediation and the tensions surrounding them; for example, wearing an iPod changes the way we experience the world and the way we interact with people.

Journeys and Discovery. The goal of our research is to design technology as an enabler to evoke experiences, triggering and cementing new associations with a place being visited. Ubiquitous computing offers much scope for facilitating and connecting past and present user experiences, through enabling people to 'dive in and step out' [c.f. 1], reflecting on their discoveries. Benford et al. [4] have developed the framework of trajectories to characterize user experiences as journeys through hybrid structures, interrupted by transitions, and in which the interactivity and collaboration are orchestrated over space and time, involving multiple roles and interfaces [4, 5]. The orchestration of what the players have to do is pre-planned to instill certain forms of pleasure, learning and surprise. A question this raises is how much of the experience should be engineered for and how much should be left as ambiguous or uncertain. Should the user experience be scripted to a high level to ensure that participants are guaranteed a certain kind of experience [5] or should it be left more open-ended, so that surprising and serendipitous experiences can result [7]? The benefit of the former is that participants are reassured and know what they have to do, where to go and the objectives of the game. However, they may not discover much for or about themselves that makes them marvel or enables them to make deep and long-lasting connections. To enable people to take the initiative and discover new connections requires thinking about how to design for chance and serendipity [12]. This, in turn, involves more than simply designing for fortunate accidents to happen. For example, Andre et al. [2] argue that it is important to design not just for the discovery of new information but also the insights drawn from those discoveries. These can be gleaned through discoveries that are not only serendipitous but also through the movement between different types of experiences, some of which are systematically reasoned about and others driven by curiosity [24].

Evocative Computing. The focus of our research is how to design for more open-ended experiences, encouraging people to discover and make connections, themselves. Although they may be more disorienting and confusing than scripted ones, we argue they can lead to deeper and more memorable experiences. In particular, we are interested in enabling people to make connections with their own personal life, emotions, memories and thoughts while exploring cultural heritage sites as places of the past. We argue that such connections can lead to people gaining deeper insights into the historical context of the site in relation to their own lives, and in doing so making the experience of exploring the historical site a memorable one. Our approach, evocative computing, is intended to provide opportunities to support the creation of associations that can leave a lingering impression on an individual.

3 The Setting

The setting for our main study was an old Victorian cemetery – a cultural heritage site with a rich social history [11, 13]. When walking through the cemetery we may read the inscriptions on the headstones, wondering who the person was, how old were they when they died, what they died from, what kind of life they lived and who they were related to. When we come across a headstone that shows that someone died young, it can stop us in our tracks as we try to imagine what was the cause. To expand upon our notion of evocative computing, we ask: How can we move beyond the inscriptions to allow visitors to delve deeper [8, 19], not only to make connections about the lives of the people buried there but also to what is present around them and to their own personal lives?

The cemetery (see Figure 1) is situated in a small city in the UK. There are around 20,000 burials in the cemetery, most in unmarked graves, with around 3,500 marked by headstones. Some of the headstones are very old, overgrown, and covered in lichen. The people buried there were from a variety of backgrounds and part of the pleasure of visiting the cemetery is to imagine and discover more about them.

Fig. 1. Headstones in the cemetery

3.1 The Evocative Technology Set-Up

We assembled a number of interactive displays and mobile devices that were distributed indoors and outdoors for people to look at and interact with when in different locations. The technology was adapted from a previous project [10] for the specific context and challenges of the cemetery. The idea was to provide multiple perspectives when either in the cemetery or an adjoining building, in order to encourage diving in and stepping out [1], and in so doing, enabling serendipitous associations. The set-up comprised smartphones, iPads, video links and a shared multi-touch surface. Interlinked software apps and access to digital databases were developed to run across them. A variety of lightweight interactions were made possible that only required minimal learning. A photograph could be taken, then annotated with some text, and linked with a grave on a digital map. It was considered important that interactions required minimal learning, to enable the participants to feel comfortable using them, with only a quick introduction. The iPads and smartphones were intended to be used outside – to take photos of headstones and other things of interest (see Figure 2).

A room in an adjacent building to the cemetery was also set up as an indoor space for stepping back and reflecting on what had been seen in the cemetery. The technology comprised a Surface Tabletop computer mirrored onto the wall, laptops and a screen displaying a live stream from the site. A phone was provided for discussions with those outside (see Figure 3).

The photos that were taken by people outside could be uploaded in real time to appear on an interactive bird's eye view map of the cemetery displayed on the tabletop (see Figure 4). When indoors, participants were able to see photos being added onto the tabletop view of the graves – initially appearing in the form of coloured dots. Touching a dot resulted in bringing up the image taken there, and any information connected by the participants. The smartphones also allowed for texting and phone calls to be made between those inside and those outside in order to exchange information and generally plan the activity between group members. Live video images from a camera outside were shown on a wall display indoors, providing those indoors with a live roaming feed for allowing them to see what those outdoors were doing.

Fig. 2. Using mobile devices to look up database records and to photograph the graves

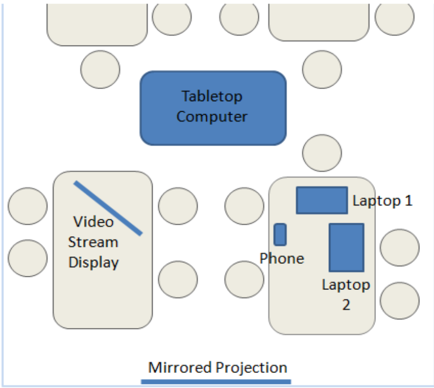


Fig. 3. Schematic diagram of indoor technology set-up

A shared content management system was created, based on the Drupal platform [14], which could be accessed from a wide range of devices to allow data to be up loaded and referred to from multiple locations. Accounts of the grave inscriptions and other parish records or copies of old newspaper clippings could be accessed. Other publicly available sources of information online, for example the digital records from the Commonwealth War Graves Commission, could also be accessed. The photos taken outdoors and the digital records were geo-referenced and mapped within the visualizations for the different devices. The GPS locations of the groups outdoors were automatically sent to a web server once every minute by the mobile phones. This location data was then displayed on the tabletop as part of the visualization. This enabled those indoors to see where the groups were outside as they moved around the cemetery. They could also see each photo they had collected, where they had collected it and any information each team had attached to that geo-referenced image.

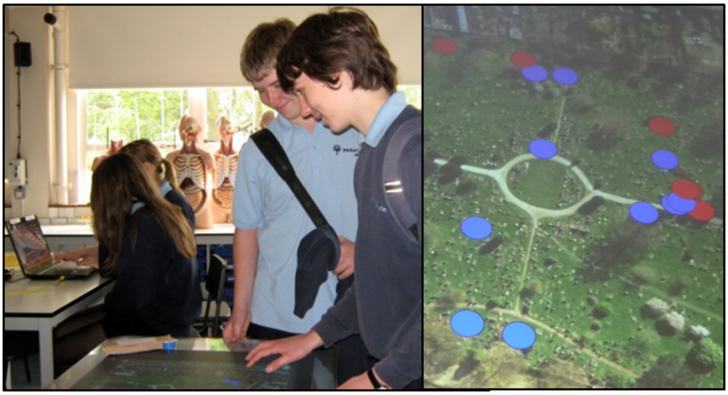


Fig. 4. Two pairs of students collaborating around the tabletop and laptops (left) and bird's eye view of the cemetery on the tabletop with dots representing the graves (right)

3.2 The Design Process

Several visits to the cemetery were made with numerous meetings held with representatives from groups who are interested in and/or use the cemetery. One group of local volunteers was the Friends of the Cemetery who work together with the Parochial Burial Grounds Management Committee and the City's Council to protect and enhance the cemetery for the public benefit. They view the cemetery as a special type of green space – a place of remembrance, spirituality, history and wildlife.

An initial prototype was trialled in a smaller graveyard to see how easy it was to use and switch between the various tools and displays. We invited the Friends of the Cemetery to trial the prototype set-up. Small groups of 2-3 participants went outside, took photos of interesting sightings and added notes. Each group then came inside and were shown the data they had collected, overlaid on the bird's eye view of the graveyard. A strong 'wow' reaction was observed; as groups experienced, for the first time, the transition from taking photos outside to seeing them linked with where they had taken them on a large-scale map. They were very moved by seeing their personal photos now as digital dots of where they had been (Figure 4). Clicking on them elicited spontaneous conversations about what they had seen, imagined and done there.

4 User Study

The user study was run in two parts. The first part took place over a period of 4 weeks during which four sessions were held, each with 12 to 22 participants visiting the cemetery. The second part took place a year later, when we revisited participants and interviewed them about their experiences and memories. Leaving aside the group who focused on wild-life, we report here on how two groups that already made visits to the cemetery for different reasons, had their visits enhanced through the technology:

(i) School children A group of 22 children, aged 14 to 15, who were studying drama, took part. Their drama teacher had used the cemetery before in her courses as a source of inspiration for writing and performing plays about people, although not yet with this particular group of children.

(ii) Local volunteers A group of 20 adults, who had a strong interest in local history, and, in particular, how stories involving the people who are buried in the cemetery inform us of the past took part. They had been working for some time on ways to digitally map the graves and on collecting materials.

Activity Design. We considered it important to let the teacher and the person running the local volunteer group decide on the focus of their explorations, in order to give them autonomy over what they wanted to engage with in the setting so as to fit in with their own interests and activities [15]. Our role was to provide technology to evoke and connect with personal experiences and help people explore and discover for themselves. The teacher chose to develop her class around the cemetery where the pupils had to create a play about the experience of soldiers in WWI and to contrast this with more recent wars, such as the Gulf War. The cemetery has a number of WWI graves of soldiers who died serving during this war, which are located throughout the cemetery.

The local volunteers wanted to identify graves in a particular part of the cemetery. Some twenty years earlier another history group had compiled lists with details of the inscriptions on each grave. This information had been transformed into database records, but no photographs were included nor was it clear where precisely each grave was located. They wanted now to determine the location of each grave and mark them on a large aerial photograph of the cemetery as part of their local history project.

The Sessions. At the beginning of each session an initial demonstration of the devices was given. The large groups were then broken up into smaller groups of 3-4 and left to work under the guidance of their leader/teacher. Participants were encouraged to make use of both the indoor room and the cemetery, which were only a few minutes walk from each other, enabling the smaller groups to visit both in a staggered way. Each session lasted between 1.5 to two hours.

The children began by looking at the headstones to see who might have died during WWI, and photographing them. One group remained indoors, researching background information on families, regiments and war locations. Throughout the afternoon groups around to enable all to switch between being indoors and outdoors. The volunteers in the cemetery started by photographing graves and then used the iPads to look up records in an attempt to link them together. They also interacted with the map on the tabletop indoors and guided the others outdoors on missing information.

Fig. 5. School children showing their photos and describing their discoveries to a researcher in the cemetery

5 Methodology

We used a mix of methods in order to study how the participants made use of the devices, how they approached their activities and what it was they talked about to each other. All the interactions with the technology were recorded using still photographs, video and notes. A team of researchers roamed around the cemetery observing the groups, and chatting to them informally (see Figure 5). Whenever groups moved from the outdoor to the indoor area, or vice versa, they were accompanied by one of the researchers, during which they would talk informally about what they had been doing and found. At the end of each session participants were asked to share their thoughts and impressions through a feedback form. We also held a group discussion with the teacher and children to talk about their findings and experiences.

A year later we went back to talk to the children, the teacher and the volunteers who were available. We asked them about their experiences of walking around the cemetery, using the technologies, conversations they had and whether anything had followed on from it. We spent a day at the school where the school children had now completed their drama coursework. We read the monologues they had written and also met with the next cohort of children who were just starting out on their drama course and listened to how they approached the task.

The data collected was analyzed thematically. These were refined and verified by a comparison of themes produced across the research team. The focus of our analysis was on how the participants made discoveries, *in situ*, in particular, how they used the devices and displays to make connections between their own lives and the histories of the lives they were discovering.

6 Findings

Our analysis revealed observable moments where participants discussed their 'personal evocative memories' with each other. We also observed 'digital situated awareness' of where they were and what they had just done. Below we present our findings in terms of a set of themes related to evocation and memory: (i) digital delight of being there, (ii) revealing more, (iii) serendipitous discovery, (iv) triggering and extending the personal, (v) personal memories and making connections, and (vi) seeing things differently.

Digital Delight of Being There. Similar to the finding of the pilot study discussed earlier, key moments that occurred within both the school children's and the volunteers' sessions were, when after being outside taking photos, they came inside and were struck by seeing a representation of their own collected data on the big display on the wall or the tabletop. Many experienced a sense of delight on seeing the 'dots grow'; representing their photographs and where they were positioned on the map of the cemetery in real time at the place they had taken them.

This indoor-outdoor set-up provided them with a new way of seeing their photos transformed into a collection of shared representations on a map, showing their discoveries. They looked at the map as more dots appeared, seeing what photos the others outside were adding. In doing so, they vicariously understood the process of being out there taking a photo while seeing it become a collective data point. This shared awareness gave them a powerful sense of connection - being present and stepping back, enabling them both to remember the details surrounding taking the photo and to reflect upon the bigger picture. Evidence of their joy and wonderment at this was made through their comments to each other and the researchers standing by. For example,

> *'Hey look at that! That's the one we took, remember? When we were near that bench. …'*

and from the organizer of the volunteers:

> *'this is amazing … there's been so much activity this afternoon, we've made so much progress, so many photos taken, and you can see all the activity on the map…..'*

The groups were also able to see at a glance how many photos they had taken relative to the other groups – which was particularly fascinating to the school children as they were seen discussing this with their teacher. It also led them to get in touch with the other groups outside, texting them to say how well each group was doing and what photos they were taking.

Revealing More. One of the volunteer groups tried to fathom out which inscription entry in the database matched with which grave, and where each grave was located. Another group was seen spending a lot of time around a particular grave trying to decipher its inscriptions. The stones were covered with a layer of lichen, a soft mossy type of fungus, making it particularly difficult to discern the writing. They were seen feeling the stone, with their fingers, to make sense of the engravings. Later on, the same group was seen using the tabletop indoors. They were looking at the photograph they had taken of the gravestone that was so difficult to decipher. By zooming in on the photograph, and enlarging it, they noticed that it was easier to decipher some of the lettering, exclaiming how the zooming action on the tabletop enabled 'magnification of the picture' that resulted in further conversations about the person buried there. This was a very pleasing moment for the group as the tabletop was able to support them in an activity that was causing them difficulty outdoors and showed a use of the tabletop that they had not anticipated. The moment was powerful, particularly since they had only just come from outside, having spent considerable amount of time around the gravestone. While outside, they had used both the iPad and the smartphone to try and identify the gravestone, but it wasn't until they were inside, that they managed to successfully do so.

Serendipitous Discovery. Whilst looking at the map of the cemetery, one of the participants accidentally pressed a dot on the map, which was associated with a different grave, near to the one they had just identified. This brought up the associated description of a famous local historian, John Seeley, who was buried there. Another volunteer (who was also a historian by profession) glanced at the description, read it and jumped up in amazement:

> '*Seeley… Is that really him??'*
>
> '*Seeley … he's a really really great man! Professor John Seeley … he's the big researcher into the British Empire.'*

He continued to explain how university buildings in the locality are named after him and regaled stories of Seeley's infamous dinner parties and what happened at them. He then looked at the tabletop to find out exactly where this grave was in relation to the others on the map. All of the group were drawn in by his anecdotes and stood around the tabletop discussing where the grave might be through studying the pattern of dots. It transpired that the grave was right next to where he had been exploring all day but hadn't noticed the gravestone until this moment:

> *'I'll have to go back and find him now. Well that's amazing – not quite like finding Shakespeare but… yes, it's on the same scale. Heavens, as it happens, I'm actually teaching about him right now.'*

He then left the room to go back to the cemetery and found the grave. His excitement at finding it was palpable and after spending 20 minutes out there returned to tell everyone about his discovery.

Triggering and Extending the Personal. A year later, we met up with him and asked if he had looked back on that afternoon's discovery. He replied emphatically that finding that grave had been a special moment for him. He had referred to it on several occasions during his lectures and at international conferences. He also reflected more widely on the value of local buildings and graveyards and their potential for teaching history to students through situated discovery:

> *'For many years I have had something brewing at the back of my mind, about trying to do something like this in teaching. Taking students out to places, local building, and letting them do research about people that lived and worked in the area. For me these ideas sit at the back of my mind for a while. I get an idea and it all gets put into a store, perhaps for many years, slowly taking shape. And then something like this happens …'*

Hence, finding the particular grave had triggered something greater in him, as he was now actively reorganizing his teaching schedule so that his students would have more opportunities to go out to visit the local cemetery. His new approach to teaching had been evoked through the powerful moment in the cemetery.

The dots in the visualization appeared to have had a strong impact in this experience. It was he who brought up the role of the digital map application and it was clear that he could still imagine the visualization in front of him. He mentioned how he had expected the dots on the map to present an anonymous image of the cemetery: expecting each dot to represent some facts about an unknown, local, rather insignificant person. However, to his surprise, one of the dots sprung up the details of a person who was, and still is, of great importance to him personally, someone who had been tremendously influential to him throughout his whole working life. Despite all he already knew about the person, the technology had supported a new, evocative experience and with lasting effects.

Personal Memories and Making Connections. As anticipated, the school children, when wandering around a part of the cemetery used the iPads to help them find out more about the war graves. Sometimes this led them to think back on their own family histories. For example, one group found two graves that matched with names on a list, photographed them and then uploaded them to the system. Another group indoors confirmed that they were on the right track. They then stumbled across a big family grave, with a long list of names and dates, including one for 1917. This grave was made of crumbly stone and the inscription revealed that the soldier died abroad in 1917. While photographing this grave, one boy began to talk of his grandfather and that he had served in the Balkans during WW2. This sharing of a personal memory triggered a further discussion of war amongst the group, and in particular the different countries that were involved in the two wars.

Throughout the day, there were several accounts of where the school children switched between everyday moments, personal memories and their drama project research. In addition, their experiences were embedded in the goings on around them (e.g., stroking a dog, chatting with homeless people sitting on a bench). The combination of these with their discoveries when switching between the cemetery and the set-up indoors made a big impression, creating an overall engaging and memorable 'time'. The situated experiences, arguably, increased the intensity and recollection of them as subsequent personal memories. Such memories were often emotive, heightened by the context of death and according to the teacher, a few weeks after the event, the visit to the cemetery was still being talked about frequently amongst the school children in the classroom, in the corridors and even at home.

When we met the children a year later, they were eager to tell us about the monologues they had written and about the characters they had chosen as a result of the visit to the cemetery. They were also reflective and mentioned how they had made connections between the war, the visit to the cemetery and their own families:

> *"I talked to members of my family about the people - about their experience of knowing and losing - a person in the war."*

> *"In the summer, when I was in Germany, I spent more time in a graveyard there. In the village graveyard, where both my grandparents and my great-grandparents are buried. I visited graves and looked at them. In Germany, because of the way they do graves, like, they have different rules. There wasn't any war people buried there. But that's to do with their traditions. But it was just, I know, it was just different."*

Exploring a cemetery during the vacation was not a regularly occurring activity for him – in fact he had never looked at these family graves before. Clearly it linked to the earlier experience of walking around the cemetery with his school friends the previous year. The personal memories of his grandfather *then* had helped him make connections between himself in the graveyard, with his friends, and the war and the countries that were involved in the war – but he was now extending this experience, by following it up with explorations of his own family.

Seeing Things Differently. Several times, throughout their explorations of the cemetery, the school children reflected upon what they noticed and looked up using the iPads and tabletop apps. For example, one group when switching indoors after exploring outdoors looked up background information of the regiments. During the group discussion one boy remarked:

> "*It is odd - that when you are in the cemetery, you only think of the person's death - where they died, when and how they died - and yet, when you are here (inside) and look into it more, you think of their life".*

Having the same information – a soldier's name – can be used through multiple representations: inscribed on a gravestone, listed as a digital record, or as part of a description of a regiment and its action during war. These multiple representations allowed participants to look through different lenses at the same person, often having a powerful evocative effect.

During our discussion with the school children a year later it was noticeable that they did not specifically refer back to the technology they had used (phones, iPads and interactive tabletop), suggesting it was largely transparent to their ongoing activities. The exception, however, was the dots visualization on the tabletop, which continued to stand out:

> *"What was so startling was that you go out to find out about a world event – a world war – which is a large international event, where countries are fighting. War is about big numbers, about stats. (…) - when you're in the cemetery you see the graves – well, you saw the dots on the map – you see the scale of it. It's a big scale."*

The teacher remembered how the activity greatly impacted on her drama course: *"These were remarkably reflective pieces that this group wrote. They were very thoughtful. They really went into the characters and I am really pleased with what they produced."* For her, like the historian who had changed his approach to teaching, the experience had led her to reflect on her teaching: *"It is also exactly how I think they should research (site specific), hands on, out there learning*" and "*Whereas before, I would have encouraged pupils to make lots of notes when visiting places, to help them start preparing for their narratives – I now tell them to bring their phones to school, so that they can make photos of the things that strike them as special or unusual, in order to be able to remember where they were."*

When meeting again with some of the local volunteers, they talked purposefully about their experiences of locating and identifying the graves in the cemetery. Underlying their concerns about data collecting and how to store data, there was a sense that the sessions in the cemetery had helped cement friendships across the group – they looked back on the sessions with great fondness, recalling moments, and people and who had said what when.

7 Discussion and Conclusion

Our study has shown how an assortment of technology can be designed and used in a historical setting by groups of visitors to discover much more about the lives and social history of people buried in a cemetery that they were able to relate to their own personal histories. Moreover, the experiences were found to be long lasting; through using the devices and displays distributed indoors and outdoors, the participants made a variety of connections between their own lives and the lives of the people they were exploring that one year later, they were still talking about.

Our research raises the question of how do memories created through the evocative computing approach differ from those arising from just visiting a cemetery or other place of interest/historical site, without the range of technologies we assembled, and perhaps only having a mobile phone to hand. Clearly, it is possible to connect personal memories with the surroundings of an ongoing activity unaided by technology and to take pictures that can be looked at later. What role did the technology play in making the experiences so engaging and memorable? Below, we discuss possible ways our evocative computing approach enriched the user experience.

(i) Multiple and Memorable Representations A technology that played a central role in the linking of discoveries and personal experiences was the tabletop visualization, where the dots appeared in real-time representing pictures taken and information recorded. The participants were able to see the dots as something other than places on a map where photos were taken; more akin to a synecdoche between their discoveries, activities and memories: when, where and what the photos they had taken meant to them. The dots were the most remembered part of the technology set-up when talking about the experience a year later. They acted like a memory cue, similar to Mancini et al's [18] idea of a memory phrase, which enables people to go back to the memory of particular events and retrieve salient and emotive aspects of their experience in detail. Whereas Mancini et al deliberately asked people to make up a memory phrase in their study, asking people to remember what they were doing at various times of the day for a project on mobile privacy, the 'line of dots' emerged serendipitously for our participants. For example, for one boy, the dots led him to marvel that worldwide events were happening on his doorstep: "you think of WW2 and you know it's big stats, but then it is happening in my locality, that struck me – it's here". For a volunteer, it was the epiphany of the unexpected: he had expected the dots to represent local people, with local issues and local stories – and instead discovered that the grave of a revered historian was on his doorstep.

(ii) Intertwining Place, Information and Narrative The findings from our study show how the volunteers began to understand how the role of technology is not just about placing documents and photos in databases and archives, but that it can also support their physical interactions with aspects of the cemetery, and each other.

(iii) Connecting Personally Meaningful, Local and Global Research in HCI on memories and cultural heritage, historical events and places has tended to be separate. The connections between these are rarely made, and often local places, close by and with the potential to be exceptionally evocative, are ignored thus giving little importance to people's autonomous engagement with cultural heritage in the context of their own lives and the specific settings of where they live [15]. Here, we demonstrated how strong links can be seen between them. For example, the school children realized how the cemetery could trigger personal memories that made war history more meaningful and closer to home.

Our findings suggest, more generally, that the evocative computing approach can lead to meaningful and lasting experiences. It provides a way of thinking about how to enable visits to digital heritage sites, such as castles or ruins, to be enhanced not just during them, but also to create long lasting memories in a person's life. Specifically, by having multiple connected technologies at hand, different entry points for serendipitous experiences are made possible. The kinds of 'wow' moments we recounted may not always arise – but that is the essence of designing for serendipity. You cannot design directly for but you can provide opportunities. As our study revealed, participants readily switched between these entry points, uncovering information as and when they wanted – thus revealing the hidden which they could make into new persistent connections. Evocative computing works as an approach to support open-ended and user-created trajectories across space and time – that is different from trajectories or tasks designed and orchestrated by others (e.g. teachers, curators, designers).

In sum, the technology can act as a kind of glue enabling people to make connections between different aspects of their experiences later on. This suggests that when designing user experiences we should consider a range of temporal aspects, not just the immediate experiences of using technologies but also what the long lasting impact will be of interacting with them. Hence, an evocative computing approach should consider providing conditions rather than pre-defining expected outcomes. Finally, technology has often been designed to enhance specific experiences taking place at specific moments. Often the joy, wonderment and pleasure of an event are assumed to occur if people simply interact with an experience. However, our research has shown the value of enriching a personal and emotive moment through technology that then makes this a lasting, meaningful experience.

Acknowledgements. We thank the Parochial Burial Grounds Management Committee for their permission to use the cemetery and Anglia Ruskin University for use of their premises. We also thank Sarah Tovell (Cambridge City Council) and Kay Blayney and pupils (Parkside Federation Academies) and the Friends of the Mill Road Cemetery for their enthusiastic engagement with the project. Many more people helped to make this project possible, including Margaret Widdess, Fabrizio Manco, Seán Lang, Victoria Smith, Simon Ruffle, Canan Blake and John Lea. The research was funded by the EPSRC OTIH project (EP/H022589/1). Estefanía Martín and Pablo A. Haya acknowledge the ASIES (TIN2010-17344) project, from the Spanish Ministry of Science and Innovation that partially funded their participation.

References

1. Ackermann, E.: Perspective-Taking and Object Construction: Two Keys to Learning. In: Kafai, Y., Resnick, M. (eds.) Constructionism in Practice: Designing, Thinking, and Learning in a Digital World, pp. 25–37. Routledge (1996)
2. André, P., Schraefel, M., Teevan, J., Dumais, S.T.: Discovery Is Never By Chance: Designing for Un-Serendipity. In: Proceedings of the Seventh ACM Conference on Creativity and Cognition, pp. 305–314. ACM (2009)
3. Barnes, M.: 'Magical' Moments in Mathematics: Insights into the Process of Coming to Know. For the Learning of Mathematics 20(1), 33–43 (2000)
4. Benford, S., Giannachi, G., Koleva, B., Rodden, T.: From Interaction to Trajectories: Designing Coherent Journeys through User Experiences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 709–718. ACM (2009)
5. Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., Row-Farr, J.: Can You See Me Now? ACM Trans. on Computer-Human Interaction 13(1), 100–133 (2006)
6. Bennett, J.: The Enchantment of Modern Life: Attachments, Crossings and Ethics. Princeton University Press (2001)
7. Chalmers, M., MacColl, I.: Seamful and Seamless Design in Ubiquitous Computing. Technical Report Equator-03-005 (2004)
8. Chonchúir, M., McCarthy, J.: The Enchanting Potential of Technology: a Dialogical Case Study of Enchantment and the Internet. J. of Personal Ubiquitous Computing 12(5), 401–409 (2008)
9. Ciolfi, L., Bannon, L.J., Fernström, M.: Including Visitor Contributions in Cultural Heritage Installations: Designing for Participation. Museum Management and Curatorship 23(4), 353–365 (2008)
10. Coughlan, T., Adams, A., Rogers, Y., Davies, S.: Enabling live dialogic and collaborative learning between field and indoor contexts. In: The 25th BCS Conference HCI, pp. 88–98 (2011)
11. Curşeu, P., Pop-Curşeu, I.: Alive after Death: An Exploratory cultural Artifact Analysis of the Merry Cemetery of Săpânța. Journal of Community & Applied Social Psychology 21, 371–387 (2011)
12. de Bruijn, O., Spence, R.: Serendipity within a ubiquitous computing environment: A case for opportunistic browsing. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 362–370. Springer, Heidelberg (2001)
13. Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J.D., Gandy, M.: Exploring Spatial Narratives and Mixed Reality Experiences in Oakland Cemetery. In: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 51–60. ACM (2005)
14. Drupal, `http://drupal.org/`
15. Giaccardi, E.: Reframing Heritage in a Participatory Culture. In: Giaccardi, E. (ed.) Heritage and Social Media: Understanding and Experiencing Heritage in a Participatory Culture. Routledge (2012)
16. The Handheld Museum, `http://museumsandmobiledevices.wordpress.com/`
17. Hornecker, E., Marshall, P., Rogers, Y.: From Entry to Access – How shareability comes about. In: Proc. of the 2007 Conference on Designing Pleasurable Products and Interfaces, pp. 328–342. ACM (2007)

18. Mancini, C., Keerthi, T., Rogers, Y., Price, B., Lukazs, J., Bandara, A., Joinson, A., Nuseibeh, B.: From spaces to places: Emerging contexts in mobile privacy. In: Proc. of the 11th International Conference on Ubiquitous Computing, pp. 1–10. ACM (2009)
19. McCarthy, J., Wright, P.: Technology as Experience. MIT Press, Cambridge (2004)
20. McCarthy, J., Wright, P., Wallace, J., Dearden, A.: The Experience of Enchantment in Human–Computer Interaction. J. for Personal Ubiquitous Computing 10, 369–378 (2006)
21. McLoughlin, M., Ciolfi, L.: Design Interventions for Open-Air Museums: Applying and Extending the Principles of "Assembly". In: Proc. of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 553–556. ACM (2011)
22. Rogers, Y., Lim, Y., Hazlewood, W., Marshall, P.: Equal Opportunities: Do shareable interfaces promote more group participation than single users displays? Human-Computer Interaction 24(2), 79–116 (2009)
23. Ross, P., Overbeeke, C., Wensveen, S., Hummels, C.: A designerly critique on enchantment. Personal Ubiquitous Computing 12(5), 359–371 (2008)
24. Tieben, R., Bekker, M.M., Schoute, B.A.M.: Curiosity and Interaction: Making people curious through interactive systems. In: Proc. BCS HCI, pp. 1–10 (2011)
25. Turkle, S.: What Makes an Object Evocative? In: Turkle, S. (ed.) Evocative Objects: Things We Think With. MIT Press (2007)
26. Wakkary, R., Hatala, M., Muise, K., Tanenbaum, K., Corness, G., Mohabbati, B., Budd, J.: Kurio: a Museum Guide for Families. In: Proc of the 3rd International Conference on Tangible and Embedded Interaction, pp. 215–222. ACM (2009)
27. Woodruff, A., Aoki, P.M., Grinter, R.E., Hurst, A., Szymanski, M.H., Thornton, J.D.: Eavesdropping on Electronic Guidebooks: Observing Learning Resources in Shared Listening Environments. In: Proc. 6th Int'l Conf. on Museums and the Web, pp. 21–30 (2002)

Systematic Integration of Solution Elements: How Does Digital Creativity Support Change Group Dynamics?

Florian Perteneder[1], Susann Hahnwald[2], Michael Haller[1], and Kurt Gaubinger[2]

[1] University of Applied Sciences Upper Austria, Media Interaction Lab, Hagenberg, Austria
mi-lab@fh-hagenberg.at
[2] University of Applied Sciences Upper Austria, School of Engineering/Enviromental/Sciences, Wels, Austria
{susann.hahnwald,kurt.gaubinger}@fh-wels.at

Abstract. In practice, most creativity techniques are still performed with traditional tools, such as pen and paper, whiteboards, and flipcharts. When transforming these techniques into a digital environment, the reduction of organizational overhead is the main goal to foster accessibility. Still, we do not know if overhead reduction fosters creativity or if it eliminates an important part of the creative process. To get a deeper understanding of these effects, we compare the performance of the creativity technique SIS (Systematic Integration of Solution Elements) in a traditional setting with a setup based on multiple interactive surfaces. By using a mix of diverse evaluation methods, we show how the use of a digital interactive creativity room can really foster creativity and produce better results.

Keywords: Creativity, Design, Creativity Techniques, Interactive Environment, Systematic Integration of Solution Elements, Collaboration.

1 Introduction

Creativity and innovation are predictors of success in our knowledge-based society. With the increasing availability of digital whiteboards, the development of tools to support creativity has become an emerging field [14, 15, 16, 18].

Since the ’70s, psychologists and practitioners have put a lot of effort in developing numerous methods for supporting creativity more effectively [13, 26, 27]. Although there are a lot of creativity techniques for versatile needs and tasks, only a small set of traditional tools and media, such as pen and paper, sticky notes, flipcharts, whiteboards, and pinboards, are used. Due to the limited possibilities to edit and copy content, especially complex creativity techniques often require a huge amount of organizational overhead (e.g. copying content, placing flipchart papers on pinboards). Digital implementations usually aim to limit this overhead and increase the ease of use [4, 20]. However, it is not clear if the efforts of digitizing the creative process lead to the desired effect of decreased task execution time or if they even have negative side effects on creativity and inspiration. New media in creative processes change the balance of power, involvement and satisfaction of participants, as well as

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 547–565, 2013.

the general dynamics of collaborative sessions. To provide a theoretical understanding for developing adequate and practical interaction processes and applications, it is necessary to carefully study the effects of digital solutions and their impact on group creativity.

In this paper, we analyze the impacts of a digital environment on group dynamics and creative outcome when performing a complex creativity method. Therefore, we performed an experiment to compare the traditional paper-based way [12, 34] of performing the creativity technique SIS (Systematic Integration of Solution Elements) [36] to its implementation in a digital, interactive environment (cf. Fig. 1). We use a combination of different evaluation approaches to get a deeper understanding of the emerging side effects that come with overhead reduction. Finally, we present the results of our analysis and discuss their implications on environments that are supposed to support collaborative creativity, and on creative tasks in general.

Fig. 1. Performing the SIS method in an interactive environment. Participants can discuss and present their ideas using digital paper and a large interactive whiteboard.

2 Related Work

2.1 Creativity Techniques

While Osborn's idea of verbal brainstorming [23] is widely known and used, many studies showed that this collaborative way of idea generation is not the best choice regarding quantity of ideas [24] due to negative social effects (production blocking, evaluation apprehension, and free riding) [8, 37]. Individual idea generation, also known as nominal brainstorming, is a way to overcome these issues and to increase the number of ideas produced [2, 32]. Nevertheless, verbal brainstorming is still preferred in many practical situations as it yields diverse perspectives when team members provide complementary skills and expertise [30].

Soon after the introduction of brainstorming, psychologists as well as practitioners developed new and improved creativity techniques [13, 26, 27, 36], focusing on different categories of problems. Most of these techniques include aspects of verbal brainstorming (collaborative), nominal brainstorming (individual), or both. The combination is considered as the best solution, especially when dealing with complex problems.

2.2 Supporting Creativity Techniques in Interactive Environments

A considerable amount of recent work has explored digital support for creativity, focusing on brainstorming or discussion support. Most systems or tools, however, focus on the implementation of one specific creativity technique. As we identified both individual and collaborative work as crucial parts of creative work, we pay special attention to this aspect when presenting existing work. In addition, we were interested in flexible and open approaches that support a variety of different creativity techniques.

Individual vs. Collaborative Work. There are multiple concepts that support both individual as well as collaborative aspects of the creative process [4, 14, 16, 18]. While some concepts provide shielded, private space for undisturbed, individual work [14, 16], other implementations provide private, non-shielded space for individual content creation [4, 18]. In general it seems as if the need for shielded space rises when tasks become more complex and require e.g. extensive sketching instead of simple input of single words.

Post Brainstorm [15] forgoes any dedicated features to support individual phases during the session. Since the system was designed as a substitution for traditional media used within the brainstorming sessions at IDEO, it offers a lot of flexibility regarding the import and arrangement of content, which is important for their creativity method.

Flexibility. Although there is some research about systems supporting very specific creative methods [20] and about providing certain improvements to the ideation process [1, 35], many systems try to provide at least some flexibility to the user. Warr and O'Neill [38] recommend providing users with flexible tools, such as free-hand drawing tools. TEAM STORM [16] aims not to impose structures by providing flexibility regarding private or public work. However, flexibility regarding a change of the underlying creativity technique is not intended.

Concepts based on the use of virtual sticky notes [14, 18] or paper strips [4] can provide a fast way of interaction but have issues with supporting concepts that require more extensive sketching.

3 Systematic Integration of Solution Elements

The creativity method *Systematic Integration of Solution Elements* (SIS) or *Successive Integration of Problem Elements* (*Successive Element Integration* – SEI) – was designed by Schlicksupp for extracting synergies from interdisciplinary teams [36]. SIS is a creativity technique to be used on problems that require rather complex solutions and offer only a restricted number of possible solutions, such as industrial or product design tasks. The basic idea of this method is to merge the benefits of individual (cf. nominal brainstorming) solutions to an integrated solution in collaborative work (cf. verbal brainstorming). Due to the integration of all individual solutions and the positive way of analyzing only their benefits (and not their drawbacks), SIS is a technique that leads to high identification of each participant with the final result.

3.1 The SIS Process

The SIS creativity technique is designed for 4-8 participants. The attendees develop individual solutions during the first working phase and integrate these ideas collaboratively during the moderated later phases. Usually the individual solutions as well as the integrated solutions consist of sketches with additional explanations.

After the initial problem framing, SIS defines a specific process that consists of three phases (cf. Fig. 2).

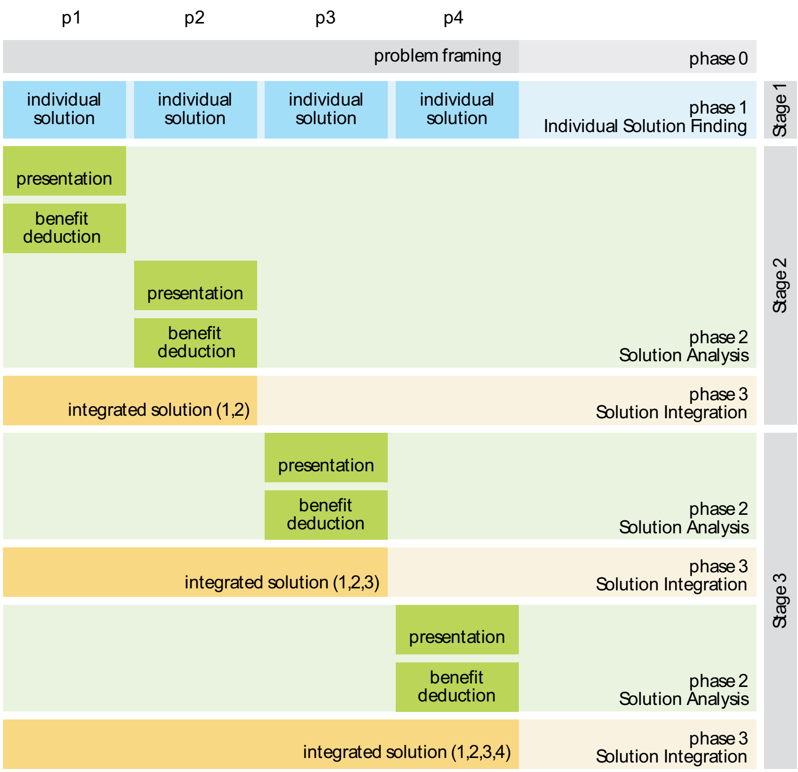


Fig. 2. Schematic representation of the different phases of the SIS process when used within a group of 4 participants. For our experiment we divided the entire process into three (artificial) stages (at the very right) to perform process-based measurements. Between the stages, we asked participants to complete questionnaires to measure their psychological state.

Phase 1 – Individual Solution Finding is characterized by an individual idea generation phase. Each participant comes up with a potential solution for the predefined problem on a sheet of paper. The goal is to get a wide variety of possible solutions by overcoming negative group dynamic effects.

Phase 2 – Solution Analysis consists of two parts. First the solution generated in phase 1 is presented by its creator. Second, beneficial ideas and advantages from the particular solution are extracted collaboratively and written down (on a flipchart or whiteboard) by the moderator.

Phase 3 – Solution Integration describes the process of combining the benefits of different solutions in one integrated solution. As this integration is performed collaboratively, the final solution does not only benefit from diverse ideas during generation, but also from different perspectives and from the expertise of a multidisciplinary team. This helps to consider the pros and cons of certain features in order to find the best solution for the final result.

As depicted in Fig. 2, phase 2 and 3 alternate according to the number of participants.

4 SIS Implementation for a Digital Environment

In a traditional setup, when performing the SIS method, all individual sketches are drawn by using pen and paper [12, 34]. Due to limited editing possibilities, sketches have to be redone and feature lists have to be copied manually when new solutions are integrated. This causes enormous organizational overhead. Our main goal is to provide a digital environment that minimizes this overhead.

The SIS method involves multiple types of interaction, which also appear in other creativity techniques. Similar to traditional, non-digital tools, the interactive environment as well as the applications were not designed to specifically support SIS but to provide a flexible solution for a variety of creative activities.

4.1 Environment

It is essential to adequately support individual and private content creation (phase 1) as well as collaborative, usually moderated work in a group (phase 2 and 3). For this reason, we decided on a twofold solution: We use digital paper (using Anoto pens[1]) to facilitate private content creation in combination with a large-scale interactive whiteboard that offers public space for discussion or collaborative work.

Digital Paper (phase 1). We chose digital paper for individual work for a number of reasons. Primarily, it captures handwriting on paper and enables displaying the content to the public. On the whiteboard, created content can be edited, copied and moved. Second, other than with digital solutions, such as tablets, people are very familiar with paper, so it requires almost no learning to use this technique. Third, paper

[1] www.anoto.com

provides adequate privacy during content generation. In addition, the possibility to take it to another place can be indirectly used for shielding during content creation. Finally, the approach is easily scalable for larger groups.

Interactive Whiteboard (phase 2 and 3). To support collaborative work, we use a large-scale multi-user interactive whiteboard based on the system presented in [17]. Due to its flexibility, it serves multiple purposes: It can be used for public note taking and to display prepared content, such as presentation slides. In addition, it can be used to display ideas, sketched on the above mentioned digital paper. As it is a digital device, it also offers benefits in terms of editing and space management capabilities compared to traditional media.

4.2 Applications

Digital paper is a suitable solution for phase 1 as it enables free-form sketching as well as handwritten content. In contrast, the requirements for the whiteboard software are diverse. First, participants have to be able to present their individual solutions to others. Second, there should be a way to emphasize and mark the benefits of a certain solution. In addition, the creation of integrated solution directly on the whiteboard should be supported. To meet these requirements two different applications were provided.

Paper Application. The first application allows users to display the content written on paper to the entire group (cf. Fig. 3, left).

It is very important to provide smooth and simple transitions of content between different devices and media. The bare possibility of providing ways to move and copy content in a digital system is no guarantee of actual overhead reduction compared to a traditional setting. For this reason, all pages that contain content are displayed per default in a small preview on the whiteboard without any additional interaction necessary. To ensure privacy during the individual content creation, the creator is able to trigger the public visibility by using a printed button directly on the paper sheet.

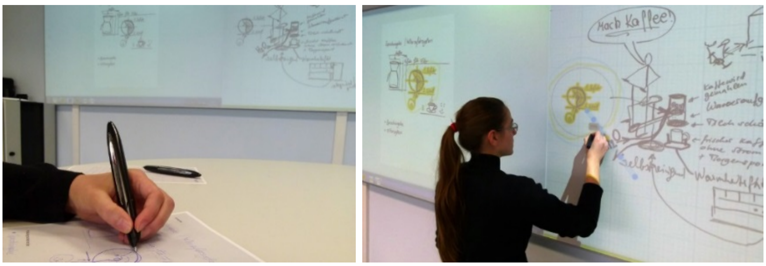


Fig. 3. Content written on digital paper is displayed on the whiteboard for presentation *(left)*. Content generated on paper is clipped into the Sketching Application for further usage *(right)*.

Individual and also groups of pages can be maximized for presentation or comparison (cf. Fig. 3, *left*). In addition, the application provides the possibility to select and copy content into a free-form sketching application for further editing (cf. Fig. 3, *right*).

Sketching Application. This application provides the possibility to create and edit content simultaneously for multiple users. Among others, editing includes erasing, selection, transformation and duplication of content. Moreover, the canvas is divided into pages. This way, the application provides effortless spatial navigation. During the SIS sessions, the page navigation was typically used to shift previous solutions and to generate new space (cf. Fig. 4).

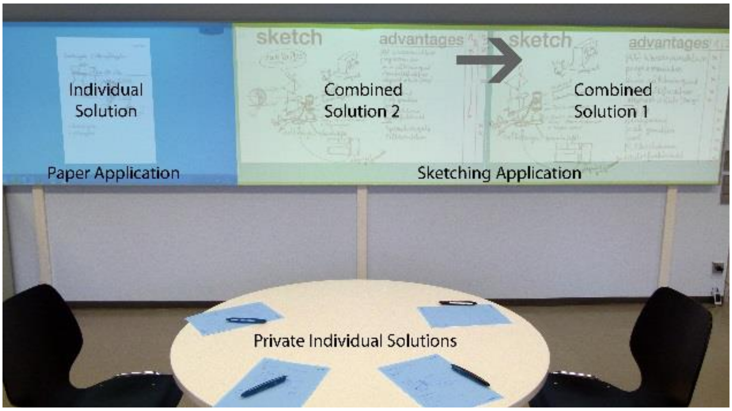


Fig. 4. Typical configuration of the digital environment. One third of the whiteboard is used for the paper application (blue) to display the previously created individual solutions; two thirds are used for the sketching application (green) to create combined solutions. In the course of the session the combined solutions are shifted to the right (symbolized by the arrow) to gain new space.

5 Evaluation Methodologies

There are two types of methods that can be used to evaluate the effectiveness and efficiency of creativity techniques. While outcome-based approaches focuses on the final results produced during the process [29], process-based approaches focuses on the ideation process itself.

5.1 Outcome-Based Approach

Outcome-based approaches focus on evaluating the final result of the ideation process. For assessing ideation quality, researchers usually are guided by the following four steps: First, unique ideas from an ideation session are identified; second, a quality score is assigned to each individual idea (usually done by domain experts, who understand and interpret the ideas [21]); third, by using one of the four approaches discussed below, a metric value is computed, which is used, forth, to make statistical comparisons between treatments of every session threshold [25].

There is a huge variety of criteria to give a score to each individual idea: The creativity of an idea is usually assessed through novelty - how unusual or unexpected an idea is compared to the other ideas [29] - and quality (feasibility or the readiness for implementation and the detail of description [8, 9]) [21]. Other criteria are utility [21], effectiveness [32], the value ideas could create [3], the importance of an idea within a specific context [33], and the magnitude of impact an idea might have [7]. Usually semantic differentials, such as Likert scales, or rubrics that evaluate one or more dimensions are used to measure those criteria [25].

5.2 Process-Based Approach

To fully evaluate the effectiveness and efficiency of a creativity technique and to understand the reasons for specific results, the overall process needs to be analyzed. By using process-based approaches, such as the concept of flow, occurrences of cognitive processes inherent to creative thought are evaluated. Other common evaluation methods are, e.g. video analysis and the 'think aloud' technique [28].

Concept of Flow. How does the user feel ideating? How does the user feel at the beginning, the middle and the end of the task? To answer these questions about how participants experience ideation, Csikszentmihalyi's concept of flow [10] can be used. It describes a complex psychological state that is characterized as a situation of perceiving an optimal and enjoyable experience by engaging in an activity with total involvement, concentration, and enjoyment. This results in intrinsic motivation and a sense of time distortion. Being in the flow contains: Having clear goals, focusing attention, losing self-consciousness, having an altered sense of time, enjoying the

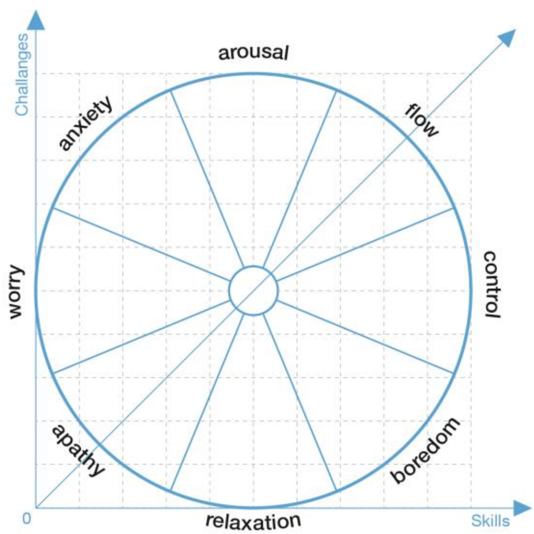


Fig. 5. The Flow Wheel adapted by Dorta [11]. If adequate skills (x-axis) match adequate challenges (y-axis), the user comes into the flow state that is considered most productive.

sense of control, and a merge of action and awareness. In summary, it describes the perfect balance between one's skills and the situations challenges, leading to an autotelic experience [6]. Therefore, one way to measure the psychological state of participants is to measure the balance between the skills and challenges, which is divided into eight possible dimensions: *apathy, worry, anxiety, arousal, flow, control, boredom,* and *relaxation* [22]. These parameters can be used as a barometer for reflecting the success of the ideation [10]. The Flow Wheel (cf. Fig. 5) asks participants to relate their amount of skills to the challenge during the recent activity by putting a single dot into one of the eight dimensions [11].

Open Coding Video Analysis. Video analysis is a popular approach to gather qualitative data of complex processes and to analyze them quantitatively.

The concrete nature of the findings is often uncertain before the analysis. In this case, open coding video analysis is a promising way to get a deeper understanding of the observed process [31, 19]. In this method, events in the videos are categorized by using codes that are defined during the coding process. This way, the approach is very flexible but also time consuming, as passages of the video have to be recoded as new codes emerge. Consolvo et al. [5] refer to a similar technique called LSA (Lag Sequential Analysis), as a valuable technique to generate quantitative and statistical data to observe ubicomp environments.

5.3 Our Evaluation Approach

Both categories, process-based as well as outcome-based evaluation methods, were used for evaluation. To get a broad insight into the ideation process, we chose to use a variety of process-based evaluation techniques. We used the Flow Wheel to find out about the participants' psychological state during the different stages of SIS. In addition, we used open coding video analyses to investigate overhead reduction and capture effects on group dynamics that might be a result of the changed environment. The evaluations were completed by an expert evaluation. Experts with different backgrounds (design, technic, and marketing) were asked to rate different aspects of the final results. Based on the above mentioned literature about outcome-based methodology we decided to use five criteria: *maturity, usability, consumer benefit, level of detail*, and *novelty*.

6 Experiment

To explore the digital implementation of the SIS method in use and to compare it with a traditional implementation, we conducted an experiment with students from the department of Innovation and Product Management. The main objective of this study was to get insight on how the use of an interactive, digital environment alters creative processes and group dynamics compared to a non-digital solution. Moreover, we wanted to find out how a digital environment fosters creativity and if it helps to get better ideas.

6.1 Reasons for Observing SIS

There are multiple reasons why we chose to use the creativity method SIS for our experiment. Most importantly it is a rather complex technique and involves a lot of different components such as individual and moderator guided collaborative work that also occur in other creativity techniques. In addition, the tools used in the traditional setting (flipcharts, pencil and paper) are rather generic and not tightly bound to this creativity method. Moreover, SIS does not limit the input to a specific type, like e.g. Method 635 [26], as the users can use sketches or handwriting to phrase ideas. Therefore, we believe that it is an appropriate technique to be studied as it is possible to draw conclusions for a variety of other creativity techniques as well.

6.2 Participants

32 of first-year students from the local university participated in the study. The participants were divided into eight groups of four, as four is a commonly used number to study various effects on group dynamics in small groups [2, 8]. There were seven females and 25 males between the age of 18 and 38 ($M = 24.2$, $SD = 5.1$). Participants in a group were either familiar or very familiar with each other. All participants made use of computers on a regular basis and had prior experience with pen-based interactive whiteboards. However, they were neither familiar with the used applications nor with the SIS technique and its digital implementation.

6.3 Moderators

As the moderator is a crucial factor in creativity methods alike SIS we decided to use different moderators in the experiment. This way we tried to balance the effects of single persons on the overall results. Overall five moderators (two female, three male) conducted one or two sessions each. The moderators stayed with the group when switching between the classical and the digital condition. They were not selected from the participant pool. The chosen moderators were more advanced students, who had considerable experience in presentation and workshop moderation. All moderators had prior training with the interactive environment and were familiar with SIS.

6.4 Apparatus

In the classic condition two flipcharts were used for public note taking and sketching (cf. Fig. 6. left). In addition, sheets of DIN-A3 paper were used by the participants during the Individual Solution Finding phase. Finally, the room was equipped with two pinboards to provide additional space to mount produced content.

The digital condition was conducted on a large (5.2 m × 1.17m) interactive whiteboard, driven by three Hitachi ultra-short-throw projectors with a total resolution of 3,072 × 768 (cf. Fig. 6. right). The whiteboard system was capable of handling simultaneous multi-user input through multiple Anoto pens (ADP-301). In addition, sheets of paper (DIN-A4) were provided to capture handwriting. For this purpose, multiple Anoto pens (ADP-201) equipped with a ball-pen refill were provided.

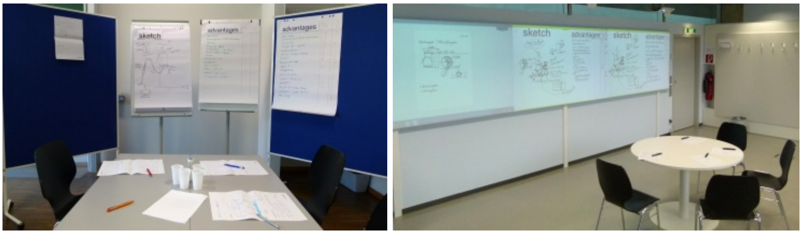


Fig. 6. The classical condition (two flipcharts, two pinboards, paper DIN-A3) *(left)*. The digital condition (one large interactive whiteboard, digital paper DIN-A4) *(right).*

6.5 Tasks

Due to the use of within-subject design, we had to define two different tasks of a similar degree of complexity, the participants were equally familiar with. Therefore, we asked the participants to design an easy-to-handle drip coffee maker in one and a novel vacuum cleaner in the other condition. We chose these tasks due to their medium complexity that enables participants to come up with different solutions but also with the realistic prospect to complete the tasks within one hour.

6.6 Procedure

Every group of four completed two tasks (~60 min each) in total, one using the classical and one using the digital environment. This resulted in a two hour session per group with a short break between the two conditions. Each group was assigned to a moderator that were told to keep track of the overall duration. The tasks and also the order of the conditions were counterbalanced. The SIS process and also the ways of analysis were the same for each method. Both tasks were unveiled to the participants one week before the study, so that they had some time to think about possible solutions.

To evaluate the ideation process using the Flow Wheel, the SIS process was divided into three stages (cf. Fig. 2). At the end of each stage, the participants were asked to complete a questionnaire to measure their mental and psychological state. At the end of both sessions, a questionnaire was handed out to the participants and moderation team to compare the experiences during the sessions. It included several questions focusing on problem description and understanding, visualization, further processing of generated ideas, documentation, and interaction. All sessions got videotaped for later analysis.

7 Results

In the final questionnaire, the digital condition was preferred by far to the classical setting, especially regarding overview, further processing, and fast visualization of the collected concepts. Also the expert evaluation confirms the advantage of the digital condition. To find out more about the causes of these positive results, we carefully analyzed the process using the above mentioned methodologies.

7.1 Outcome-Based Results Based on Expert Evaluation

All 16 integrated solutions were visually standardized after all sessions had been completed, to avoid mapping of solutions to a specific condition. In addition, a short explanation was added by the moderator. A heterogeneous group of experts (2 senior designers, 2 senior technicians, 2 marketing experts) rated these solutions independently on a 10-point scale by using the five above mentioned criteria (maturity, usability, consumer benefit, level of detail, and novelty). We can make several interim conclusions from the scores in Fig. 7.

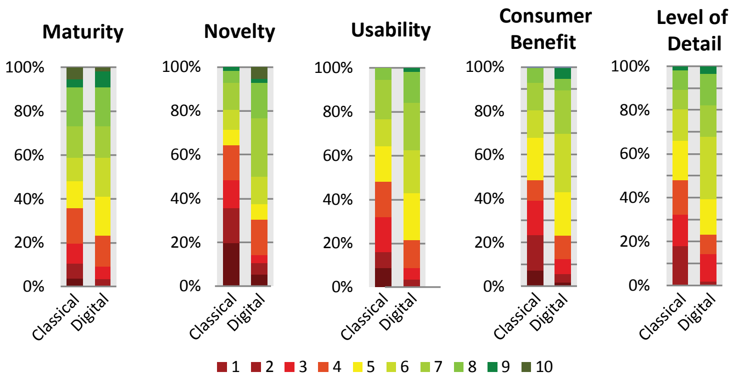


Fig. 7. Outcome-based expert evaluation of the final, visually standardized results. Each colored section represents one particular value on a 10-point scale (1=very bad, 10=outstanding).

A Wilcoxon Signed-Rank Test comparing both techniques showed no significant difference in the maturity (z = -0.771, p = .441). The aspect of novelty was rated better in the digital condition ($\tilde{X}_{\text{digital}} = 6.5$) than in the classical condition ($\tilde{X}_{\text{classic}} = 4$), z = -2.524, p = .012. The outcome of the digital condition was also considered superior by the experts in terms of usability ($\tilde{X}_{\text{classic}} = 5$ vs. $\tilde{X}_{\text{digital}} = 6$, z = -2.38, p = .017), consumer benefit ($\tilde{X}_{\text{classic}} = 5$ vs. $\tilde{X}_{\text{digital}} = 6$, z = -2.521, p = .012), and level of detail ($\tilde{X}_{\text{classic}} = 5$ vs. $\tilde{X}_{\text{digital}} = 6$, z = -1.68, p = .039). In the following section we are going to examine the reasons for these results.

7.2 Process-Based Results Based on the Flow Wheel

The diagrams in Fig. 8 show that the overall distribution of the psychological states is not entirely different for the two conditions. In both, we can see that most participants went from *flow* at the beginning (stage 1) towards the *control* segment at the end (stage 3). As the conditions did not really differ up to this point, this concord is not a surprise. However, it is interesting to see that, when using the classical SIS technique, already in the middle of the session (stage 2) quite a notable number of participants were in a state of boredom. This trend continued in stage 3, as the challenge level continued dropping. In contrast, in the digital condition, participants experienced a better balance between skills and challenges. Therefore they were kept longer and more consistently in the flow and control dimensions.

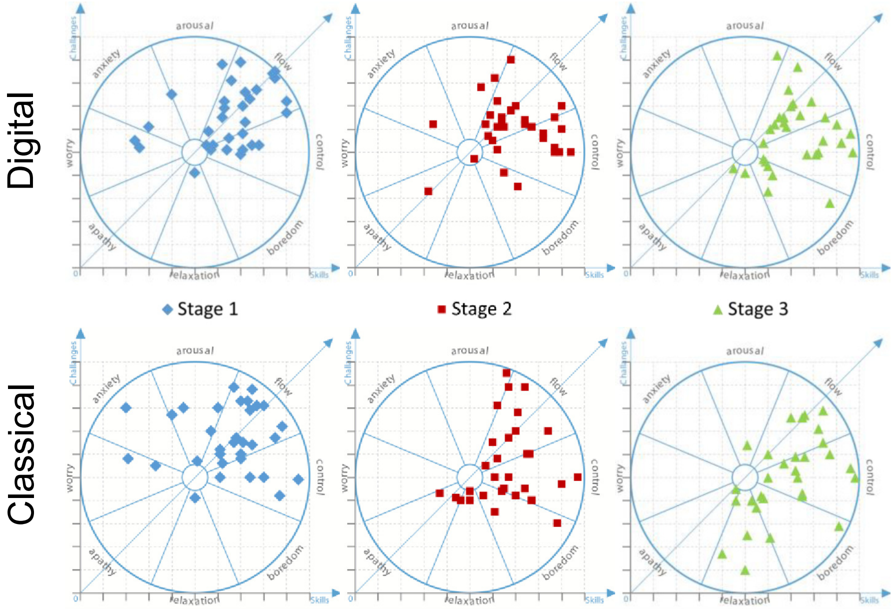


Fig. 8. Psychological state of the participants during the different stages of the session

Fig. 9 shows that participants were immediately able to start their task (stage 1) and to get into the state of flow. Stage 2, in contrast, was new to them. In the digital condition, the majority stayed in flow (15) and control (9) and only one continued to be anxious. Most of the participants stayed in flow (14) or control (8) also in stage 3, only eight participants felt either bored or relaxed. In contrast Fig. 9 shows a much higher drop-off of participants in the flow for the stages 2 and 3 in the classical condition.

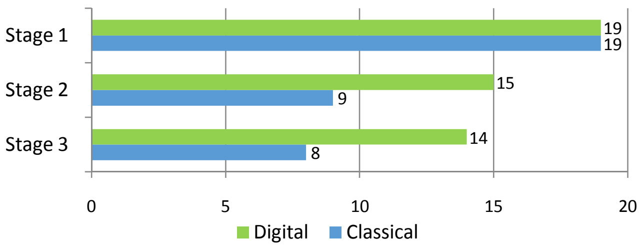


Fig. 9. Number of the participants in the flow state during the different stages

7.3 Process-Based Results Based on Video Analysis

The collected video data was subdivided into fragments of five seconds and analyzed by using open coding [31, 19]. For each fragment the corresponding phase of the SIS technique was determined. Moreover, the physical activity (e.g. sitting, standing, environment adaption) as well as the core activity (e.g. listening, talking, private / public writing, idle) was identified for each participant and the moderator.

Overhead Reduction. Regarding the reduction of overhead, the video analysis gave us the expected clear confirmation. The average amount of overhead decreased from 8:21 min (*SD* = 1:02 min) in the classical condition to 4:44 min (*SD* = 0:44 min) in the digital condition (cf. Fig. 10). The measured time includes activities, such as moving flipcharts and pinboards as well as pining sheets of paper and copying content in the classical condition. In the digital condition, navigation and moving, duplicating, and transforming content were activities considered as overhead.

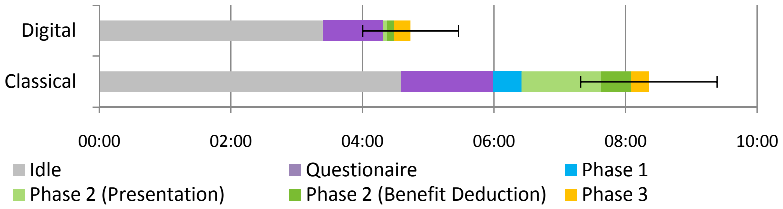


Fig. 10. Overhead Comparison: Idle refers to the time required for organizational tasks that block the creative process. Time needed to complete in-process questionnaires as well as phase 2 was also often used for organizational tasks.

Effects on the Overall Process. While the reduction of overall organizational overhead is quite high (43%) in the digital condition, the reduction of the *Idle* time - the timespan where the creative process has to be stopped to perform tasks considered as overhead - is much smaller (26%). In both conditions, the time to complete questionnaires was used for organization. In addition, in the classical condition phase 2 was heavily used to copy or duplicate content while the participants were busy with presenting. On multiple occasions in the classic conditions, the moderator asked the participants to continue with the discussion, while he/she was still performing organizational tasks.

Fig. 11 shows the temporal sequences observed from two sample groups (1, 2) under the digital (D) and classical (C) condition. Although the overall sequence of the phases was more or less the same, certain differences can be observed.

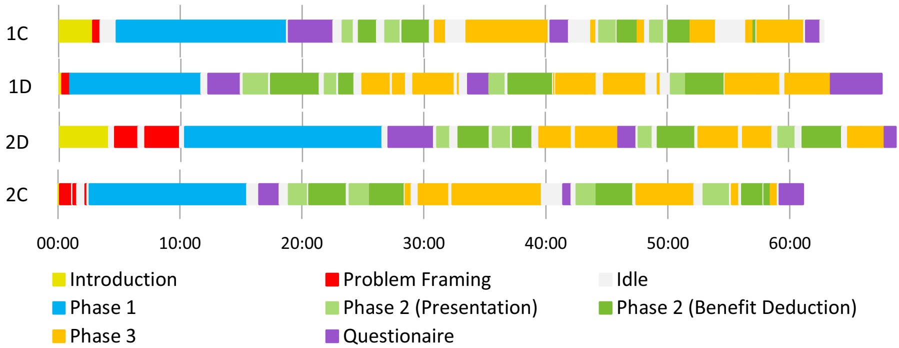


Fig. 11. Four samples of temporal sequences. 1D and 2D represent digital, 2C and 1C classical conditions.

1C and 2D were the first conditions conducted in the selected group. This is why the introduction was a lot longer. It is also noticeable that the illustrated digital conditions (1D, 2D) had a longer overall duration. Although this finding corresponds with the overall, average durations (classic: 54:06 min; digital: 56:47 min), there are too many different influences to identify a clear trend. For instance, the comparatively long problem framing phase in 2D was due to multiple questions of the participants. Furthermore the integration of the third solution in 1C and the forth solution in 2C were extremely short due to large benefit overlaps between individual and combined solutions.

Comparing the idle phases between the classical and digital conditions shows the average number of idle phases per session is exactly the same ($M = 18.25$; $SD_{digital} = 2.6$; $SD_{classical} = 0.5$). The phases, however, last 7.1% longer in the classical condition. This difference seems to be the result of a small number of extraordinary long idle phases that occur when the moderator is no longer able to handle the overhead simultaneously with running the session. These long idle phases appear in nearly all classical conditions (cf. Fig. 11). In contrast, the idle phases' durations are more equal in the digital conditions.

Effects on Participants' Performance. Since there are a lot of variables (e.g. personality, involvement) that affect participants' performance, the results in this section have to be handled with care. By all means, the observed sessions in both conditions were very dynamic. Especially when the moderator was busy with writing or organizing, subgroups that actively discussed issues or did something else, unrelated to the actual problem of the session, emerged very fast.

Fig. 12 unveils that the communication of the moderators in phase 2 (Benefit Deduction) and phase 3 increases in the digital condition. Most likely this effect occurred due to the reduced amount of overhead performed in these phases. Moreover, increased attention during phase 2 (Presentation) might also affected this result as the moderators had a better understanding of the solutions. Surprisingly the number of the statements of the participants in phase 2 (Benefit Deduction) rose simultaneously. We suppose this is at least partly due to the increased activity of the moderator.

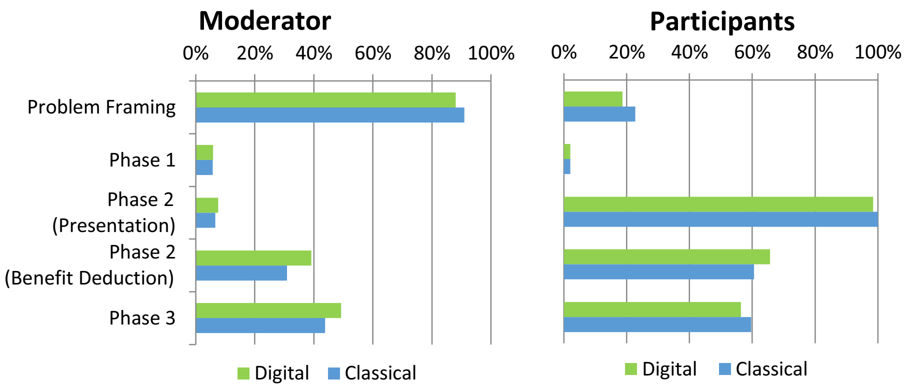


Fig. 12. Percentage of speech in the different phases and conditions. Due to the possibility of simultaneous comments, the combined time for moderators and participants may exceed 100%.

8 Discussion

In general, the analyses confirm that digital environments are not only capable of substituting the traditional paper-based environment, but also provide additional values reducing the overhead and can lead to better results. Especially the results of the expert evaluation are very convincing.

The results of the flow analysis (cf. Fig. 8 and Fig. 9) shows that participants stayed considerably longer in a state of flow during the digital condition. Excitement about the novelty of the digital approach might be one possible explanation. However, as all participants were already familiar with the interactive environment, the factor of novelty was eased to a great extent. Therefore, we assume that the ability to stay in the flow for a longer period was primarily caused by other reasons, such as changed interaction patterns between the group and the moderator.

In the classical condition, the moderators attempted to keep the creative process alive by doing organizational work during the actual creative phases. Investigating the amount of speech (cf. Fig. 12), we can see that this additional work leads to a decreased participation of the moderator especially in the second part of phase 2 and 3. Without a moderator leading the discussion, the results of most discussions between the participants ended quickly or the participants didn't stay focused. The experiment confirmed that a strong moderator is a very important part in the creativity technique. With decreasing his/her workload, which results in a higher participation and availability, the group productivity and process quality can be increased significantly.

In the light of this finding it is even more important to improve both the digital environment and the interaction concepts to decrease the required overhead even further. However, as currently most of the organizational tasks are accomplished by the moderator, also the adaption of the creativity techniques themselves should be considered. Digital environments enable smooth transitions between different private and public media and thus support more dynamic workflows. This could help the group to reduce their dependency from the moderator or to distribute the organizational overhead among all participants. This way, it would be possible to use the observed dynamic creation of subgroups and discussions more productively and sustainably.

9 Conclusion and Future Work

In this paper, we compared the performance of the creativity technique SIS in a traditional, paper-based setting to a digital environment with focusing on reducing the organizational overhead. Using a set of versatile methodologies, we did not only prove the advantage of a digital environment, but also investigate the negative effects of overhead when performing a creativity technique. Our findings show that the reduction of overhead does not necessarily result in a decrease of the task execution time. It rather leads to a higher participation of the moderator, higher likeliness for the participants to stay in a state of flow, and finally better and more advanced results.

In future work, we think it is important to intensify the research on interface and interaction techniques that support smooth transitions and effortless content management in digital environments to further reduce unnecessary overhead. Moreover, it will be interesting to develop new concepts of creativity techniques that enrich the dynamic possibilities of digital, interactive environments.

Acknowledgements. The presented research has received funding from the European Union's Seventh Framework Programme FP7/2007-2001 under grant agreement n° 318552. The experiment was realized in the "sprint>lab" that is supported by the Austrian Research Promotion Agency (FFG) within the research program COIN. The authors would like to thank the participants as well as the moderators, who helped us to realize the extensive study. We would also like to thank our colleagues from the Media Interaction Lab for their support with the implementation and their useful comments.

References

1. Bao, P., Gerber, E., Gergle, D., Hoffman, D.: Momentum: Getting and Staying on Topic during a Brainstorm. In: CHI 2010, pp. 1233–1236. ACM Press, New York (2010)
2. Baruah, J., Paulus, P.B.: Effects of training on idea generation in groups. Small Group Research 39(5), 523–541 (2008)
3. Boehm, B., Gruenbacher, P., Briggs, R.: Developing groupware for requirements negotiation: Lessons learned. IEEE Computer 18(3), 46–55 (2001)
4. Clayphan, A., Collins, A., Ackad, C., Kummerfeld, B., Kay, J.: Firestorm: A brainstorming application for collaborative group work at tabletops. Technical Report 678, University of Sydney (2011)
5. Consolvo, S., Arnstein, L., Franza, B.R.: User Study Techniques in the Design and Evaluation of a Ubicomp Environment. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 73–90. Springer, Heidelberg (2002)
6. Csikszentmihalyi, M.: Artistic problems and their solution: an exploration of creativity in the arts. PhD Thesis. University of Chicago, Chicago (1965)
7. Dennis, A.R., Valacich, J.S., Nunamaker Jr., J.F.: An experimental investigation of the effects of group size in an electronic meeting environment. IEEE Transactions on Systems, Man, and Cybernetics 20(5), 1049–1057 (1990)
8. Diehl, M., Stroebe, W.: Productivity Loss in Brainstorming Groups: Toward the Solution of the Riddle. Journal of Personality and Social Psychology 53(3), 497–509 (1987)
9. Diehl, M., Stroebe, W.: Productivity Loss In idea-generation groups: Tracking down the blocking effect. Journal of Personality and Social Psychology 61(3), 392–403 (1991)
10. Dorta, T., Lesage, A., Pérez, E., Bastien, J.M.C.: Signs of Collaborative Ideation and the Hybrid Ideation Space. In: ICDC 2010, Kobe, Japan, pp. 199–206. Springer (2010)
11. Dorta, T., Pérez, E., Lesage, A.: The Ideation Gap: Hybrid tools, Design flow and Practice. Design Studies 29(2), 121–141 (2007)
12. Eversheim, W.: Innovation Management for Technical Products: Systematic and Integrated Product Development and Production Planning, p. 384. Springer (2010)
13. Geschka, H., Schaude, G.R., Schlicksupp, H.: Modern Techniques for Solving Problems. Chemical Engineering, pp. 91–97 (1973)
14. Geyer, F., Pfeil, U., Höchtl, A., Budzinski, J., Reiterer, H.: Designing Reality-Based Interfaces for Creative Group Work. In: C&C 2011, pp. 165–174. ACM Press, New York (2011)

15. Guimbretière, F., Stone, M., Winograd, T.: Fluid interaction with high-resolution wall-size displays. In: UIST 2001, pp. 21–30. ACM Press, New York (2001)
16. Hailpern, J., Hinterbichler, E., Leppert, C., Cook, D., Bailey, B.P.: TEAM STORM: Demonstrating an Interaction Model for Working with Multiple Ideas during Creative Group Work. In: C&C 2007, pp. 193–202. ACM Press, New York (2007)
17. Haller, M., Leitner, J., Seifried, T., Wallace, J., Scott, S., Richter, C., Brandl, P., Gokcezade, A., Hunter, S.: The NiCE Discussion Room: Integrating Paper and Digital Media to Support Co-Located Group Meetings. In: CHI 2010, pp. 609–618. ACM Press, New York (2010)
18. Hilliges, O., Terrenghi, L., Boring, S., Kim, D., Richter, H., Butz, A.: Designing for collaborative creative problem solving. In: C&C 2007, pp. 137–146. ACM Press, New York (2007)
19. Isenberg, P., Tang, A., Carpendale, S.: An exploratory study of visual information analysis. In: CHI 2008, pp. 1217–1226. ACM Press, New York (2008)
20. Liikkanen, L.A., Kuikkaniemi, K., Lievonen, P., Ojala, P.: Next Step in Electronic Brainstorming: Collaborative Creativity with the Web. In: CHI EA 2011, pp. 2029–2034. ACM Press, New York (2011)
21. Liikkanen, L.A., Hämäläinen, M.M., Häggman, A., Björklund, T., Koskinen, M.P.: Quantitative Evaluation of the Effectiveness of Idea Generation in the Wild. In: Kurosu, M. (ed.) HCD 2011. LNCS, vol. 6776, pp. 120–129. Springer, Heidelberg (2011)
22. Massimini, F., Carli, M.: La selezione psicologica umana tra biologia e cultura. In: Massimini, F., Inghilleri, P. (eds.) L'esperienza Quotidiana. Franco Angeli, Milan (1986)
23. Osborn, A.F.: Applied Imagination: principles and procedures of creative thinking. Scribner, New York (1953)
24. Paulus, P.B., Brown, V.R.: Toward More Creative and Innovative Group Idea Generation: A Cognitive-Social-Motivational Perspective of Brainstorming. Social and Personality Psychology Compass 1(1), 248–265 (2007)
25. Reinig, B., Briggs, R., Nunamaker Jr., J.F.: On the Measurement of Ideation Quality. Journal of Management Information Systems 23(4), 143–161 (2007)
26. Rohrbach, B.: Kreativ nach Regeln – Methode 635, eine neue Technik zum Lösen von Problemen. Absatzwirtschaft 12, 73–75 (1969)
27. Schlicksup, H.: Innovation, Kreativität und Ideenfindung, pp. 124–126. Vogel Business Media, Würzburg (2004)
28. Shah, J.: Experimental Investigation of Collaborative Techniques for Progressive Idea Generation Techniques in Engineering Design. ASME Design Theory and Methodology Conference (1998)
29. Shah, J.J., Smith, S.M., Vargas-Hermandes, N.: Metrics for measuring ideation effectiveness. Design Studies 24(2), 111–134 (2002)
30. Sutton, R.I., Hargadon, A.: Brainstorming Groups in Context: Effectiveness in a Product Design Firm. Administrative Science Quarterly 41, 685–718 (1996)
31. Tang, A., Tory, M., Po, B., Neumann, P., Carpendale, S.: Collaborative Coupling over Tabletop Displays. In: CHI 2006, pp. 1181–1290. ACM Press, New York (2006)
32. Taylor, D.W., Berry, P.C., Block, C.H.: Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking? Administrative Science Quarterly 3, 23–47 (1958)
33. Valacich, J.S., Dennis, A.R., Connolly, T.: Idea generation in computer-based groups: A new ending to an old story. Organizational Behavior and Human Decision Processes 57(3), 448–467 (1994)

34. VanGundy, A.B.: Techniques of Structured Problem Solving, 2nd edn., pp. 168–169. Van Nostrand Reinhold, New York (1988)
35. Wang, H.-C., Cosley, D., Fussel, S.R.: Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In: CSCW 2010, pp. 103–106. ACM Press, New York (2010)
36. Warfield, J.N., Geschka, H., Hamilton, R.: Methods of Idea Management. Academy for Contemporary Problems. Columbus (1975)
37. Warr, A., O'Neill, E.: Understanding Design as a Social Creative Process. In: C&C 2005, pp. 118–127. ACM Press, New York (2005)
38. Warr, A., O'Neill, E.: Tool Support for Creativity using Externalizations. In: C&C 2007, pp. 127–136. ACM Press, New York (2007)

Accessibility of Public Web Services: A Distant Dream?

Kristiina Nurmela[1], Antti Pirhonen[2], and Airi Salminen[2]

[1] Department of Mathematical Information Technology, University of Jyväskylä, Finland
[2] Department of Computer Science and Information Systems, University of Jyväskylä, Finland
{kristiina.k.nurmela,antti.pirhonen,airi.salminen}@jyu.fi

Abstract. Today, many public services are available online through Web sites. The accessibility of the sites, also to people with disabilities, is important because the accessibility concerns equality of citizens, a cornerstone of democracy. In the current study we carried out a meta-analysis of 17 studies concerning the accessibility of the Web sites of public administration. Furthermore, we assessed the accessibility of Web pages of 12 ministries of the Finnish government. The assessments were based on the Web Content Accessibility Guidelines (WCAG). The results showed that in terms of the WCAG guidelines, the average accessibility of public Web sites is poor. Moreover, there was no improvement in the accessibility in the 2000's and many of the accessibility failures were so simple that they could have been easily avoided. This may indicate that the building of information society is driven by technology, rather than principles of democracy and well-being.

Keywords: Accessibility, public administration, WCAG.

1 Introduction

Building the so-called information society is usually conceptualised primarily as a technological task [1]. This can be seen, for instance, in the national information society programmes. Finland was among the first ones to prepare one, in the mid-1990s [2]. That report contained a technically oriented futuristic vision. More recently, the official objectives have been more human-centric. The [Finnish] National Knowledge Society Strategy for 2007-2015 is entitled "A renewing, human-centric and competitive Finland" [3]. The objective for developing the society is expressed as "A good life in the information society" [3]. However, the approach of this document is still very techno-optimistic. Like in many other countries, high expectations have been set concerning the opportunities of networked society for democracy [see e.g. 4]. This kind of enthusiastic visions should be taken dubiously, if they don't include critical analysis.

Public Web services are meant, at least in principle, for all citizens. They should be in a form which is accessible for everyone. For instance, people with disabilities should be taken into account in the design of the services.

We wanted to find out to what extent the accessibility has been taken into account in the design of public Web sites. Well-known criteria for Web content accessibility

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 566–578, 2013.

have been defined in the guidelines of the World Wide Web Consortium (W3C), called Web Content Accessibility Guidelines (WCAG). Our study was carried out in two steps. First we accomplished a meta-analysis of 17 studies concerning the accessibility of the Web sites of public administration. Then we assessed the accessibility of Web pages of 12 ministries of the Finnish government. The assessments were based on the WCAG guidelines. Before describing the two parts of the study and their results, we first discuss the core concepts of our study: disability, accessibility, and evaluation of accessibility.

1.1 Perspectives on Disability

The concept of disability (along with related concepts like impairment and handicap) has been a topic of lively debate for decades, at least from the early 1960s [5, 6]. Analysis of all the nuances of the debate is out of focus of the current study. We contend ourselves with the introduction of two, clearly distinctive main approaches: the medical and the social approach to disability.

The medical model of disability is probably the most mundane approach in conceptualising what disability is all about. In it, people are classified on the basis of medical criteria to those who are disabled and those who are not. In the approach, disability is seen primarily as individual's problem which should be overcome. Due to the complex relationship between an individual and the society, individual's problem obviously concerns society as well.

In the traditional medical definition of disability, the stumbling block, causing strong arguments is obvious:

"In the context of health experience, a disability is any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being" [7]

As can be seen, this definition goes "out of the frying pan into the fire": while defining disability, it introduces another, similarly disputable and even ethically questionable concept of *normality* of a human being. Even though the expression has been smoothened in the updated version of the same document [8], the individual's perspective is still present.

The contradictory approach is commonly called as *social model of disability*. In this perspective, the relationship between an individual and the society is approached from the opposite direction: disability is seen to result from the structures which prevent some individuals from participating in the function of the society. In other words, no one is disabled as such; the disability appears in the discordance between the individual and the environment.

The contradiction between the social and medical views of disability is nicely compromised in the Convention on the Rights of Persons with Disabilities by United Nations [9]. In its Article 2, the concept of *universal design* is defined to mean:

"*… the design of products, environments, programmes and services to be usable by all people, to the greatest extent possible, without the need for adaptation or specialized design. "Universal design" shall not exclude assistive devices for particular groups of persons with disabilities where this is needed.*"

In other words, the priority is to design for all, yet admitting that in some cases, special arrangements are necessary.

Even though practical implications for the design of public services may be in many cases the same regardless of the perspective, the distinction between the medical and social approaches is important in the long run.

1.2 Perspectives on Accessibility

Regardless of the perspective on disability, accessibility is a fundamental issue and criterion in the development of public services. This is based on wide agreement about the rights of the disabled people, formalised in the Convention on the Rights of Persons with Disabilities [9], in Article 9:

"*States Parties shall also take appropriate measures:*

…

(f) To promote other appropriate forms of assistance and support to persons with disabilities to ensure their access to information;

(g) To promote access for persons with disabilities to new information and communications technologies and systems, including the Internet;

(h) To promote the design, development, production and distribution of accessible information and communications technologies and systems at an early stage, so that these technologies and systems become accessible at minimum cost."

In other words, importance of information access has been seen such a central right, that it is stressed by expressing the issue from the points-of-view of information, related technology and even the development of technology.

Sir Timothy Barnes-Lee himself, the director of the World Wide Web Consortium (W3C), has been one of the key persons behind recent accessibility efforts e.g. in the UK [10]. Indeed, the Internet and the World Wide Web are the technologies which have become essentials for citizens. It is therefore quite natural, that much of the accessibility efforts concern the World Wide Web.

1.3 Evaluation of Accessibility

Web accessibility is usually bind to the accessibility guidelines of W3C, called Web Content Accessibility Guidelines (WCAG), version 1.0 from 1999 [11] and 2.0 from 2008 [12]. The two versions have a lot in common; the later one, however, is applicable in a wider range of devices. For instance, version 2.0 contains four general accessibility principles hierarchically above the detailed guidelines; the principles are thus applicable even when the technology specific, detailed guidelines appear irrelevant.

The WCAG guidelines provide a number of checkpoints that can be used to check the accessibility of Web content. Each checkpoint is associated with one of three priority levels. Level 1 criteria or checkpoints define the basic and most critical features that accessible content *must* satisfy, level 2 checkpoints refer to the criteria that the content *should* satisfy, and level 3 to the criteria that *may* be satisfied. The checkpoints in guidelines are constructed in a hierarchical manner so that the requirements concerning the same feature get higher when proceeding from level 1 to level 2, and further to level 3. For the content to be tested, for example, a Web page or a Web site, the guidelines define three conformity levels: A, AA, and AAA. In order to be on level A, all level 1 criteria have to be satisfied. On level AA, all the criteria of levels 1 and 2 have to be satisfied. Finally, on level AAA, all criteria of all levels have to be satisfied.

There are several other guidelines and criteria for accessibility, but in this study, we focus on WCAG.

WCAG can be applied both in the design of accessible Web content as well as in the evaluation of accessibility of existing content. The evaluation methods can be divided into automated, semi-automated, and manual. Since in the definition of WCAG it is stated that the criteria has to be testable [12], the success criteria are concrete and many of the criteria can be tested automatically.

2 Meta-analysis of Previous Public Administration Accessibility Studies

In order to get an overall view of the current state of accessibility in public administration Web sites, we carried out a meta-analysis of existing relevant accessibility studies which have been reported in scientific forums.

2.1 The Data

The criteria and method for the article sample was as follows: They had to be published in the proceedings of international conferences, journals, or in the final reports of large research projects. The reports were searched from the databases of ACM, SpringerLink, and Elsevier. Google Scholar was also utilised, as well as the reference lists of relevant articles. The keywords used in the searches were ”WCAG 1.0”, ”WCAG 2.0”, ”accessibility”, ”Web sites”, ”public government” and ”public administration”. Since WCAG 1.0 was released in 1999, only articles which have been published after that were included.

We found 23 studies in total. In a closer inspection, it was found out that four of them did not report the application of WCAG detailed enough for the needs of this analysis. In addition, one report was excluded because of quality problems (did not actually use WCAG 2.0 criteria even if it was argued in the report that it did). The final list of 17 included studies is provided in Table 1.

Table 1. The included studies

Authors & Countries	**Method:** A=automatic M=manual	**No. of analysed pages**	**A-level WCAG 1.0/ 2.0**	**AA-level WCAG 1.0/ 2.0**
Abanumy et al. [13], Saudi-Arabia & Oman	A (Bobby[1])	27	0	*
Al-Khalifa [14], Saudi-Arabia	A, M	36	*/0	*/0
Basdekis et al. [15], Greece	A, M (Bobby)	256	14	1
CabinetOffice [16], EU	A, M	436	3	0
Choudrie et al. [17], Singapore, Australia, Canada, Hong Kong & Finland	A (WebXACT[2])	5	40	0
EU [18], EU & USA	A, M (TAW[3])	102	0/0	0
Goette et al. [19], USA	A (Bobby)	51	70	2
Kuzma et al. [20], USA	A (Truwex)	50	12/0	*
Lazar et al. [21], USA	A, M (A-Prompt[4])	50	2	*
Loiacono et al. [22], USA	A (Bobby)	221	28	*
MeAC [23], 25 EU-countries, Australia, Canada & USA	A, M	336	5.3	0
Paris [24], Ireland	A (Bobby)	26	14	0
Potter [25], USA	A (Bobby)	63	19	2
Shah et al. [26], Nepal	A (Bobby)	27	11.1	*
Shi [27], China & Australia	A (Bobby)	CN 30 AU 8	CN 3 AU 87.5	*
Shi [28], China	A (Bobby)	339	0	*
Yu et al. [29], USA	A (Kelvin[29])	272778	0	0

[1] http://www.erigami.com/truwex/
[2] http://oa.mo.gov/itsd/cio/architecture/domains/interface/PC-Accessibility-MOITAccessiblityStds-WatchfireWebxact060606.pdf
[3] http://www.tawdis.net/
[4] http://www.aprompt.co.uk/45/Website_Accessibility_Testing/

The names in the third column refer to the automated tool used. The figures in the 4th and 5th column indicate the percentages. "*" stands for N/A.

2.2 The Level of Accessibility

As can be seen in Table 1, in four of the included studies none of the analysed pages met even the WCAG level A. In three studies, less than ten per cent of the pages reached level A. In six studies, the percentage of level A pages was between 10 and 25. In four studies only the percentage of pages that satisfied level A criteria, was above 25.

In 10 of the included studies the conformance to level AA was assessed. In them, the percentage of the level AA pages was 2 at its best. In six studies the conformance to level AAA criteria was assessed as well, but not a single page met that level.

It has to be kept in mind that for reaching the compliance level A the page cannot fail in a single level 1 criterion. For reaching the compliance level AA, the page has to satisfy all level 1 and 2 criteria. And only pages satisfying all of the criteria, on all three levels, reach level AAA compliance. This means that one failure on level 1 is enough to prevent calling a page WCAG compliant, on any level, even though all other criteria were satisfied. Therefore it is important to take a look at the results concerning individual success criteria and the most typical failures concerning them.

2.3 Typical Accessibility Failures

WCAG 1.0 Priority 1 Checkpoints

- Checkpoint 1.1: "*Provide a text equivalent for every non-text element.*" This is essential e.g. for blind users applying screen readers. All of the included 17 studies mentioned this as one of the main accessibility problems.
- Checkpoint 12.1: "*Title each frame to facilitate frame identification and navigation.*" Even though the use of frames is not as usual as previously, they remain one of the most common accessibility challenges. Screen readers, for example, do not necessarily read the contents of frames. The titling of frames was found as a problem in 7 studies.
- Checkpoint 6.3: "*Ensure that pages are usable when scripts, applets, or other programmatic objects are turned off or not supported. If this is not possible, provide equivalent information on an alternative accessible page.*" Often, scripts are disabled, so if the use of a Web site requires the functioning of scripts, they are an accessibility problem for anyone. Six of the studies found problems in the use of scripts.

WCAG 1.0 Priority 2 Checkpoints

Ten of the included studies assessed level AA conformity. The most common failures concerned:

- Checkpoint 3.4: "*Use relative rather than absolute units in mark-up language attribute values and style sheet property values*." From the point of view of accessibility, it is important that the user-interface elements are scalable to different needs and terminals. Eight studies mentioned this as an issue.
- Checkpoint 9.3: "*For scripts, specify logical event handlers rather than device-dependent event handlers*." Device dependent technology is problematic since especially disabled people may be unable to use standard devices and have to use customised technology instead. The violation of this guideline was detected in five studies.
- Checkpoint 13.1: "*Clearly identify the target of each link*." Important for users with cognitive disabilities and users with visual impairments in particular. Classified as a problem in four studies.
- Checkpoint 3.2: "*Create documents that validate to published formal grammars*." Valid code ensures that it works in different platforms, including assistive technologies. Either HTML or CSS code was found invalid in four studies.
- Checkpoint 12.4: "*Associate labels explicitly with their controls*." This is important especially for the visually impaired users, who are unable to utilise visual cues in associating e.g. a form field and its help text.

WCAG 1.0 Priority 3 Checkpoints

In the six studies assessing level AAA conformance the priority 3 failures mostly concerned

- Checkpoint 4.3: "*Identify the primary natural language of a document*."
- Checkpoint 5.5: "*Provide summaries for tables*."
- Checkpoint 10.5: "*Until user agents (including assistive technologies) render adjacent links distinctly, include non-link, printable characters (surrounded by spaces) between adjacent links*."

WCAG 2.0

Surprisingly, only one of the studies contained analysis of the accessibility problems on the basis of version 2.0 of WCAG. Therefore, it was not possible to make any comparisons or generalisations about them.

2.4 Summary of the Meta-analysis

Table 1 shows that in average the level of accessibility in terms of WCAG is low. It also shows major differences in the results, especially concerning the level A compliance of public sector Web pages of the case countries or areas. Based on this meta-analysis we cannot, however, make any kind of ranking of the countries or areas in respect to their WCAG level A compliance. Likewise, any other kind of direct comparisons among the included studies is problematic due to their evident differences; without going into details in the differences in the data and analysis methods of the included studies, it is easy to see that already in the number of pages assessed in the studies there is a huge range: from 5 to 272778. The earliest of the studies has been published in 2002, the latest in 2011. Our expectation was that some improvement might be visible during the time frame from 2002 to 2011. To our surprise no such improvement could be seen. In other words, there is no sign in this sample that accessibility had improved over recent years. Another surprise was related to the individual failures and their types. Most of the violations of WCAG checkpoints were plain and apparent. Since the form of WCAG is extremely pragmatic and concrete, most of the identified accessibility violations would have been easy to avoid. This may indicate an attitude problem rather than lack of resources. The issue is further discussed in the last section of this paper.

3 Accessibility of the Web Sites of Finnish Ministries

After the meta-analysis indicating major accessibility problems of the Web sites of the public administration in the USA and many other countries we wanted to assess the accessibility level of the Finnish public administration.

3.1 The Data and the Analysis

We chose the Web sites of the Finnish ministries for the assessment because the Finnish Government has emphasized for several years the importance of building the Finnish Information Society. In the Finnish Government there are 12 ministries. From the ministry Web sites we analysed 108 pages altogether. The chosen pages included the front page of each of the ministries, and the second-level pages which were linked from the main navigation bar of the front page.

In order to gain commensurability with analysis of previous studies, we used a similar kind of approach: We analysed the WCAG compliance and searched for typical accessibility failures. Most of the analysis was carried out with the help of automated accessibility tool, called Worldspace [30]. We ended up to an automated tool since this strategy is usually high in reliability. Automatic analysis is prone to validity problems [31], though, which has to be taken into account in the interpretation of results. In addition to the application of Worldspace, the validity of the HTML code was checked with the markup validator of W3C. From the available automated tools Worldspace was chosen since it is free, covers both versions of WCAG, is included in the listing of W3C and is easy-to-use. It covers all conformity levels of WCAG 1.0 and levels A and AA of version 2.0.

Since automated analysis is purely based on the HTML code, ideally all the results would be manually checked. In the current study, in terms of the resources available, full scale manual inspection was not an option.

3.2 Detected Accessibility Failures

Criteria: WCAG 1.0

Our analysis shows that out of the 108 Web pages, only 9 reached the level A conformance of WCAG 1.0. Of these 9, 8 were pages of the Ministry of Transport and Communications, which is a positive indication – at least the Ministry which is responsible for the domestic Internet policy has applied the accessibility guidelines. Unfortunately, even this ministry's expertise has not yielded above level A: no single page in the whole sample reached the AA conformance, not to speak about AAA.

The common violations of WCAG 1.0 are quite in accordance with the previous studies. We now list the ones which came up in more than half of the pages, or at least three most common problems in each level.

The foremost violation of the first level checkpoints concerned the checkpoint 1.1: "*Provide a text equivalent for every non-text element.*" The failure to meet this criterion was detected in 80 per cent of the pages. The second most common issue on the first level was about the checkpoint 12.1: "*Title each frame to facilitate frame identification and navigation.*" The difference between the frequency of 1.1 problems and 12.1 problems was huge: only 10% of the pages violated guideline 12.1. The third most common violation was about 6.1: "*Organize documents so they may be read without style sheets*", which was violated in 8% of the pages. In other words, just by providing text equivalents for all non-text elements, the accessibility rating of these pages would have been hugely higher.

On the second priority level the problems were more evenly distributed among checkpoints:

- 13.1: 91% of the pages violating ("*Clearly identify the target of each link*")
- 3.2: 91% ("*Create documents that validate to published formal grammars*")
- 12.4: 85% ("*Associate labels explicitly with their controls*")
- 3.5: 68% ("*Use header elements to convey document structure and use them according to specification*")
- 3.4: 63% ("*Use relative rather than absolute units in markup language attribute values and style sheet property values*")
- 10.1: 62% ("*Until user agents allow users to turn off spawned windows, do not cause pop-ups or other windows to appear and do not change the current window without informing the user*")

Of these, only 3.5 and 10.1 did not come up in the list of most common accessibility issues in the review of previous studies. This shows that the types of accessibility problems are quite similar on the Web sites of public administration in different countries.

On the third priority level, the number of problems was huge, the most common ones distributed as follows:

- 10.5: 95% ("*Until user agents (including assistive technologies) render adjacent links distinctly, include non-link, printable characters (surrounded by spaces) between adjacent links*")
- 9.5: 88% ("*Provide keyboard shortcuts to important links (including those in client-side image maps), form controls, and groups of form controls*")
- 13.6: 88% ("*Group related links, identify the group (for user agents), and, until user agents do so, provide a way to bypass the group*")

- 2.2: 81% ("*Ensure that foreground and background color combinations provide sufficient contrast when viewed by someone having color deficits or when viewed on a black and white screen*")
- 4.3: 71% ("*Identify the primary natural language of a document*")

A good example is the most common of the level 3 failures. Guideline 10.5 had been very simple to follow. Whether it is a question of negligence, ignorance or something else, obviously cannot be answered on the basis of this study.

Criteria: WCAG 2.0
In terms of the WCAG version 2.0, none of the pages of our sample reached even the lowest level of conformance (A). In the review of previous studies, the conclusions were similar; only very few pages of public administration reached the A level.

According to the results of the automated analysis, the most common accessibility problems concerned criteria on the first priority level, in particular, the following criteria:

- 1.3.1: 93% ("*Information, structure, and relationships conveyed through presentation can be programmatically determined or are available in text*")
- 3.3.1, 89% ("*If an input error is automatically detected, the item that is in error is identified and the error is described to the user in text*")
- 1.1.1, 77% ("*All non-text content that is presented to the user has a text alternative that serves the equivalent purpose, except for the situations listed below*")
- 3.1.1, 71% ("*The default human language of each Web page can be programmatically determined*")
- 4.1.2, 67% ("*For all user interface components…, the name and role can be programmatically determined; states, properties, and values that can be set by the user can be programmatically set; and notification of changes to these items is available to user agents, including assistive technologies*")
- 2.4.4, 61% ("*The purpose of each link can be determined from the link text alone or from the link text together with its programmatically determined link context, except where the purpose of the link would be ambiguous to users in general*")

Surprisingly, only one problem was detected by Worldspace concerning the second level success criteria. The failure concerned the criterion 1.4.3: "*The visual presentation of text and images of text has a contrast ratio of at least 4.5:1*"(…plus some UI-element specific refinements). From the tested pages 49% failed to meet the criterion.

4 Discussion

There are today a great number of different public services available through Web sites. The limitations in the accessibility of the content of a site may fully prevent using the main functionalities of the site. Our study reveals that the level of Web content accessibility of public administration is in low level. Since accessibility of public Web services is essential in terms of the principles of democracy, improving the accessibility is a necessity.

We used WCAG as a criterion for accessibility in this study. Those guidelines have their strength in being concrete and measurable. They do not, however, take a stand on more ambiguous qualities like attitudes or ethical values. Therefore, there may be a risk that the usage of guidelines is mechanical fixing of errors. In other words, application is first designed and only after the initial design accessibility guidelines are utilised. An analogous problem has already been handled in human-computer studies for quite a while ago: When personal computers became common, user-interface was seen as a means to make the complicated computer technology usable for everyone. It was found soon, however, that a nice user-interface cannot compensate fundamental usability problems. The solution was to change the process: rather than creating unusable technology and only after that try to make it usable with a fancy user-interface, we should include the user's point-of-view in all stages in design.

We argue that the approach described above should be applied to the design of Web content. Rather than creating inaccessible content and later try to fix it with the help of guidelines, the content should be designed accessible in the first place. The form of technology should not be the primary criterion for the development of public services – accessible service may not always be the fanciest looking and not even the cheapest option.

It can be concluded that the construction of accessible services is not a huge technological challenge. The biggest challenge is in the attitudes. Designers of public services should see their role as constructors of democratic society rather than only as application designers [32].

References

1. Halttunen, V., Maksimainen, J., Pirhonen, A.: Less, Slower, Better. Do Information Society Visions Have Healthy Alternatives? In: Suomi, R., Ilveskoski, I. (eds.) Navigating the Fragmented Innovation Landscape. Proceedings of the Third International Conference on Well-Being in the Information Society (WIS 2010), pp. 77–88. TUCS General Publication no. 56, University of Turku (2010)
2. Finland's Way to the Information Society, Ministry of Finance, Edita, Helsinki (1995)
3. Renewing, A.: Human-Centric and Competitive Finland. The National Knowledge Society Strategy 2007-2015. Government Policy Programmes, Prime Minister's Office (2006), http://vnk.fi/julkaisukansio/2006/tietoyhteiskuntaneuvosto/tietoyhteiskuntastrategia/en.pdf
4. La Porte, T.M., Demchak, C.C., de Jong, M.: Democracy And Bureaucracy In The Age Of The Web: Empirical Findings and Theoretical Speculations. Administration & Society 34(4), 411–446 (2002)
5. Oliver, M.: Defining Impairment and Disability: Issues at Stake. In: Barnes, C., Mercer, G. (eds.) Exploring the Divide: Illness and Disability, pp. 39–54. The Disability Press, Leeds (1996)
6. Oliver, M.: Understanding Disability: from Theory to Practice. MacMillan, London (1996)
7. WHO (World Health Organisation) International Classification of Impairments, Disabilities, and Handicaps: A Manual of Classification Relating to the Consequences of Disease. World Health Organization, Office of Publications, Geneva, Switzerland (1980)

8. WHO (World Health Organisation) International Classification of Impairments, Disabilities, and Handicaps: A Manual of Classification Relating to the Consequences of Disease. Final draft, WHO/EIP/GPE/CAS/ICIDH-2 FI/ 01.1 World Health Organization, Office of Publications, Geneva, Switzerland (2001)
9. United Nations. Convention on the Rights of Persons with Disabilities (2006), http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf
10. http://Data.gov.uk
11. W3C. Web Content Accessibility Guidelines (WCAG) 1.0 (1999), http://www.w3.org/TR/WCAG10/
12. W3C. Web Content Accessibility Guidelines (WCAG) 2.0 (2008), http://www.w3.org/TR/WCAG20/
13. Abanumy, A., Al-Badi, A., Mayhew, P.: e-Government Website Accessibility: in Depth Evaluation of Saudi Arabia and Oman. The Electronic Journal of e-Government 3(3), 99–106 (2005)
14. Al-Khalifa, H.: The Accessibility of Saudi Arabia Government Web Sites: An Exploratory Study. Universal Access in the Information Society 11(2), 201–210 (2010)
15. Basdekis, I., Klironomos, I., Metaxas, I., Stephanidis, C.: An Overview of Web Accessibility in Greece: A Comparative Study 2004–2008. Universal Access in the Information Society 9(2), 185–190 (2010)
16. Cabinet Office: eAccessibility of Public Sector Services in the European Union (2005), http://webarchive.nationalarchives.gov.uk/+/http://www.cabinetoffice.gov.uk/e-government/resources/eaccessibility/content.asp (accessed January 10, 2013)
17. Choudrie, J., Ghinea, G., Weerakkody, V.: Evaluating Global e-government Sites: A View Using Web Diagnostic Tools. Electronic Journal of E-government 2(2), 105–114 (2004)
18. EU: Study on "Web Accessibility in European Countries: Level of Compliance with Latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to Implement Those Specifications" (2009), http://ec.europa.eu/information_society/activities/einclusion/library/studies/docs/access_comply_main.pdf (accessed January 10, 2013)
19. Goette, T., Collier, C., White, D.: An Exploratory Study of the Accessibility of State Government Web Sites. Universal Access in the Information Society 5(1), 41–50 (2006)
20. Kuzma, J., Weisenborn, G., Philippe, T., Gabel, A., Dolechek, R.: Analysis of U. S. Senate Web Sites For Disability Accessibility. International Journal of Business Research 9(6), 174–181 (2009)
21. Lazar, J., Beere, P., Greenidge, K.-D., Nagappa, Y.: Web Accessibility in the Mid-Atlantic United States: A Study of 50 Homepages. Universal Access in the Information Society 2(4), 331–341 (2003)
22. Loiacono, E., McCoy, S.: Website Accessibility: A Cross-Sector Comparison. Universal Access in the Information Society 4(4), 393–399 (2006)
23. MeAC: Measuring Progress of eAccessibility in Europe. Assessment of the Status of eAccessibility in Europe (2007), http://ec.europa.eu/information_society/activities/einclusion/docs/subgroupmtg_jan10/meac_study/meac_report_06_11_final.pdf (accessed January 10, 2013)

24. Paris, M.: Website accessibility: A Survey of Local e-Government Websites and Legislation in Northern Ireland. Universal Access in the Information Society 4(4), 292–299 (2006)
25. Potter, A.: Accessibility of Alabama Government Web sites. Journal of Government Information 29(5), 303–317 (2002)
26. Shah, B.P., Shakya, S.: Evaluating the Web Accessibility of Websites of the Central Government of Nepal. In: ICEGOV 2007 Proceedings of the 1st International Conference on Theory and Practice of Electronic Governance, pp. 447–448 (2007)
27. Shi, Y.: E-Government Web Site Accessibility in Australia and China: A Longitudinal Study. Social Science Computer Review 24(3), 378–386 (2006)
28. Shi, Y.: The Accessibility of Chinese Local Government Web Sites: An Exploratory Study. Government Information Quarterly 24(2), 377–403 (2007)
29. Yu, D.X., Parmanto, B.: U.S. State Government Websites Demonstrate Better in Terms of Accessibility Compared to Federal Government and Commercial Websites. Government Information Quarterly 28(4), 484–490 (2011)
30. Worldspace, `http://www.deque.com/`
31. Vigo, M., Brajnik, G.: Automatic web accessibility metrics: Where we are and where we can go. Interacting with Computers 23(2), 137–155 (2011)
32. Pirhonen, A., Maksimainen, J., Sillence, E.: The Design of Information Technology as Design of the Information Society. In: Li, H. (ed.) Studies on Inequalities in Information Society. Proceedings of the Fourth International Conference on Well-Being in the Information Society (WIS 2012), pp. 125–132. TUCS Lecture Notes, No 16, University of Turku (August 2012)

Augmenting Accessibility Guidelines with User Ability Rationales

Christophe Ponsard[1], Pascal Beaujeant[2], and Jean Vanderdonckt[2]

[1] CETIC Research Centre, Gosselies, Belgium
[2] Louvain School of Management, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
christophe.ponsard@cetic.be, {pascal.beaujeant,jean.vanderdonckt}@uclouvain.be

Abstract. Designing accessible web sites and more generally Internet-connected devices remains a challenging task nowadays. A number of guidelines (such as the WCAG2) are now widely available and recognised. To better cope with the quickly evolving technological landscape, these guidelines are also being formulated in technology-neutral terms. However this is still leaving the user dimension largely implicit, which makes it difficult to understand exactly which kind of user a given website is hindering.

This paper describes how to capture and use rational links between guidelines and user capabilities/impairments by combining a set of complementary models (user, task, user interface, guidelines). The process of building those accessibility rationales relies upon available user and guidelines ontologies and also on obstacle identification and resolution techniques borrowed from the requirements engineering domain. This resulting enriched guidance enables a number of interesting new scenarios to better help web developers, analyse guidelines or make comparisons between guidelines.

Keywords: Accessibility, Assessment, Web, User-Model, Task Model, Ontology, Guidelines, WCAG.

1 Introduction

Keeping the web accessible to all and especially to people with disabilities has become fundamental given the importance the web as gained in many aspects of life such as education, employment, commerce, health care, government, leisure, and more. Considering sight and hearing impairments which are among the most problematic when considering web usage, about 6.4 percent of the population age 15 years and older and more than 20% of the population over 65 is concerned in the US, according to a 2010 report [2]. Those figures can be transposed to other developed countries and are far higher in developing countries.

Applying accessibility principles is also widely recognised as enhancing the overall usability experience for everyone whether or not suffering impairment.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 579–586, 2013.

Accessibility standards such as the WCAG [13] or section 508 [11] are widely recognised and formulated as practical guidelines to ease adoption.

Despite those facts, recent reports keep showing that the level of accessibility remains very low. For example, a 2008 European Commission study highlighted that only 2.6% of key public and commercial websites in Member States were accessible, while only 5.3% of government websites were accessible [6]. Explaining the gap between the need and the current lack of accessible websites is not easy and multiple factors are involved. The fast pace of technical evolution does not help but beyond that, one of the key reasons frequently mentioned is the lack of information and training of web designers.

Accessibility guidelines have quite matured and now clearly separate the accessibility principles and guidelines from specific techniques to enforce them in specific technologies [13]. However, while web designers are well armed to face “how” to apply the principles, the support to help them understand “why” they need to apply specific guidelines is still largely missing; e.g. the WCAG2, focussed upon in this paper, does not explicitly relate the need of captions in video content with deafness.

In order to provide better support for the web designer, this paper proposes to enhance the existing guidelines with such rationales. We show how to systematically produce them based on the existing guidelines combined with user impairment ontologies and user interface modelling techniques abstracting from the pure implementation level (such as abstract user interfaces and tasks models). In order to represent and reason about the accessibility related knowledge, we rely on semantic web techniques like RDF/OWL representations, queries and ontology interconnections [10]. Those techniques also enable an access as open linked data.

The structure of this paper follows our research approach. First, in Section 2, we survey the available models and related ontologies which can be exploited and connected to better support the design process. Section 3 then details how additional links can be inferred to bridge important gaps such as the rationales between guidelines and user impairments. Section 4 then illustrates a number of scenarios on how the enriched guidelines bring added value. Section 5 reviews some related works; and finally section 6 concludes by highlighting how this work can be further developed.

2 Survey of Accessibility Related Ontologies and Models

This section presents the characteristics (strength/limitations) of existing accessibility and impairment ontologies as well as relevant User and UI modelling techniques.

2.1 Accessibility and Impairement Ontologies

Web Content Accessibility Guidelines. WCAG guidelines are expressed as principles in four major categories: perception, operability, understandability and robustness. Those are refined into 12 major guidelines which are then supported by specific techniques, either generic or specific to a technology. Related checks are also defined to keep the traceability between these guidelines. Its structure can be represented as a goal refinement tree depicted in Figure 1 and explained in more details in section 2.2.

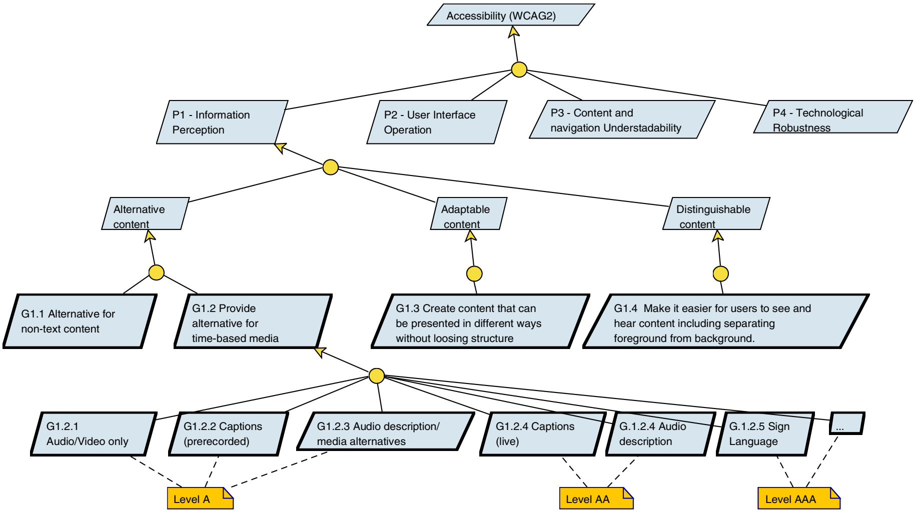


Fig. 1. Partial goal model for WCAG2 guidelines

The guidelines are mostly available as well-organised structured hypertext on the W3C website. The AEGIS project has also made them available as ontology under the OWL format which can be used for semantic processing [1]. Figure 2 illustrates this ontology in the Open Source Protégé tool.

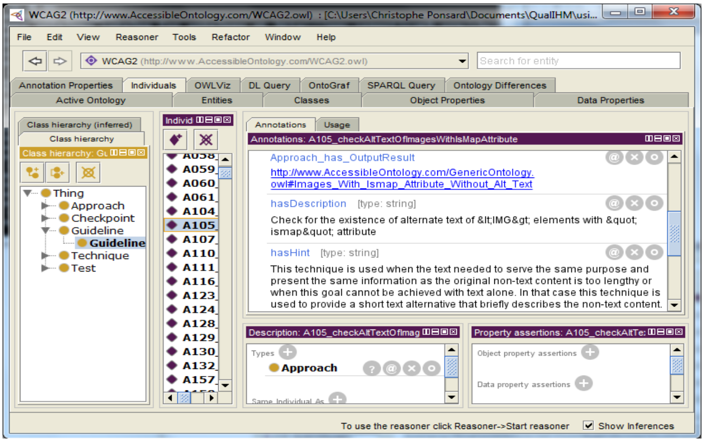


Fig. 2. WCAG2 Ontology

International Classification of Functioning, Disability and Health (ICF) [5]. It is a universal classification of disability and health for the definition, measurement and policy formulations in health and health-related sectors. The naming reflects its philosophy oriented to the measuring functioning in society, no matter what the reason for one's impairments. This is reflected by its conceptual structure represented in

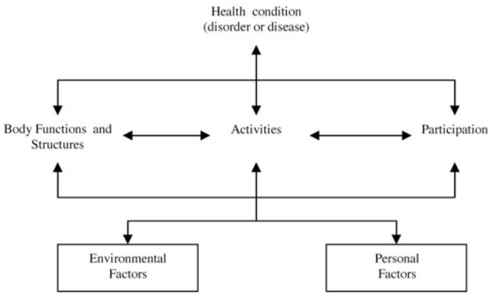


Fig. 3. ICF Concepts and Relationships

Figure 3. It is therefore relevant to consider this classification also in the specific context of IT usage and web accessibility. An extensive OWL ontology has been released by the BioPortal, composed of about 1600 classes and 3250 individuals [3].

2.2 Models

User Interface Models provide representations of the user interaction with a computer program or another reactive device with the purpose of supporting the UI design and analysis process. Different aspects of human-machine interaction can be captured by a set of the following complementary models and are supported by specification languages like UsiXML [12].

- **Task model** enables the description of high-level user requirements in terms of activities to be performed by the user and/or by the system in order to reach some goal [14]. They have a hierarchical graphical syntax with a well-defined semantics.
- **Abstract UI model** defines interaction spaces grouping subtasks according to various criteria and independently of any context/modality of use.
- **Concrete UI model** define widgets layout and navigation. While making the Look & Feel explicit, it is still a mock-up rendered in a non-operational environment.

Goal **models** are used in the larger context of requirements engineering of system [15] while task models are more specifically targeting UI design. Goals capture properties to be achieved by users together with the systems. Goals can be refined hierarchically and can be seen to some point as a generalisation of tasks. Goals support reasoning about obstacles and conflict by taking into account known domain properties (in our context: UI characteristics and user impairments).

3 Producing Guidelines Rationales

In this section, we present the process to generate accessibility guideline rationales, based on existing techniques used among other in requirements engineering such as obstacle analysis and abduction techniques based on domain knowledge [15]. This rationale generation process is depicted in Figure 4 and relies on the following steps:

1. Content types are extracted from the models, possibly at different level of abstraction (e.g. task model, abstract UI model if available or concrete UI)
2. Obstacle analysis is performed based on the user capability ontology (e.g. ICF).
3. Matching is made with existing guidelines provided by the guideline library (e.g. WCAG2).

The process can be complemented by a bottom-up phase, starting from the existing guidelines and attaching them to obstacle or completing existing obstacle trees.

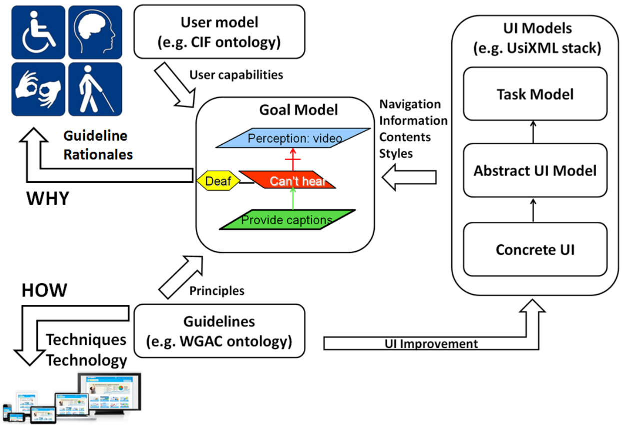


Fig. 4. Rationale Generation Approach

This process can be applied systematically by using the rich ICF and WCAG ontologies identified in the section 2. The captured knowledge can then be bundled into a merged ontology enriched with the extra links illustrated in the following example:

Task – understanding some textual content (typically as a step in workflow such as making a choice acknowledging terms of use)

- **obstacle**: textual content not understood because user may have cognitive limitations or be non-native (including signing deaf)
- **resolution**:
  - identify user language ➔ **guidelines 3.1.1 (global) and 3.1.2 (parts)**
  - propose alternative language ➔ (out of scope of WCAG)

  or support simplified content ➔ **guideline 3.1.5 (reading level)**

Fig. 5. Example of rationale generation

The systematic application of the above process revealed two interesting points. First, it acts as a **completeness check** to detect partly addressed impairments in some usage context. Second, **some models seem to relate more specifically to certain functional impairment categories** (see Table 1). This can help having a special focus when building those models or even support the decision to build such models.

Table 1. Impacted functions in relation with UI models

UI model	UI concepts (added from previous level)	Impacted functions
Task	Information flow, navigation flow	Cognitive
Abstract UI	Controls, content types	Motor, Vision, Hearing
Concrete UI	Look and Feel (styles, colors)	Vision

4 Usage Scenarios

The availability of the guideline rationales enables new usage scenarios for different target groups (web designers, developers, assessors,...). The scenarios reported here were identified together with Anysurfer (`http://www.anysurfer.be`), a major association active in accessibility assessment and training. Some subjective validation elements are also reported.

4.1 Providing User-related Explanations for Web Designers

Currently, the application of guidelines produces a global ranking such as A, AA, AAA with limited explanations in terms of passed/failed checks. Problems can be traced to guidelines and corrective actions. The rational information enables to:

- state which kind of user is impacted by a failed check. The feedback is this helps web designers to better identify the type of impaired users and to better understand and to apply the relevant parts of guidelines.
- provide finer grained assessment going by functional impairment going beyond "black box" A/AA/AAA labels. Some development can for example target a specific user group for which extra requirements have to be implemented.

This support to web designers can be part of an accessibility assessment report but larger gain is expected when it is applied at design stage.

4.2 Aligning Guidelines for Web Designers

This paper is focusing on the WCAG2 which are the major reference guidelines. There are however other guidelines such as section 508 in the US and BS 8878:2010 Web accessibility in UK. For global websites or websites in countries transitioning to WCAG, it is more effective to be able to compare between guidelines rather than performing multiple independent assessments.

Table 2. Guideline Alignment example

ICF	Obstacle to address	WCAG2	Section 508
b.21021 Seing fion: colour vision	Colour only information	G1.4.1	1194.21i
b.21022 Seing fion: contrast sensitivity	Contrast enhancement	G1.4.3 G1.4.6	1194.21g

The proposed process can be applied to multiple guidelines with the same structure, resulting in a natural alignment of guidelines as illustrated in Table 2.

4.3 Checking Gaps in the Guidelines for Accessibility Working Groups

As mentioned in Section 3, the completeness check can point out some obstacles not addressed by the considered guidelines. The validation with accessibility assessors revealed some possible reasons for this: the topic might be related to usability rather than accessibility, or corrective measures might have been considered too advanced/costly to implement. In all cases, it was found interesting to report this information for helping in the future evolution of the guidelines. Some issues can also be reported to web developers as recommendations going beyond the current guidelines.

5 Related Work

Applying ontologies to accessibility has already been proposed. For instance, [16] shows how they help in addressing limitations of natural language but without referring to specific guidelines. This is now a recognised approach as shown in Section 2.

The idea to connect ontologies has been investigated in [7]. A generic accessibility pattern connecting user, capability, interface element and information has been proposed and instantiated to various examples (memory/recall, perception). The work however does not explore the use of renown and standard ontologies like the ICF and WCAG as proposed in our work.

An alternative approach proposed in [9] is to build a specific ontology formally describing the whole information about user's impairments, and the available interface characteristics. This allows a personalised accessibility assessment but the presented scenario does not seem to support the detection of partly covered impairments. We also believe it is better to avoid mixing ontologies addressing different domains to ease future evolution, for example considering different sets of guidelines.

The project ACCESSIBLE developed a harmonised methodology for measuring accessibility, including guideline alignment [8]. Though quite similar, that work looks ICF-driven (top-down) while our approach can also be guidelines driven (bottom-up). The work also covers the alignment of guidelines such as WCAG1/2 and section 508.

AEGIS project developed an Open Accessibility Framework relying on ontologies [1]. It supports the mapping from requirements/constraints of users to characteristics not only of web applications but also desktop and mobile applications. The AEGIS generic accessible user interaction model shares similarities with our work although it is more directed towards personal customization scenarios.

6 Conclusion and Perspectives

The paper shows how accessibility guidelines can usefully be extended to better support accessibility rationales providing better explanation why specific guidelines are required by linking them to impairments. A number of interesting scenarios exploiting these rationales have been identified and investigated.

At this point, the resulting combined ontology is still partial but available from http://www.accessible-it.org/ontologies. On-going work is to enrich them and achieve integration with existing frameworks in order to conduct validation experiment inside a local cluster of web SMEs. We also plan to investigate more specific user capability models such as the DSM-5 [3] dealing with cognitive disorders.

Acknowledgement. This work was funded by the QualIHM project (grant nr 1217570) as part of the collective research programme of the Walloon region. We warmly thank the Belgian AnySurfer association and the reviewers for their insightful feedback.

References

1. AEGIS Project. Open Accessibility Everywhere, http://www.aegis-project.eu
2. Americans With Disabilities: 2010, report from the US Census Bureau (July 2012)
3. American Psychiatric Association, DSM-5 Development, http://www.dsm5.org
4. BioPortal, ICF Ontology, http://bioportal.bioontology.org/ontologies/1411
5. International Classification of Functioning, Disability and Health (ICF), http://www.who.int/icidh
6. European Commission, Web-accessibility in European countries: level of compliance with latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to implement those specifications, SMART 2008/0068 (2008)
7. Karim, S., Latif, K., Tjoa, A.M.: Providing universal accessibility using connecting ontologies: A holistic approach. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 637–646. Springer, Heidelberg (2007)
8. Mourouzis, A., Kastori, G.-E., Votis, K., Bekiaris, E., Tzovaras, D.: A Harmonised Methodology towards Measuring Accessibility. In: Stephanidis, C. (ed.) Universal Access in HCI, Part I, HCII 2009. LNCS, vol. 5614, pp. 578–587. Springer, Heidelberg (2009)
9. Oikonomou, T., Kaklanis, N., Votis, K., Tzovaras, D.: An Accessibility Assessment Framework for Improving Designers Experience in Web Applications. In: Stephanidis, C. (ed.) Universal Access in HCI, Part I, HCII 2011. LNCS, vol. 6765, pp. 258–266. Springer, Heidelberg (2011)
10. OWL Web Ontology Language Reference, http://www.w3.org/TR/owl-ref
11. US Government, Section 508, https://www.section508.gov
12. Vanderdonckt, J., Limbourg, Q., Michotte, B., et al.: USIXML: A User Interface Description Language for Specifying Multimodal User Interfaces. In: W3C Workshop on Multimodal Interaction, Sophia Antipolis (2004)
13. Web Content Accessibility Guidelines (WCAG), http://www.w3.org/TR/WCAG/
14. W3C, MBUI - Task Models, Working Draft (2012), http://www.w3.org/TR/task-models
15. van Lamsweerde, A.: Requirements Engineering: From System Goals to UML Models to Software Specifications. Wiley (2009)
16. Xiong, J., Farenc, C., Winckler, M.: Towards an Ontology-Based Approach for Dealing with Web Guidelines. In: Hartmann, S., Zhou, X., Kirchberg, M. (eds.) WISE 2008. LNCS, vol. 5176, pp. 132–141. Springer, Heidelberg (2008)

Lessons Learned from Crowd Accessibility Services

Hironobu Takagi, Susumu Harada, Daisuke Sato, and Chieko Asakawa

IBM Research – Tokyo, 5-6-52 Toyosu, Koto-ku, 135-8511 Japan
{takagih,haradas,dsato,chie}@jp.ibm.com

Abstract. Crowd accessibility services for people with disabilities, driven by crowd-sourcing methods, are gaining traction as a viable means of realizing innovative services by leveraging both human and machine intelligence. As the approach matures, researchers and practitioners are seeking to build various types of services. However, many of them encounter similar challenges, such as variations in quality and sustaining contributor participation for durable services. There are growing needs to share tangible knowledge about the best practices to help build and maintain successful services. Towards this end, we are sharing our experiences with crowd accessibility services that we have deployed and studied. Initially, we developed a method to analyze the dynamics of contributor participation. We then analyzed the actual data from three service deployments spanning several years. The service types included Web accessibility improvement, text digitization, and video captioning. We then summarize the lessons learned and future research directions for sustainable services.

Keywords: Crowd-sourcing, accessibility, digital book, captioning, Web accessibility.

1 Introduction

Accessibility technologies have advanced significantly over the years, but there still remain many issues that technology alone has not been able to solve. Screen readers and refreshable Braille display technologies opened up a wide array of information to people with visual impairments, but areas such as the automatic conversion of visual information into textual descriptions remain major challenges. Automatic speech recognition and remote captioning technologies have expanded the communication possibilities for people who are deaf or hard of hearing, but for informal speech or conversations in noisy environments, automatic recognition has yet to attain practical levels of reliability.

In recent years, there have been rapid developments of crowd-sourcing technologies to bridge the gaps in mechanical computing technologies, based on harnessing and combining the perceptual, cognitive, and intellectual abilities of human beings. Pioneering services such as the ESP Game [1] have demonstrated the power of this approach. Crowd accessibility services have emerged as a category of crowd-sourcing service focused on supporting people with disabilities. General crowd-sourcing services sometimes are designed to train machine intelligence systems instead of

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 587–604, 2013.

supporting humans. For example, building a dictionary for machine translation can be crowd-sourced to people without directly specifying the human beneficiaries or their specific objectives.

In contrast, crowd accessibility services focus on supporting specific people with disabilities. A captioning service seeks to support people who are deaf or hard of hearing, or a book digitalization service can focus on supporting readers and learners with visual disabilities [2, 8]. One category of new services is called “vision as a service”, and it helps blind people to recognize all of the things around them [3, 9]. The existence of human beings who need support strongly distinguish accessibility services from other services.

In various ways, crowd accessibility services can enable dream-like services for people with disabilities. However, it is sometimes difficult to sustain these services because of the small sizes of the target populations. For sustainable services, the supply of workers and the work requests should be balanced, but this is often difficult for crowd accessibility services. This characteristic generally obliges service providers to depend upon unpaid crowd volunteers.

At the same time, crowd accessibility services also have characteristics that present special opportunities. Since crowd accessibility services tend to be focused more on social services and not commercial aspects, we have learned that these contributors tend to be more motivated by altruistic factors rather than by monetary gains or by the diversion and novelty offered by quick and simple tasks, which are among the motivators for traditional crowd-sourcing services.

How can these services succeed and endure while recruiting new and active contributors? How can we motivate them? How can we insure the quality of outputs? Little work has been done to analyze the issues that are specific to crowd accessibility services, since the analysis methods themselves are challenging given the diverse characteristics of the services and the limited availability of data for such analyses.

In this paper, we offer some answers to these questions by analyzing data from three services of our own. After the review of related work, an analysis method is proposed by taking into account the practical availability of data. After examining our data in the observations section, our lessons learned focus on four topics. Before offering our conclusions, a wide range of future research directions are discussed for social infrastructures based on crowd accessibility approaches.

2 Related Work

2.1 Crowd Accessibility Services

Crowd accessibility services are within the broad category of crowd-sourcing services, but are characterized by their focus on people with disabilities as the beneficiaries of the services. Image labeling for accessibility is one of the popular applications for crowd accessibility. Dardailler [7] pioneered the approach of labeling images using the power of the crowd in the ALT-server proposal that stores alternative text on a remote server. The ESP game [1] used the gamification approach in labeling images to support both image searches and accessibility. This meant the service was not purely a crowd accessibility service, but suitable for general and accessibility applications. Optical character recognition (OCR) technology was combined with crowd-sourcing in the WebInsight system [4]. This was the first approach of combining human intelligence with computer intelligence to improve accessibility. Social Accessibility [17, 18] focused on comprehensive and high quality Web accessibility improvements including image labels. VizWiz [3] explored real-time labeling by recruiting Web workers from the Amazon Mechanical Turk service to answer vision-related questions submitted as photos by blind users using smartphones. [9] is a commercial service to label photos in real-time by combining an image recognition engine with crowds of workers. This service tried to expand the scope from a pure accessibility application to a more general augmented reality application by using general purpose image recognition technologies. In spite of such technical advances such as gamification, OCR integration, authoring tools, and image recognition integration, sustainability still remains an important issue to be addressed by crowd accessibility services. Bigham et al. [5] reviewed the history of crowd accessibility services, summarizing the features and characteristics of 15 services and proposing 13 design dimensions with which to compare the services, such as motivation, accuracy, and reliability. [2, 19].

2.2 Methods to Analyze Participation

This section describes some methods to analyze participation in the crowd services from two perspectives, objective (such as statistics and log analysis) and subjective (such as interviews or surveys). Wikipedia is the largest service that is powered by a crowd of volunteers without monetary incentives, which has led to many analyses of its participants. According to an internal article titled "Wikipedians", there are about 16 million registered accounts and about 0.3 million of them edit Wikipedia articles on a monthly basis [20]. The number of edits per account ordered by magnitude follows a Zipf distribution, which forms a straight line on a double-logarithmic graph [10]. Swartz [16] reported that about 73.4% of Wikipedia editing is done by the top 1,400 users, and most of the remaining editing is minor, such as fixing incorrect spellings. This phenomenon is also described as a 90-9-1 rule [11], since 1% of participants make almost all of the contributions, 9% make minor contributions, and the remaining 90%, the "lurkers" [13], only read the results. Stewart et al. [15] showed that the distribution is different in a closed service limited to company employees, reporting values of 33-66-1.

Objective study is key to gaining an overall picture of the participation in these services, though subjective studies reveal many of the psychological factors driving participation. Nonnecke et al. [14] studied why lurkers lurk in crowd-sourcing services. Bryant et al. [6] interviewed active Wikipedians to analyze how they had evolved from lurkers to leading contributors of Wikipedia. A typical user initially visits Wikipedia to obtain information, then begins making minor contributions and learning about the rules and conventions of the community, and finally becomes registered as a contributor. Their aim changes from just reading or polishing the articles to improving the quality of Wikipedia itself.

Table 1. Evaluated Services

	Service type	Primary target users	Start date	Service duration	Scope	Quality assurance
EBIS	OCR correction	Blind and other print disabilities	2011/8/23	2011/12/15	Internal	Expert check
CCES	Captioning	Deaf and hard of hearing	2011/11/11	2011/12/28	Internal	Expert check
Social Accessibility	Web accessibility metadata authoring	Blind and people with visual impairments	2008/7/8	2010/3/31	External	User report

3 Evaluated Services

Over the past several years we deployed three crowd accessibility services, as summarized in Table 1. For each service, the following sections provide a summary of the service, characteristics of the supporters, task descriptions and definitions of the task units in the logged data, and methods for quality assurance.

3.1 EBIS

EBIS [8] was originally created to rapidly digitize physical books for the blind and other people with print disabilities, focusing on Japanese books that have up to 10,000 characters in various styles, which makes the process quite difficult compared to languages that use phonetic alphabets. The books are unbound and scanned to create images, and then processed through an OCR engine to generate the initial digital output. The raw output from the OCR engine is not accessible, since it tends to contain various character recognition errors and structural problems. There are several steps, but we focused on a crowd-sourcing process.

EBIS is using a check-by-expert approach as the quality assurance method. A skilled contributor was assigned as a proofreader to each book to identify any errors that might have escaped the eyes of the contributors in the earlier phases. This phase was managed manually and with mail-based information exchange. EBIS used gamification based on points and grades. According to contribution, each user earned some number of points and climbed the grade ladder based on the accumulated points.

The system was announced internally within IBM as part of the company's centennial 'celebration of service' event as one of the encouraged volunteer opportunities. This meant the contributors were unpaid volunteers recruited from a large pool of potential participants.

3.2 CCES (Collaborative Captioning Editing System)

CCES is a crowd-sourcing system that adds captions to digital videos [12]. Adding text captions to digital video tends to be a time consuming task, especially when the task work is being performed ondone for content for which no text transcript is available. The CCES service aims to speed up this process by splitting up the video into short segments, automatically segmenting it at detected breaks between phrases, and offering a user interface through which a captioner can type in the text as they listen to the clip. Each CCES task consists of a 30-second clip of video content that is to be transcribed and submitted by the captioner. The 30-second clip is actually split into roughly 10 or so sub-clips that approximate phrase utterances, and the supporter can review and transcribe them one at a time.

The expert check method was used as the quality assurance method in CCES. The system has a user interface to allow administrators to check and fine-tune captions. For example, administrators can adjust the timing, position on screen, length of one caption, and other attributes that are used for caption color coding, such as the gender of a speaker. The system was also announced internally within IBM as part of the company's centennial 'celebration of service' event. As of the writing of this paper, the CCES service has been running for about one month in its newest version.

3.3 Social Accessibility (SA)

The Social Accessibility [17] service was one of the first crowd accessibility projects we developed at our group. It provides a mechanism through which Web consumers such as blind computer users can identify accessibility problems with certain webpages, and then submit requests for improvements to a central server. The request is then made available to contributors who use a Web interface to create a fix for the accessibility problem by creating metadata that augments the original webpage, perhaps by adding alternative text to images. Since the metadata is stored on our server without altering the original website and the Web consumer can retrieve the metadata using our browser plugin, the service enables quick turnaround and wide reach.

The SA service was available to the public, with anyone able to sign up either as a requestor or a contributor. A requestor could submit a request by simply using a browser plugin and pressing a special hotkey when an inaccessible page was encountered. A contributor could view the pending requests with our online interface, and also view the webpage with the problem together with a panel for entering the various kinds of metadata information. A task in SA was a collection of metadata that a contributor submitted for one particular page in one session.

SA employed the user report as its main quality assurance mechanism. Each end user could report incorrect or suspicious errors whenever they used metadata within a seamless user interface with the same tool that applied the metadata to the webpages. Volunteers could also check and improve the quality of the metadata by using the tool that created the metadata, but this was also uncoordinated volunteer work.

4 Analysis Method

Given the diversity of crowd accessibility services, a standard analysis method itself is hard to develop. Therefore, we have developed an analysis method by focusing on the factors affecting the sustainability of crowd accessibility services. The method consists of two types of analysis to cover various aspects, such as transitions and gaps involving participation statuses (Transition Analysis).

The number of new contributors and the ratio who become committed contributors are the critical factors for assessing sustainable contributor participation. One major problem for each analysis method is the limited availability of data that applies across services. Each output of a comparative analysis method should be based on data that is available for all of the crowd accessibility services.

To determine specific challenges for the services, the gaps (or barriers) for participants should be assessed. To increase the number of new contributors, we should determine the gaps facing each new contributor. To increase the number of committed contributors, the gaps in returning should be studied. The transition analysis was designed as a tool to identify these gaps within services. The dynamics of participation can be considered as a series of transitions among participation statuses. Fig. 1 shows our definition of the transitions and Fig. 2 is a bar chart of a model of the statistical transitions among the statuses. The total number of people in all of the statuses refers to the total number of participants at the end of that day.

Visitors are people who visited the service, but did nothing. This is the initial state for a participant. If a person finds and visits the service, but decided not to contribute, then the person's status is "visitor". A **registered visitor** is a person who has registered with (or subscribed to) the service but has not yet contributed. This level of involvement is close to the definition of "lurkers" in social media (Section 2.2). In the case of a crowd accessibility service, people may not be simply lurking, but actively trying to figure out ways to contribute. Ideally, this status should be minimized. The number can be easily calculated from the contributor table in each service. Usually, the contributor table records the registration date as a property. This same data can be used to count the number of changes in the status of a registered user.

A **new contributor** is a person who is contributing to the service for the first time and an **experienced contributor** is someone who has contributed to the service at least twice. In Fig. 2, all of the contributors are new contributors on the first day, and after the second day there will be a mix of returnees (experienced contributors) and new contributors. These two types of contributors indicate different steps in the service. The gap between registered visitors and new contributors reflects the difficulties in starting to contribute after subscribing to the service. An iceberg chart visually shows the time changes of the ratio between registered visitors and newcomers (bottom-side dark gray and upper-side black areas). The gap can be also measured as a ratio:

Registration-contribution ratio = (total number of contributors) / (total number of registered visitors)

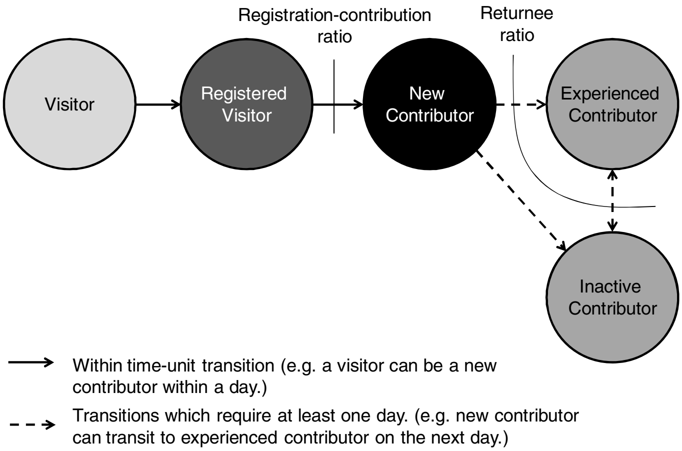


Fig. 1. Transition Diagram among Participation Status

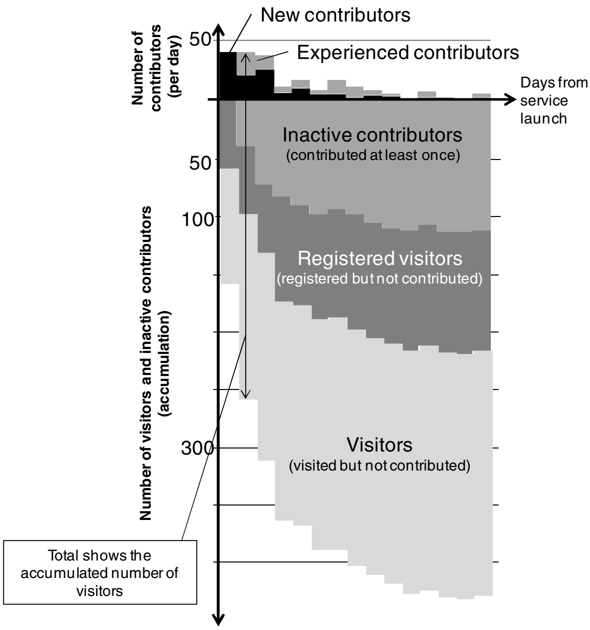


Fig. 2. Iceberg Visualization of Transitions (Contributor Status Chart)

In contrast, the number of returnees (experienced contributors) is a metric to assess the degree of activity of committed contributors. By checking the changes over time in the numbers of new contributors and experienced contributors, we can visually assess the ratio of returnees. Thus, the returnee ratio from new contributors to experienced contributors can be seen as a metric to assess the difficulty of the transition to become a committed contributor.

Returnee ratio = ((total number of contributors) – (number of one-day contributors))/ (total number of contributors)

The definitions of new contributors and experienced contributors take into account the ease-of-calculation based on the properties of output data in a usual crowd accessibility service. For example, metadata is the output in SA, and each piece of metadata has properties including a link to its creator and its creation date. Any service with functions to track "who" worked on "which" task units will have such a data table.

An **inactive contributor** is a person who has not done any work on a given day. This definition means that the active contributor statuses (new and experienced) are temporary and that most of the contributors are in this inactive contributor status.

4.1 Task Analysis

The number of completed tasks in a unit of time is a clear quantitative metric for the activity level of a service. This task analysis focuses on measuring and visualizing the dynamics of completed tasks by new contributors and experienced contributors separately.

The definition of a task unit varies according to the target of a service. One transcribed line of text can be one task unit for captioning services, and one confirmed character can be one task unit for an OCR correction service. The definition can also be affected by the characteristics of the task management for the service. For example, when an OCR correction service manages correction work in units of "one page" then each page can be a task unit. A page can be an acceptable unit as long as the granularity of the task unit is sufficiently fine for the analysis.

5 Observations

We examined three of our crowd accessibility services using the analysis methods described in Section 3. In this section, we will present the analysis of the data for each service. The results will be summarized in the following section as a set of lessons learned.

5.1 EBIS

Fig. 3 shows the analysis results for EBIS. Fig. 3(b) shows several multi-day peaks in the completed tasks. The Within each of these periods, contributors successfully processed 10 to 25 books. The task shortage was strongly linked to the very success of the service. Contributors actively worked and completed all of the available tasks in a short period of time. As seen in the aggregate task count chart in Fig. 3(b), the EBIS service had other periods when there were no tasks being performed by the contributors. The charts show that work typically occurred in spurts spanning two to four days. The contributor data shows that majority of the contributors worked for

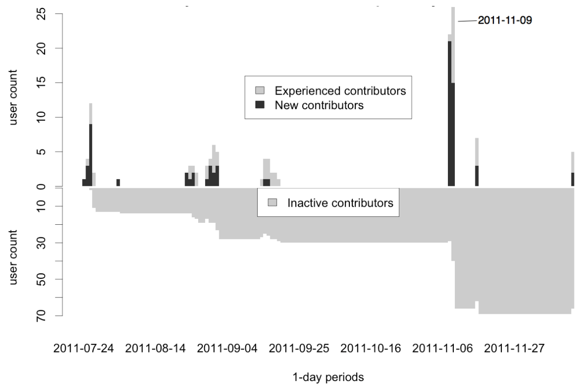


Fig. 3. (a). Contributor Status Chart for EBIS (2011/7/24 ~ 2011/12/14)

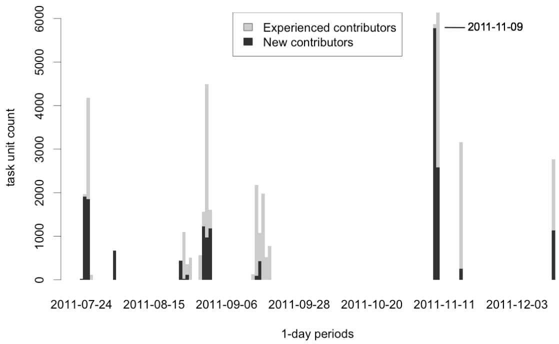


Fig. 3. (b). Aggregate Task Count Chart for EBIS (2011/7/24 ~ 2011/12/14)

no more than three days, roughly corresponding to the average duration of a spurt. The contributor status charts in Fig. 3(a) also show that when the second spurt occurred around November 8, 2011, almost all the contributors were new first-time contributors, meaning that almost none of the previous contributors returned.

The first spike on November 8 was created by an announcement to all employees on the front page of a portal. The small peak on November 17 was created without any announcement, but just by people who discovered the availability of the service. That is why the majority of those contributors were returnees, as shown in Fig. 3(b). The last peak was created by an email announcement to experienced contributors about the availability of new books. In this way, the service successfully engaged with the experienced contributors and they completed the tasks much more quickly than the administrators could prepare additional books.

The underlying cause of the task shortage was the bottleneck due to expert phases in the process. The availability of books was limited by the performance of the experts, and this meant the contributors had to wait until each batch of processed books was proofread before being able to continue with the next batch. This can also be seen in Fig. 3(b), showing that when a new batch of books became available on

November 8, 2011 (after a 2-month hiatus), almost all of the tasks were completed by new contributors and few of the previous contributors returned.

We received various informative comments as feedback from the contributors. One category of comments was about the effectiveness of the gamification feature that allowed the contributors to climb up a grade ladder by completing more tasks (Section 3.1). Some of contributors mentioned that the indication of the remaining task units to go up a level was highly motivating. Not only because of the explicit gamification feature, but many contributors mentioned the task itself gave them an impression of a "game" and that helped them to concentrate on the tasks. Many contributors mentioned that the user interface was similar to the "brain training games" on mobile game consoles (such as the Nintendo DS).

5.2 CCES

Fig. 4(a) shows the contributor status chart for CCES. It can be seen that throughout the duration of the service, there was a relatively steady flow of contributors performing tasks each day, both new and experienced contributors. A survey was conducted to get feedback from contributors. Among all of the contributors, 92.2% of the contributors answered good (51.0%) or very good (41.2%). The other options are 5.9% neutral, 0% bad and 0% very bad with 2.0% no answer. This high ratio of acceptance backs up the result of the steady flow of contributors.

The dips around November 26, 2011 and December 4, 2011 coincided with weekends, reflecting the fact that most contributors performed the tasks on weekdays. The graph also shows that each day there were at least some new contributors joining to perform tasks (indicated by the black bars). Such a participation pattern indicates a successfully operating service, with contributors' task output remaining relatively steady and new contributors steadily joining the service. Fig. 4(b) shows the same tendency as Fig. 4(a). New and experienced contributors share the tasks in a balanced way. The direct interpretation of the observed data is that the service was not sticky enough to encourage contributors to complete many tasks, even for experienced

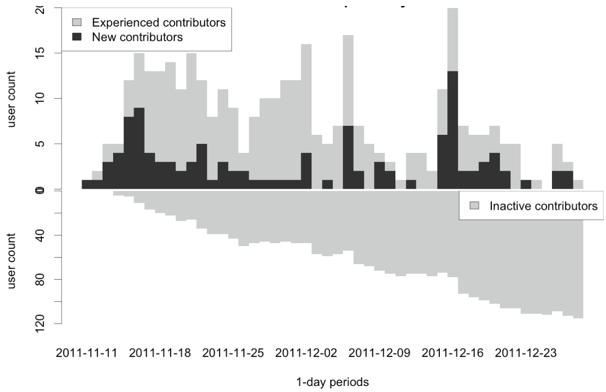


Fig. 4. (a). Contributor Status Chart for CCES (2011/11/11 ~ 2011/12/28)

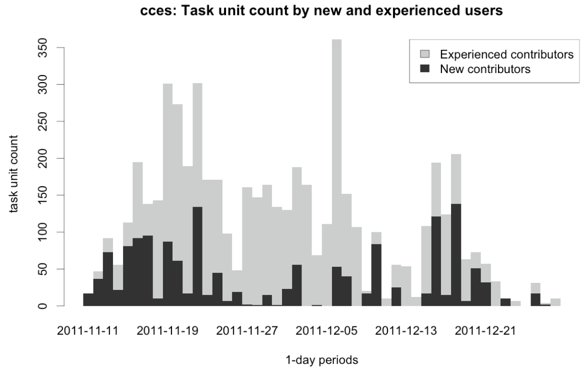


Fig. 4. (b). Aggregate Task Count Chart for CCES (2011/11/11 ~ 2011/12/28)

contributors. Instead, many new people joined the service and each quickly completed a small number of tasks. The phenomenon was not statistically significant and we clearly need more data to study such differences.

5.3 Social Accessibility (SA)

Two key characteristics of the SA service that distinguish it from the other services we have deployed were the length of time the service was available and the contributor pool. As seen in Fig. 5(b), the SA service continued to observe contributor activity for well over six months. When the SA service was launched, it was announced not only within Japan but also globally, resulting in nearly 700 registered users during the deployment period (Table 2). Among the three crowd accessibility services we deployed, SA had the largest number of contributors (98), with the largest number of contributors who contributed for more than two days (52).

The SA service was also able to stimulate a surge in contributor activity when an improved user interface was announced on November 27, 2008 (Fig. 5(a)). The surge was mainly due to experienced contributors who returned to the service as shown in Fig. 5(b). This highlights the important impact that usability enhancements can have in stimulating contributions.

While the SA service continued to yield contributions for over six months, the conversion from registered users to contributing users was relatively low compared with the other services, 14% (Table 2). From the contributor data, a majority of contributors contributed less than 10 days total. This suggests that the prolonged continuation of contributions may in large part have been due to a very small number of highly dedicated and productive contributors. This view seems to be supported by the contributor status chart (Fig. 5(a)) which shows that for the last two-thirds of the service period, the average number of contributors per day was between one and three.

The contributor chart also shows that during this period while the contributor count remained low, the number of registered visitors was continuing to increase. What this

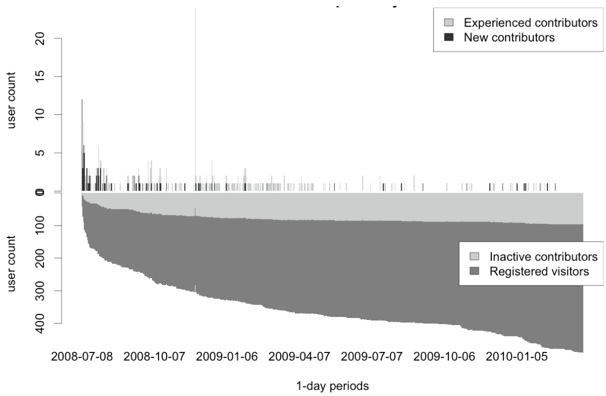


Fig. 5. (a). Contributor Status Chart for SA (2008/7/8 ~ 2010/3/29)

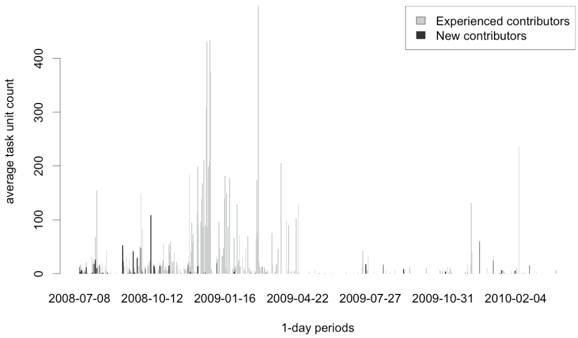


Fig. 5. (b). Aggregate Task Count Chart for SA (2008/7/8 ~ 2010/3/29)

Table 2. Contributor Status Transition Table

	Registered visitors	Registration-contribution ratio	Contributors (at least once)	Returnee ratio	Experienced contributors
EBIS	340	21%	71	41%	29
CCES	209	56%	116	52%	60
Social Accessibility	690	14%	98	53%	52

suggests is that new people were registering for the service, but were not converting into contributors. This represents missed opportunities, and may point to some limitations of the service that were turning away new contributors or inhibiting them from taking the first step.

6 Lessons Learned

6.1 Importance of Task and User Interface Design

One of success factors we found was the ease of completing tasks. One example is in CCES. One of the reasons for the high registration-contribution ratio (56%, Table 2) compared with other services may be due to the easy to understand task and productive user interface. The task goal of "captioning" may be familiar for people from seeing captioned TV programs. In addition, the interface makes it easy for a contributor to repeatedly listen to a short clip of the video and to type in the text, and provides convenient support for pausing and briefly rewinding the playback as the user types without requiring more explicit control by the user.

EBIS seems to have a similar situation. The registration-contribution ratio is lower than CCES but some of that is probably due to the periodic task shortages resulting from the task analysis (Fig. 3(b)). The success of the design can be seen from the rapid and large quantity of task completions, as mentioned in Section 5.1. Among the concerns expressed by the contributors, one of their highest priorities was "a rapid response from the server". This highlights the importance of having a streamlined user interface and a responsive system to fully benefit from the potential of large numbers of human participants.

In SA, several factors might cause the low registration-contribution ratio (14%, Table 2), such as shortages of user requests, a side effect of its being a public service (while the other services are all private), and the long term of its deployment. Another factor may be the steeper learning curve compared with CCES and EBIS, since the metadata required by the SA service was quite specific to the Web accessibility domain, and good results depended upon some knowledge of concepts such as heading navigation, alternative text, etc. Among the tasks in the services, character correction (EBIS) and captioning (CCES) are relatively simple tasks and do not require unusual expertise, which made the design of their tasks and their user interfaces easier than SA.

6.2 Necessity of Task Shortage Management

The importance of the management of tasks, especially for handling task shortages, was one of most important lessons we learned. SA used a request-answer concept as the basic crowd-sourcing model, but that means the model depends on the number of requests arriving from people with disabilities. That made the system especially vulnerable to task shortages.

EBIS experienced a task shortage (Section 5.1), but the cause was different from the SA situation. The process for quality assessment (Section 3.1) became the bottleneck of this service, and resulted in task shortages. The long gaps between short spurts of contributions were due to the fact that the contributors would quickly finish processing a batch of books in a period of a few days, but the next batch could not be made available to the contributors until the post-processing had been completed for those books. This post-processing step involved a single proofreader who carefully read through the entire book. Such proofreading can only be performed by skilled contributors when high quality is a concern, and this step could not be eliminated because the partner organizations wanted high-quality results. The gap between the crowd-handled work (with EBIS) and the proofreading was large. There are other alternative methods for such quality assurance phases, but no perfect alternative exists. Majority voting is a popular method but it sometimes requires large number of contributors if the quality criteria are high.

In contrast, CCES was managed in a better way. Right from the launch of the service, the service maintainers carefully monitored the consumption rate of the content by the contributors. For instance, during the first few days after deployment, the maintainers noticed that English content was receiving relatively little attention, and decided to augment the input data feed with more Japanese content during the first content update on the first weekend (November 18). The task count on the following day (November 19) more than doubled, apparently due to the increased availability of new content.

In summary, the lessons we learned fit into four areas: (1) tasks should be designed not to cause task shortages, (2) bottleneck phases should be minimized, (3) task shortages should be carefully monitored, and (4) methods for responding to task shortages should be predefined.

6.3 Engagement with Continuous Contributors

We learned that the top contributors played an important role across the services. In CCES and SA, they contributed to the services over many days, and achieved some of the highest average task completions per day. While it may not be prudent to depend on the emergence of such top contributors in every service, it is better for service owners to prepare for and take advantage of such contributors when they do appear. Identification is the first step to working with top contributors. The next step is to reach out to these top contributors. The service owners can derive various insights from the top contributors for improving their service. It may be possible to receive quite specific feedback regarding the service from the perspective of an experienced contributor that could be used to help improve the service for the rest of the contributors. They may also have developed various methods or strategies for approaching the tasks that may benefit other contributors as well.

For example, during the deployment of the SA service, we were able to meet with one of the top contributors by chance. We were able to extract various insights from him. He was much older than we had expected, actually one of the early baby boomers. He was a retired office worker, had lots of spare time and was trying to find something meaningful for his life. SA fit well with his current objectives. His motivations for participating in the service were very personal. The service owner team could clearly understand a role model contributor.

7 Discussion of Future Research Directions

7.1 Mechanisms to Propagate the Sense of Contributions

One of the most committed contributors to SA was neither an engineer nor a Web professional, but he had passion and time for volunteer work (Section 6.3). Since SA was a request-based service, he commented that he could directly understand the users' problems and the value of his work. His comments suggest the importance of the special sense of altruistic contribution as an incentive in accessibility services. Crowd accessibility services are characterized by the existence of human beneficiaries, the people with disabilities who are always behind the system. Because of this characteristic, we learned that a personal sense of contribution to those who need help can be an important incentive factor. In the services we studied, CCES notified the contributors when a video was posted with their captions, EBIS notified them when a book they had worked on was uploaded to the digital library, and SA had a function to notify a contributor "when someone used the metadata you created" by using a blinking icon [18]. One of the topics we have not explored was the cultural effects of the propagation of the sense of contributions. Most of the contributors for these three services were Japanese, and therefore cultural difference among countries could not be examined in this study. We suspect that the sense of altruism may differ among cultures, so further research activities in this area would be important.

7.2 Mechanisms for Accessibility Skill Development

Accessibility improvements often require specific skills for contribution. SA required basic knowledge of Web accessibility, such as appropriate alternative texts according to contexts and appropriate heading levels, and so the task familiarity for contributors was low (Section 6.1). The tasks for CCES and EBIS were more familiar than the SA work but still required specific knowledge that affected the output quality. For example, the notation of sound effects in captions, and techniques to find appropriate kanji characters from character tables. These tasks require knowledge and skill to extend them to environments for people with disabilities. Section 6.1 discussed the importance of easy-to-use interfaces and easy-to-understand tasks, but systematic mechanisms to develop "accessibility skills" in contributors offer another big area to explore.

7.3 Challenges of User-Side Service-Quality Assurance

One of the specific challenges for crowd accessibility services is the difficulty of quality assessment on the user side. SA relied on user reporting as its most important quality assurance method (Section 3.3). User reporting is a popular method used by many social services in such forms as the "illegal content report" of most video sharing services. However, for crowd accessibility services this can be difficult to implement. For SA the quality of an alternative text can be assessed by some straightforward criteria (such as spelling mistakes, lack of clarity, etc.), but this assessment is not possible for many kinds of problems. For example, if a malicious volunteer added an alternative text such as "Cancel" to an OK button, it would be extremely difficult to detect what was wrong by pushing the button. The end users need support because of their limited cognitive abilities, and that also limits their ability to assess the quality. This is one of the characteristic challenges in some types of crowd accessibility services.

7.4 Influence of Unintended Fun Factors for Increasing Engagement

We saw cases in which the unintended fun factor contributed to motivate contributors in serious accessibility services. In the CCES, the video clips were extracted from longer and sometimes enjoyable videos about substantive topics. From the comments in the survey, contributors found themselves becoming interested in the content itself as they performed the transcription tasks. A similar unintended fun factor also worked in EBIS, since in the final check the contributors read through an entire book and they could enjoy it. We found that the sense of contribution was the primary motivation for the contributors to accessibility services, but the unintended fun factor strengthened engagement with each service, thus increasing the participation. Compared to the intentional fun factors like gamification, unintended fun can easily be overlooked. The influence of unintended fun for serious accessibility services is worth investigating in future research.

7.5 Possibilities for Senior Citizens as Contributors

Another implication is the importance of encouraging potential contributors. For the leading contributor mentioned earlier, we had not expected such prominent involvement by a senior citizen. This anecdotal evidence suggests that senior citizens may be especially good candidates as volunteers for crowd accessibility services. Based on this experience, we studied several crowd-sourcing tasks including accessibility work for seniors ranging in age from 60 to 80. Half of them answered that they would be willing to do such tasks "without any fee", in contrast to a smaller percentage of younger participants who agreed to the no-fee model.

7.6 Possibilities for Internal Organizational Crowd-Sourcing

Two out of the three services were deployed within enterprise environments. (See Section 3 and Table 1.) These environments were not open to the public, so they can be regarded as constrained environments. However, our experiments showed some advantages in deploying crowd applications within an enterprise. The first advantage is the potentially large number of contributors in a large corporation, making the deployment similar to a general public environment, but with better control. We can announce, maintain, and analyze the results by using same framework as used for public services. Second, contributors within a corporation have a strong rationale not to act maliciously or to submit low-quality work in their enterprise environment partly because of their lack of anonymity. Such activities can be tracked, which is a strong incentive for the contributors to strive for high quality work. The third point is that services will be more sustainable due to the official support from the corporation as part of its CSR (Corporate Social Responsibility) efforts.

8 Conclusion

This paper discussed some lessons learned through an analysis of three crowd-accessibility services. After reviewing the related work, we developed visualization techniques that reflect the status changes of the contributors and of the completed tasks. The lessons learned were considered in the observation section by highlighting four topics, (1) design of the tasks and the user interface, (2) task shortage management, (3) contributor engagements, and (4) data collection and monitoring methods. Beyond these lessons learned, challenges and future research directions were discussed. The sense of contribution can be an important incentive especially for accessibility services. Better mechanisms for developing accessibility skills are imperative to improve the quality of the services. User side quality assurance poses unique challenges for accessibility services. The influence of unintended fun factor is worth investigation to increase engagement with the services. Also, senior citizens and corporate workers can be untapped new contributing resources.

Crowd accessibility has a huge potential to improve many situations for people with disabilities by combining human intelligence and computer intelligence. We are studying various ways to make the approach more sustainable in our society. A standard analysis method is itself challenging and methods should be evolved to unveil more specific problems within services. We hope that this paper will contribute to improving crowd accessibility strategies toward the level of more reliable and sustainable services.

References

1. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2004, pp. 319–326 (2004)
2. Benetech. Accessible books and periodicals for readers with print disabilities - Bookshare - accessible books for individuals with print disabilities, http://www.bookshare.org/
3. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: VizWiz: nearly real-time answers to visual questions. In: Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology, UIST 2010, pp. 333–342 (2010)
4. Bigham, J.P., Kaminsky, R.S., Ladner, R.E., Danielsson, O.M., Hempton, G.L.: WebInSight: making Web images accessible. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2006, pp. 181–188 (2006)
5. Bigham, J.P., Ladner, R.E., Borodin, Y.: The design of human-powered access technology. In: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2011, pp. 3–10 (2011)
6. Bryant, S.L., Forte, A., Bruckman, A.: Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In: Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2005, pp. 1–10 (2005)
7. Dardailler, D.: The ALT-server ("An eye for an alt") (1997), http://www.w3.org/WAI/altserv.htm
8. Ishihara, T., Itoko, T., Sato, D., Tzadok, A., Takagi, H.: Transforming Japanese archives into accessible digital books. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2012, pp. 91–100 (2012)

9. IQ Engines. IQ Engines: image recognition and visual search, http://www.iqengines.com/
10. Nielsen, J.: Zipf distribution (power law) of Website popularity (Alertbox Sidebar) (1997), http://www.useit.com/alertbox/zipf.html
11. Nielsen, J.: 90-9-1" rule for participation inequality: lurkers vs. contributors in internet communities (Jakob Nielsen's Alertbox) (2006), http://www.useit.com/alertbox/participation_inequality.html
12. Sobhi, A., Nagatsuma, R., Saitoh, T.: Collaborative Caption Editing System – Enhancing the Quality of a Captioning and Editing System. In: Proceedings of the 28th Annual International Technology and Persons with Disabilities Conference, CSUN 2012 (2012)
13. Nonnecke, B., Preece, J.: Lurker demographics: counting the silent. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 73–80 (2000)
14. Nonnecke, B., Preece, J.: Why lurkers lurk. In: Proceedings of the Americas Conference on Information Systems (2001)
15. Stewart, O., Lubensky, D., Huerta, J.M.: Crowdsourcing participation inequality: a SCOUT model for the enterprise domain. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP 2010, pp. 30–33 (2010)
16. Swartz, A.: Who writes Wikipedia? (2006), http://www.aaronsw.com/weblog/whowriteswikipedia
17. Takagi, H., Kawanaka, S., Kobayashi, M., Itoh, T., Asakawa, C.: Social Accessibility: achieving accessibility through collaborative metadata authoring. In: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2008, pp. 193–200 (2008)
18. Takagi, H., Kawanaka, S., Kobayashi, M., Sato, D., Asakawa, C.: Collaborative Web accessibility improvement: challenges and possibilities. In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2009, pp. 195–202 (2009)
19. Valerie, C., Harrison, A.: Bookshare crosses 100,000-member milestone - Bookshare - accessible books for individuals with print disabilities (2010), http://www.bookshare.org/_/aboutUs/2010/07/100KmemberMilestone
20. Wikipedia. Wikipedia:Wikipedians - Wikipedia, the free encyclopedia (2011), http://en.wikipedia.org/wiki/Wikipedia:Wikipedians

Designing for Different Users and Multiple Devices: A Roadmap towards Inclusive Environments

Ig Ibert Bittencourt[1,2], Maria Cecília Baranauskas[2], Diego Dermeval[3], Juliana Braga[4], and Roberto Pereira[2]

[1] Center of Excellence in Social Technologies, Computing Institute, Federal University of Alagoas, Brazil
ig.ibert@ic.ufal.br

[2] Computing Institute, University of Campinas, São Paulo, Brazil
{cecilia,rpereira}@ic.unicamp.br

[3] Federal University of Campina Grande, Paraíba, Brazil
diegodermeval@dsc.ufcg.edu.br

[4] Federal University of ABC, São Paulo, Brazil
juliana.braga@ufabc.edu.br

Abstract. The Web can be understood as an ecosystem of interconnected technological resources organized by rules, strategies, organizational structures, and mainly people. Such ecosystem is improving the potential to access knowledge everywhere and at any time. However, for different reasons, this technological extension is not reaching everybody yet. Those without access to knowledge are mainly people with disabilities or living in underserved communities. Nevertheless, the extension of Web technologies to different types of devices (such as mobile phones, tablets, TV) and their connections have the potential to increase the solutions to reach people with different needs through different devices. For this reason, several research and industrial studies have been proposed to design interfaces for multiple devices considering differences among users. In this paper, we present results of a systematic review on literature to build a roadmap towards inclusive environments. Additionally, the study also suggests existing tools to support the design of accessible applications for multiple devices. A significant result of this review is the lack of studies addressing underserved communities.

Keywords: Inclusive Environments, Portability, User Interface Design, Inclusiveness, Diversity, Multiple Devices.

1 Introduction

Several organizations around the world have undertaken work programs to establish electronic communication via Web technologies. The Web can be understood as an ecosystem of interconnected technological resources organized by rules, strategies, organizational structures, and mainly people. As a result, this ecosystem is improving the potential of access to knowledge everywhere and at any time, and is becoming a way to tackle the challenge of providing a participative and universal access to

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 605–622, 2013.

knowledge. Universal access has been considered one of the great challenges of several International Communities around the world [1] [2]. This challenge is about the use of technologies to ensure access to knowledge in a participative and personalized way for the citizen, taking into account the diversity of people and, consequently, different[1] users' needs encompassing disability issues as well as social problems (e.g., people living in underserved[2] communities).

The accessibility challenge regards ensuring access to information and functionalities to all potential users of technology. As a matter of fact, a valuable body of research and best practices has been developed to address Web accessibility. However, the challenges of Web accessibility have increased substantially due both to the extension of Web technologies to different kind of devices (e.g., mobile phones, tablets, TV) and to their possibility of interactions. On one hand, the Web movement beyond desktop to different devices increased the challenges of accessibility. On the other hand, this movement amplified the possibilities to ensure access to information independently of the place or the knowledge domain. For this reason, many research and industrial studies have been proposed to design interfaces for multiple devices and different users — such as [2] [3] [4] —; however, no systematic review has been conducted to provide an overview of the design of inclusive environments for different users and multiple devices.

The goal of this paper is to present results of a systematic review about interface design approaches addressing multiple devices and different users, and, at the same time, to identify the tools available to build accessible applications aiming to reach people with different disabilities and/or living in underserved communities.

This paper is organized as follows: Section 2 describes the research method of the study; Section 3 gives an overview of the studies included in the work; Section 4 presents the main results of this systematic review with a discussion on the reported results. Finally, Section 5 summarizes this work, presents our conclusions and points out future works.

2 The Method

The research in this work was undertaken as a systematic literature review (SLR) to provide a repeatable and formal process to document relevant papers about portability on inclusive social web. As a result of identifying, interpreting and evaluating their data, it is possible to find evidence on which to base conclusions according to the research questions. The conclusions are commonly used to support or contradict claims made by researchers, identify gaps in existing research, provide motivation for new research, and supply a context for the new research [5].

---

[1] In this paper, we consider the term "different users" or "different conditions" as the users with disabilities (sensorial impairment, motor impairment, and cognitive impairment), social problems (underserved or illiterate people) and ageing (elderly).

[2] In this paper, we consider the term underserved as the people or communities living in areas without good access to technology and/or internet, such as rural areas and low income areas in the periphery of large cities.

According to [5], a systematic review is composed by three phases (planning, conducting, and reporting) divided into several steps, which are: 1) Planning the Review (Identification of the need for a systematic review; Development of a review protocol); 2) Conducting the Review (Identification of research; Selection of the studies; Study Quality assessment; Data extraction and monitoring; Data analysis; Data synthesis); 3) Reporting the review (Report-writing).

2.1 Review Questions

The goals of this systematic review were: i) to summarize the research in this topic; and ii) to present a roadmap towards inclusive environments. As the goal of this systematic review was to gather knowledge about the design of inclusive environments for different users and multiple devices, the high-level question of this study was:

How researchers are designing applications for different users and multiple devices?

Based on this research question, other two more specific questions were raised. The questions and their motivations are described in Table 1.

Table 1. Research Questions and Motivations

Research Question	Motivation
RQ1. Which interface design approaches are being used to address multiple devices and users with different conditions?	This question provides a starting point to understand how designers conduct the interface specification for multiple devices. The answer to this question is important to understand how people with different conditions are being considered in the interface design process.
RQ2. Which tools are being used to support the design of inclusive applications for multiple devices?	This question presents the different solutions (i.e., software, frameworks, authoring tools, architectures, and so on) to build portable applications for all. The answer to this question is important to identify the different solutions regarding the adopted software engineering practice.

2.2 Sources and Search Selection Criteria

The first step to perform our review was to define the search selection criteria. Due to the fact that this review has several sources to consider, two kinds of search strategies were considered (based on [6]), which were automatic and manual search. Automatic search was done according to the specification of the search terms (i.e., search string). Although automatic search covers a large range of relevant papers, it is also important to search in specific and specialized sources to improve the coverage. For this reason, a manual search in some of the most important conferences and journals of the Human-Computer Interaction area was taken into account.

Hereafter, the search terms definition and the digital libraries (DLs) selection regarding the automatic search are explained. According to the research questions aforementioned, a set of relevant terms was defined, such as: *cross-device, disabilities, underserved communities, design interfaces, approaches, multiple devices, different users, tools,* and *inclusive*. After that, these terms were categorized and their related terms were identified. The terms were identified based on: i) expertise of the authors; ii) analysis of terms present in a HCI systematic review available at [7]; iii) *TagCloud* for HCI presented in Fig. 1 and iv) suggested topics for contributions to the *Interact 2013 Conference*.

Fig. 1. TagCloud for HCI [8]

The related terms for the main words are described as follows:

- **Tool** = environment, framework, authoring, architecture, software, ambient, "reference model";
- **Disabilities** = Inclusive, inclusiveness, inclusivity, accessibility, disability, assistive, underserved, "marginalized communities", "design for all", "universal access", elderly, "older adults", diversity;
- **Multiple Device** = "cross-device", multimodal, migration, "different devices", "device-independent", "migratory interfaces", "distributed interfaces", "plastic user interfaces", "flexible user interfaces", "flexible interfaces", "distributed user interfaces", portability, "portable web applications", "portable systems", interoperability;
- **Interaction design** = "adaptable interface", "adaptable user interfaces", "interaction resources", "responsive web design", "universal design", "inclusive design", "process model", "adaptable model", "system design", "meta-design".

By contrast, the set of digital libraries was defined according to the most popular and traditional DLs. However, the SpringerLink digital library was excluded due to search restrictions and its intersection with others DLs. The selected DLs were:

- ISI Web of Science (http://www.isiknowledge.com);
- Scopus (www.scopus.com/scopus/home.url);

- ACM Digital Library (http://portal.acm.org);
- IEEE Xplore (http://www.ieee.org/web/publications/xplore/);
- ScienceDirect (http://www.sciencedirect.com).

After the definition of relevant terms and DLs, the search string for automatic search on the mentioned digital libraries was built as follows:

```
((tool OR environment OR framework OR authoring OR architecture OR soft-
ware OR ambient OR "reference model")
AND
(inclusive OR inclusiveness OR inclusivity OR "inclusive web" OR "inclu-
sive social web" OR accessibility OR disability OR disabilities OR as-
sistive OR underserved OR "marginalized communities" OR "design for all"
OR "universal access" OR "universal design" OR "designing for diversity"
OR "design for diversity" OR "design diversity" OR diversity)
  AND
("Multiple Device" OR "Cross-device" OR "Multimodal" OR migration OR
"different devices" OR "device-independent" OR "migratory interfaces" OR
"distributed interfaces" OR "plastic user interfaces" OR "flexible user
interfaces" OR "flexible interfaces" OR "distributed user interfaces" OR
portability OR "portable web applications" OR "portable systems" OR
"information interoperability" OR "knowledge interoperability")
  AND
("Interaction design" OR "adaptable interface" OR "adaptable user inter-
faces" OR "interaction resources" OR "responsive web design" OR "univer-
sal design" OR "inclusive design" OR "process model" OR "adaptable
model" OR "meta-design" OR "meta design" OR metadesign OR "participative
design"))
```

Moreover, in order to perform the manual search, two relevant conferences and journals in the Human-Computer Interaction area were considered (see Table 2). It is important to note that other Journals and Conferences could have been considered for manual search, but the Journals were chosen based on their Impact Factor. In addition, we could not have access to the library of some journals, such as the International Journal of Human-Computer Interaction. With regards to conferences, we initially added the ACM SIGCHI Conference on Human Factor in Computing Systems, but the list of published articles in this conference would represent half of the whole search space, unbalancing the study.

Table 2. Relevant journals and conferences on the Human-Computer Interaction area

Journals
1. International Journal of Human Computer Studies
2. Interacting with Computers
Conferences
1. IFIP INTERACT
2. Cross-Disciplinary Conference on Web Accessibility

2.3 Inclusion and Exclusion Criteria

The aim of defining a criterion is to really identify those primary studies that provide direct evidence about the research questions and also to reduce the likelihood of bias [5]. Regarding the inclusion criteria, articles written in the last ten years related to any of the research questions were considered. The exclusion criteria involve papers not related to the research questions, papers which were not in English, short papers[3], duplicate studies and papers before 2002[4]. The summarized inclusion and exclusion criteria are presented in Table 3.

Table 3. Inclusion and exclusion criteria

Inclusion criteria
Peer-reviewed studies that provided answers to the research questions
Studies that focus on design approaches for multiple devices
Studies that focus on design approaches for different users' needs
Studies that focus on assistive technologies for different users' needs
Studies published since 2002
Exclusion criteria
Short-papers
Non peer-reviewed studies
Studies that are not related to research questions
Duplicated studies
Studies before 2002
Non English papers

2.4 Data Extraction

After the definition of the search and the selection processes, a data extraction process was performed by reading the abstract and screening the full-text of each one of the selected papers. In order to guide this data extraction, the data collection from Biolchini *et al.* [9] was adapted as follows:

- **Paper Information**: Study Reference (ID); Source; Year; Source Type (Journal or Conference); Affiliations; Authors list; Paper Title; Google Scholar Citation;
- **Context:** (Industry and Academia);
- **Device Types:** (Desktop, Web, Tablet, TV, Mobile Phones, PDA, Tabletop, Braille Notes);
- **Target Audience:** (Blind/Visual Impairment; Deaf/Hearing Impairment; Motor/ Mental; Underserved people; Elder);

[3] Short-papers were excluded because they usually represent ongoing research.

[4] Due to the rapid evolvement of web technologies and hardware devices, the older inclusive solutions for multiple devices lack the benefits and potential of current devices. For this reason, we decided to limit our review to the last decade.

- **Design Interface Approach:** (User-Centered; Task-Centered; Participatory; Scenario-Based; Ethnographic Methods; Design per Target; Model-based; Automatically Generated; Multi-tier; Universal Design; User Sensitive Inclusive Design);
- **Use or propose any tool?** (Yes; No);
- **Tool Type:** (API; Design Pattern; Framework; Platform; Software Product Line; Authoring; MDA; Reference Model; Middleware; Architecture);
- **Study Type:** (Controlled experiments; Quasi-experiments; Case Study; Survey; Ethnography; Action Research);
- **Subjective results extraction**.

3 Overview of the Included Studies

This section presents the included studies according to the automatic and manual search (see Fig. 2). Firstly, the automatic search was conducted at each digital library. Then, an iterative process was applied to exclude the not relevant papers based on the exclusion criteria. The exclusion criteria were applied according to the analysis of the abstract, full-text screening and finally the duplicate papers. In a similar way, the process was applied to the manual search.

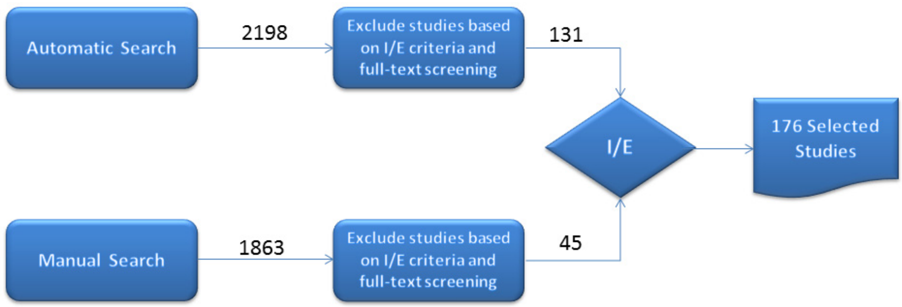


Fig. 2. Search process and selected studies

The automatic and manual query was conducted in the period between December 4th (2012) and January 11th (2013) and the data was extracted by two people. The results per each digital library, conference and journal are shown in Fig. 3.

Fig. 3 (Manual search) presents that the IFIP Interact (31%; 14 studies) and the W4A – Cross-disciplinary Conference on Web Accessibility (27%; 12 studies) – were the largest vehicles of relevant studies about portability on inclusive web. However, other vehicles were identified as relevant when the automatic search was applied, such as the ACM SIGCHI Conference on Human Factors in Computing Systems (6.25%; 8 studies), the ACM SIGACCESS conference on Computers and accessibility (4.69%, 6 studies) – and the ICCHP –International Conference on Computers Helping People With Special Needs (4.69%; 6 studies).

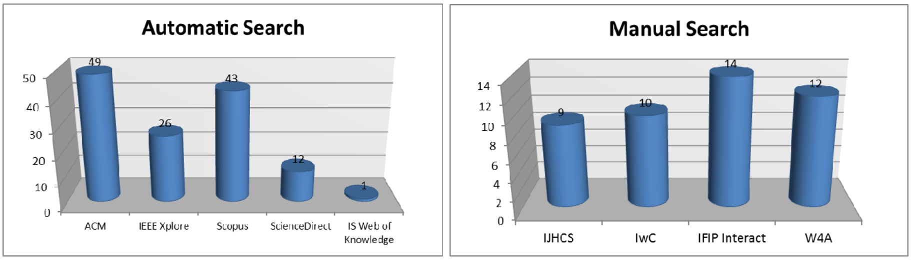


Fig. 3. Automatic and Manual results

From a temporal point of view, it is noticed in the trend curve an increasing number of publications in the context of this review since 2003 (see Fig. 4). It is important to note that there are no works from the year 2002. The first study was published only in 2003. By contrast, there is a decrease in the number of publications in some years (2008, 2009, and 2012).

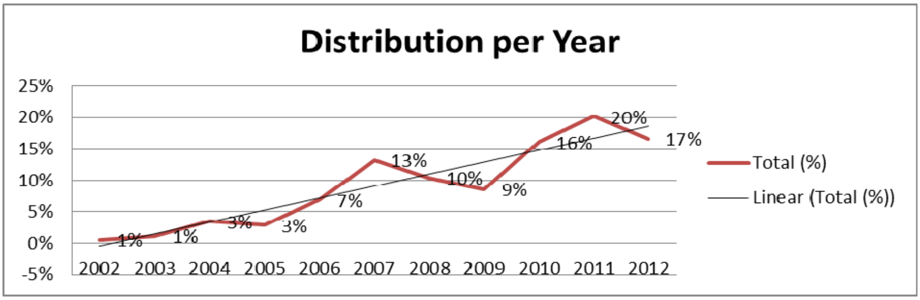


Fig. 4. Distribution of publications per year

Since this review began on December 2012, a possible cause of this decreasing in 2012 is that some papers were still under publication process, thus it is natural that some papers were not available online yet. Therefore, we can see, in general, an increase in the number of publications (based on their linear progression). The significant increase of publications on portability in inclusive web reflects the need for convergence of technologies and, at the same time, the importance of deploying inclusive solutions.

The study holds contributions from 35 countries located in all the continents. Although all the continents are represented in the included papers, there is a concentration in the Americas (26.34%) and Europe (62.44%). Table 4 presents the publications per authors country. Furthermore, according to the distribution of the included papers, most of the related papers were published in conferences (66%) instead of journals (34%), considering both the automatic and manual search.

Table 4. Publications per country

Country	Works	Total (%)
United States of America	29	14.15%
United Kingdom	27	13.17%
Brazil	16	7.80%
Spain	15	7.32%
Portugal	11	5.37%
Belgium	10	4.88%
Italy	10	4.88%
France	9	4.39%
Germany	9	4.39%
Greece	7	3.41%
Japan	7	3.41%
Austria	6	2.93%
Finland	6	2.93%
Canada	5	2.44%
Australia	4	1.95%
Sweden	4	1.95%
South Korea	3	1.46%
Norway	3	1.46%
South Africa	3	1.46%
Others	21	10.34%

4 Results

As described in Section 2, three research questions drove this systematic review. Based on the research questions, the string search was built and the type of data extraction defined. According to the data extraction, the addressed conditions of most users were Blind or Visual Impairment (24%; 58 studies), Mental (14%; 34 studies) and Motor (14%; 34 studies). One of the important points about this information is that although the number of deaf people or with hearing impairment is relatively high in the general population, the number of studies addressing this disability in the review was very low (see Fig. 5).

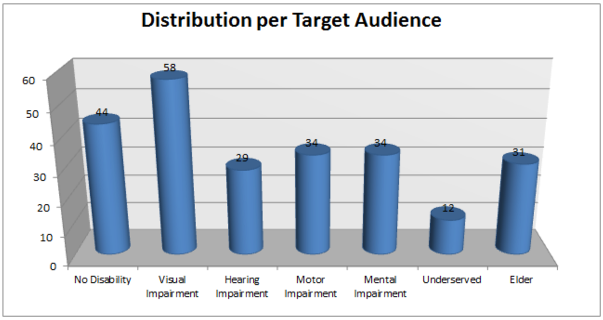


Fig. 5. Distribution of publications per target audience

The following subsections present and discuss the results for the research questions. Moreover, a specific subsection discusses the works that address underserved communities, due to the reduced amount of papers identified for this audience in the study. In the end of this section, threats to the validity of this review are also discussed.

4.1 Inclusive Applications for Multiple Devices

Table 5 presents the distribution of papers according to device type. Most of the solutions were proposed for mobile phones (29.84%; 94 studies) and Web (18.73%; 59 studies). It is worth noting that the number of solutions for tablets is still small (8.25%; 26 studies); this may happen especially because tablets have started to become popular more recently than mobile phones. Nevertheless, an increasing in their use is expected since they have the potential to provide access to underserved communities. Although there are 29 studies in other categories of Device Types, 21 studies (6.67%) are not related or did not define the device.

Table 5. Distribution per device type

Device type	Works	Total (%)
Mobile Phones	94	29.84%
Web	59	18.73%
Desktop	49	15.56%
PDA	44	13.97%
Tablet	26	8.25%
TV	14	4.44%
Others	29	9.21%

Table 6 presents the distribution per tool type. Frameworks and platforms are in the top of the list, while the Software Product Line (SPL) approach is not mentioned, suggesting a gap between the Human Computer Interaction and the Software

Table 6. Distribution per tool type

Tool type	Works	Total (%)
Framework	29	16.48%
Application	24	13.64%
Reference Model	21	11.93%
Platform	20	11.36%
Architecture	7	3.98%
Authoring	4	2.27%
MDA	3	1.70%
Design Pattern	2	1.14%
API	2	1.14%
Hardware	2	1.14%
Middleware	1	0.57%

Engineering issues. SPL is one of the most sophisticated concepts in the Software Engineering area with regard to software reuse and flexibility, since it allows a strategic reuse that provides a platform which can be easily adapted according to users' specific needs in a specific domain [10].

In addition, it can be noticed that half of the papers that presented tools provide an inclusive solution (50%). However, only 15% of the selected papers proposed inclusive tools for multiple devices, as illustrated in Fig. 6. It is important to observe that by inclusive tools we mean support tools to generate inclusive applications. Additionally, most of the inclusive tools for multiple devices are Frameworks (37%) and Platforms (32%).

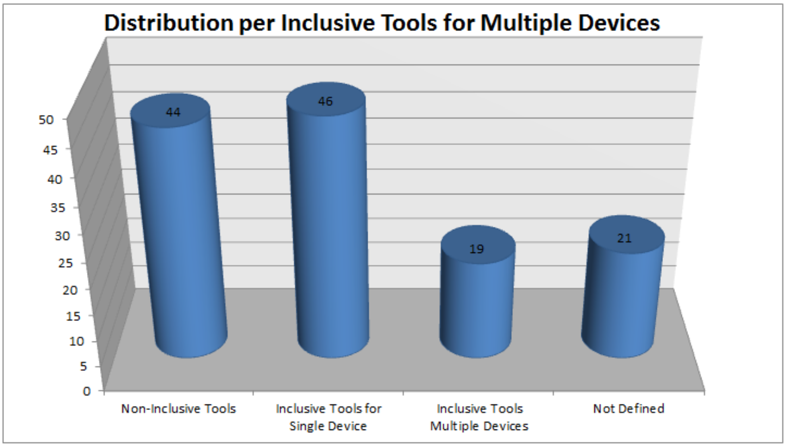


Fig. 6. Graphic about inclusive tools for multiple devices

The list of papers that propose inclusive tools for multiple devices is presented in Table 7[5] (papers which did not define the tool type were not included).

Table 7. List of support tools to generate inclusive tools for multiple devices

Study Ref.	Paper Title	Technology Type
SACM19	Distributed Intelligence: Extending the Power of the Unaided, Individual Human Mind	Platform
SACM20	Towards Ubiquitous Accessibility: Capability-based Profiles and Adaptations, Delivered via the Semantic Web	Platform
SACM27	MyUI: Generating Accessible User Interfaces from Multimodal Design Patterns	Framework
SACM62	The Potential of Adaptive Interfaces as an Accessibility Aid for Older Web Users	Platform
SACM68	Accessibility of Dynamic Adaptive Web TV Applications	Framework

[5] The list of the papers is available at `http://www.nees.com.br/interact`

Table 7. (*Continued*)

Study Ref.	Paper Title	Technology Type
SACM70	Design, Adoption, and Assessment of a Socio-Technical Environment Supporting Independence for Persons with Cognitive Disabilities	Platform
SACM75	Agent-Based Architecture for Implementing Multimodal Learning Environments for Visually Impaired Children	Framework
SIEEE07	A-CitizenMobile: A Case Study for Blind Users	Framework
SIEEE10	An Open Architecture to Develop a Handheld Device for Helping Visually Impaired People	API
SIEEE11	A Framework for Designing Flexible Systems	Framework
SIEEE30	i*Chameleon: A Unified Web Service Framework for Integrating Multimodal Interaction Devices	Framework
SSD13	Automatically generating personalized user interfaces with Supple	Platform
SSCOPUS25	Rapid Prototyping of Adaptable User Interfaces	Platform
SSCOPUS 38	A Novel Design Approach for: Multi-device Adaptable User Interfaces: Concepts, Methods and Examples	Framework
SSCOPUS 40	The Contribution of Multimodal Adaptation Techniques to the GUIDE Interface	Platform
SSCOPUS 69	Contributions of Dichotomic View of plasticity to seamlessly embed accessibility and adaptivity support in user interfaces	Architecture
SSCOPUS 80	Assistive smartphone for people with special needs : The Personal Social Assistant	Application
SSCOPUS 85	Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice	Application

Fig. 7 presents the type of empirical study applied to evaluate the tools. More than 50% of the works applied evaluation by an empirical case study. In addition, only 25% of the works applied some kind of experiment. By contrast, survey studies were 5%.

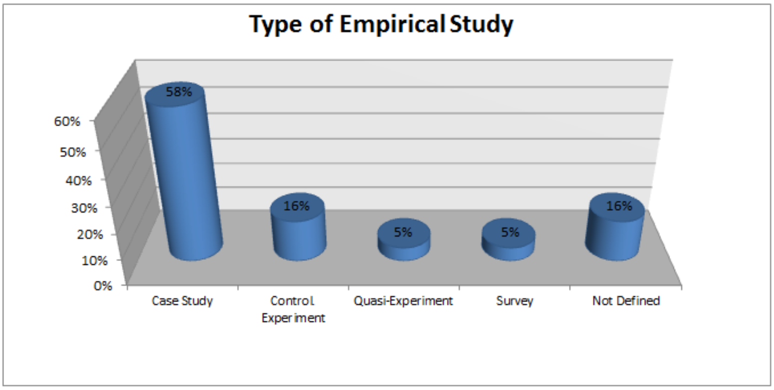


Fig. 7. Type of Empirical Study of the Inclusive Tools for Multiple Devices

It is also important to identify which of these studies address the design for diversity. For this reason, Table 8 presents them grouped by tools for one target audience and for diversity. As a result, only 4.54% (6 studies) of the inclusive tools for multiple devices (see Fig. 6) address inclusive tools for different users, while 7.63% (10 studies) address tools for one target device.

Table 8. List of inclusive studies for multiple devices based on the target audience

Technology Type	For one Target Audience	For Diversity
Framework	SACM27; SACM75; SIEEE07	SACM68; SIEEE11; SIEEE30; SScopus38
Platform	SACM19; SACM20; SACM62; SACM70; SSD13	
Application	-	SScopus80; SScopus85
API	SIEEE10	-
Architecture	SScopus69	-

4.2 Interface Design Approaches for Diversity

The distribution of the included studies according to the Interface Design approaches (Table 9) shows that more than 50% of the studies apply User-centered, participatory or automatically generated approach. However, more than 15 interface design approaches were identified, which turned the data classification, organization, and analysis much sparse.

Table 9. Distribution per Interface Design Approach

ID Approach	Works	Total (%)
User-Centered	31	19.62%
Automatically Generated	27	17.09%
Participatory	25	15.82%
Model-Based	17	10.76%
Task-Centered	15	9.49%
User-Sensitive Inclusive Design	10	6.33%
Design Per Target Device	9	5.70%
Ethnographic	9	5.70%
Scenario-Based	7	4.43%
Others	8	5.06%

Moreover, as presented in Table 10, only few studies (13.79%; 20 studies) present a design approach considering user diversity in the application development.

Conversely, although there are 19 studies addressing Interface Design approaches for diversity, only 3 (three) studies also addressed multiple devices, as depicted in Table 11. Among these studies, two of them apply automatic interface generation and

Table 10. Distribution per focus on diversity

ID Approach	No Disability	For one Target Audience	For Diversity
Participatory	4.83%	6.90%	4.14%
Automatically Generated	11.03%	4.83%	3.45%
User-Centered	7.59%	8.97%	3.45%
Model-Based	9.66%	0.00%	2.07%
User-Sensitive Inclusive Design	0.00%	4.83%	0.69%
Design Per Target Device	4.14%	2.07%	0.00%
Task-Centered	6.21%	4.14%	0.00%
Scenario-Based	3.45%	2.07%	0.00%
Ethnographic	2.07%	3.45%	0.00%

the other two use participatory design. Besides, two studies evaluated their tools by some kind of experimental research, while the other two papers performed evaluation by an empirical case study.

Table 11. List of studies from Table 10 for Diversity and Multiple Devices

Study Ref.	Paper Title	ID Approach	Study Type
SACM68	Accessibility of Dynamic Adaptive Web TV Applications	Automatically Generated	Case Study
SIEEE11	A Framework for Designing Flexible Systems	Participatory	Case Study
SScopus85	Attuning speech-enabled interfaces to user and context for inclusive design: technology, methodology and practice	Participatory	Controlled Experiment

4.3 Solutions for Underserved Communities

This section discusses the studies with a focus on underserved communities. Indeed, only 12 studies addressed this topic which is also the less discussed target audience identified in this systematic review – representing only 5% (see Fig. 5). The full list of studies attending underserved communities is presented in Table 12.

It is worth noting that less than half of the underserved people studies involve researchers from countries where the researches were developed. Another important point to highlight is the reduced number of studies about underserved people coming from the HCI community, which means there is a need for cooperative work from different communities with the Human-Computer Interaction community.

As a roadmap regarding studies about underserved communities, most of the studies proposed the use of mobile devices as the best technology. The main reason seems to be that underserved communities have bad access to electricity, internet, and so on. Moreover, we realized that several studies with regards underserved communities refer also to illiterate people. In addition, the most common user interface type

Table 12. List of studies addressing underserved communities

Study Reference	Paper Title	Countries
SACM30	Comparing Semiliterate and Illiterate Users' Ability to Transition from Audio+Text to Text-Only Interaction	Canada and India
SACM41	Designing with Mobile Digital Storytelling in Rural Africa	Australia and South Africa
SACM52	Universal Accessibility As A Multimodal Design Issue	Spain and Serbia and Montenegro
SACM58	Cultural Coding and De-Coding as Ways of Participation: Digital Media for Marginalized Young People	USA and Belgium
SACM86	Technology-Supported Cross-Cultural Collaborative Learning in the Developing World	USA
SScopus35	Pushing personhood into place: Situating media in rural knowledge in Africa	South Africa, Namibia and Denmark
SIWC14	Designing new technologies for illiterate populations: A study in mobile phone interface design	United Kingdom
SW4A10	The Spoken Web Application Framework User Generated Content and Service Creation through lowend mobiles	India
SW4A11	Developing Countries; Developing Experiences: Approaches to Accessibility for the Real World	United Kingdom
SW4A12	Designing for Auditory Web Access: Accessibility and Cellphone Users	USA
SW4A13	Exploring Web Accessibility Solutions in Developing Regions as Innovations for the Benefit of All	France
SW4A14	Designing new technologies for illiterate populations: A study in mobile phone interface design	United Kingdom

was based on Voice or Web. Although Voice approach is commonly used, it does not solve the problem of illiteracy. Another important point is that only 2 studies proposed a software solution, while the others proposed a reference model or a new hardware. Finally, only four studies address the context of developing country.

4.4 Discussion and other Related Work

This section focuses on how the systematic review tackled the three research questions and discusses about the important conclusions obtained from the analysis of the papers included in this review. As a result of this review, it was possible to identify the interface design approaches and the tools according to specific user conditions and also to specific devices.

With regards to the RQ1 (*Which interface design approaches are being used to address multiple devices and users with different conditions?*), only a very limited number of papers (3 studies, as presented in Table 11) was identified from the whole list of studies (see Fig. 2), representing 1.70%. One important point is that the Interface design approaches were only the Automatically Generated Interface (from 2012) and the Participatory Design (from 2008 and 2011). This means that the studies were focused on ID approaches to tackle the first phases of software development lifecycle (before software implementation), while other part of the studies were focused on implementation, test, and deployment phases. Only one study raised (SIEEE11) the importance of user involvement during software evolution.

Concerning RQ2 (*Which tools are being used to support the design of inclusive applications for multiple devices?*), six studies presented solutions for the design of inclusive applications for multiple devices (see Table 8), representing 3.41% of the whole data extraction (see Fig. 2). According to the data, the types of technologies used to design inclusive applications for multiple devices were Platform (from 2007), the Framework (from 2011 and 2012) and the Application (from 2009). Thus, only three types of technologies were proposed. New strategic approaches for reuse and flexibility were not addressed, such as the Software Product Line. In comparison to the Software Engineering approaches used by the identified studies, the SPL may support higher reduction in software development costs, higher increase of software quality, faster time to market and higher reduction in maintenance efforts [10]. In this way, with these advantages, it is intriguing that this approach has never been used in the design of tools to support the development of such applications. Thus, this information could demonstrate some level of detachment between HCI and Software Engineering. In contrast, it may also demonstrate that the software engineering methodologies need to advance in order to take into account inclusiveness as a very important non-functional requirement. Through the analysis of the papers included in this review, it could also be noticed the interest of several studies to evolve Web technologies. Indeed, some of these studies have explored the Semantic Web and Ontologies technologies to automatize interface generation.

Furthermore, the results reported in this review also show that the underserved audience is still under addressed by inclusive design studies (see Table 12). By contrast, the number of studies that attends the elder audience is substantial (Fig. 5) – which reveals the concern of researchers with the older population. Considering the devices used in the studies, the large presence of mobile phones in these studies can be highlighted (see Table 5). It can also be noted a smaller amount of works focusing on Tablets (compared to mobile phones); however it is expected an increase in the number of studies which address this kind of device as well as smartphones because both devices are becoming more and more popular. Additionally, it can be perceived a trending convergence of devices in the studies. All the identified inclusive design works for multiple devices (Table 7) were published in the last seven years.

For the best of our knowledge, there is no previous systematic review that answers the research questions of this work. Nevertheless, some analyzed studies present secondary studies (surveys) that address HCI aspects for the design of multiple devices and different kinds of disabilities (e.g., [11] [12] [13]).

4.5 Threats To Validity

This section discusses the threats to validity that might have affected the results of this systematic review. The review protocol was validated to ensure that the research was as correct, complete and objective as possible. However, possible limitations in two moments of the review process were identified: in the publication selection and in the data extraction.

The search for publications was performed in two major steps: (i) automatic search and (ii) manual search. In step (i) there is a limitation because the search string could not be used in SpringerLink library, which possibly leads to a reduction in the considered studies. In step (ii), it was identified a limitation concerning the papers included in the review. The manual search was performed only on a limited set of journals and conferences and it was expected that relevant studies published in other journals or conferences would be captured through the automatic search realized in the previous step. However, it cannot be guaranteed that all related papers published are included in this review. Moreover, it is possible that some kind of inaccuracy or misclassification have occurred in the data extraction performed in this systematic review, mainly because the data extraction was done individually by the researchers.

5 Conclusions

Web movement beyond desktop to different devices increased the challenges for accessibility, while, at the same time, this movement amplified the possibilities to ensure access to information for all. This paper presented a systematic review on the design for different users and multiple devices. Thus, a roadmap towards inclusive environments was drawn based on the extracted data in order to answer two research questions: “RQ1. *Which interface design approaches are being used to address multiple devices and users with different conditions*?” and “RQ2. *Which tools are being used to support the design of inclusive applications for multiple devices*?”. A systematic review protocol was defined and the automatic and manual search returned a total of 4061 studies between 2002 and 2012. After applying the exclusion/inclusion criteria, it led to the inclusion of 176 studies in the review.

The results indicated that: i) the HCI community is presenting solutions for the software development lifecycle until the software deployment, but only few researches are considering software evolution and interaction design during *use time*; ii) there is a gap between the software engineering and the HCI communities regarding the subject, because some more recent software engineering approaches for development and maintenance are not in the analyzed work yet; in the same way that software engineering approaches need to consider inclusiveness as an important non-functional requirement; iii) a connection between the HCI and the Semantic Web and ontologies communities could be identified. Some included studies explored Semantic Web and ontologies technologies to automate user interface generation; iv) it is important for the HCI community to increase the number of studies addressing underserved people and also to interact with local research communities addressing this problem; v) there is a growing curve on the number of publications on the subject addressed by this work, in the last decade.

As further work, we expect to: i) diminish the threats to validity, ii) analyze the studies according to each disability, iii) analyze each ID approach based on software development lifecycle, iv) evaluate each provided tool by developing inclusive environments, v) propose new ID approaches to cover software development, maintenance, and dynamic evolution, and finally, vi) extend the study to consider other research questions and more papers of relevant authors identified in this systematic review.

Acknowledgements. This work has been supported by the Brazilian Institutions: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processes: 150113/2013-7 and 560044/2010-0 (EcoWeb Project).

References

1. Medeiros, C.B.: Grand research challenges in computer science in Brazil. Computer 41, 59–65 (2008)
2. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for multi-target user interfaces. Interacting with Computers 15, 289–308 (2003)
3. Gajos, K.Z., Weld, D.S., Wobbrock, J.O.: Automatically generating personalized user interfaces with Supple. Artificial Intelligence 174, 910–950 (2010)
4. Falb, J., Popp, R., Rock, T., Jelinek, H., Arnautovic, E., Kaindl, H.: Fully-automatic generation of user interfaces for multiple devices from a high-level model based on communicative acts. In: 2007 40th Annual Hawaii International Conference on System Sciences (HICSS 2007), p. 26 (2007)
5. Kitchenham, B.: Procedures for performing systematic reviews, TR/SE-0401, Department of Computer Science, Keele University, Keele, Staffs, UK (2004)
6. Beecham, S., et al.: Motivation in Software Engineering: A systematic literature review. Inform. Softw. Technol (2007), doi:10.1016/j.infsof.2007.09.004
7. Almeida, L.D.A., Baranauskas, M.C.C.: Accessibility in Rich Internet Applications: People and Research. In: Proceedings of IHC 2012, 11th Brazilian Symposium on Human Factors in Computing Systems, vol. 1, pp. 3–12 (2012)
8. Buchdid, S.B., Baranauskas, M.C.C.: IHC em contexto: o que as palavras relevam sobre ela. In: IHC 2012, Proceedings of Brazilian Symposium on Human Factors in Computing Systems, Cuiabá, vol. 1, pp. 199–208 (2012)
9. Biolchini, J., Mian, P.G., Candida, A., Natali, C.: Systematic Review in Software Engineering. Engineering 679, 165–176 (2005)
10. Pohl, K.: Software Product Line Engineering. Springer (2005)
11. Loureiro, B., Rodrigues, R.: Multi-touch as a Natural User Interface for elders: A survey. IEEE (2011)
12. Economou, D., Gavalas, D., Kenteris, M., Tsekouras, G.E.: Cultural applications for mobile devices: Issues and requirements for authoring tools and development platforms. ACM SIGMOBILE Mobile Computing and Communications Review 12, 18 (2008)
13. Seffah, A., Forbrig, P., Javahery, H.: Multi-devices “Multiple” user interfaces: development models and research opportunities. Journal of Systems and Software 73, 287–300 (2004)

User Control in Adaptive User Interfaces for Accessibility

Matthias Peissner[1] and Rob Edlin-White[2]

[1] Fraunhofer Institute for Industrial Engineering IAO, Nobelstr. 12, 70569 Stuttgart, Germany
[2] Human Factors Research Group, ITRC Building, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK
matthias.peissner@iao.fraunhofer.de, epxrwe@nottingham.ac.uk

Abstract. Adaptive user interfaces offer great potential for improving the accessibility of interactive systems. At the same time, adaptations can cause usability problems, including disorientation and the feeling of losing control. Adaptations are therefore often discussed in terms of costs and benefits for the users. However, design strategies to overcome the drawbacks of adaptations have received little attention in the literature. We have designed different adaptation patterns to increase the transparency and controllability of run time adaptations in our MyUI system. This paper presents an experimental user study to investigate the effectiveness and acceptability of the proposed patterns in different cost-benefit situations and for different users. The patterns turn out to increase the transparency and controllability of adaptations during the interaction. They help users to optimize the subjective utility of the system's adaptation behavior. Moreover, the results suggest that preference and acceptance of the different patterns depend on the cost-benefit condition.

Keywords: Adaptive user interfaces, design patterns, accessibility, user study, controllability.

1 Introduction

Self-learning and self-adaptive user interfaces offer big potential for accessible technology. In contrast to one universal design for all, personalized user interfaces can offer optimal levels of usability and accessibility for diverse users and context conditions by individual design solutions [1]. For personalization in accessible technologies there are two main approaches: *adaptable* and *adaptive* user interfaces (cf. [2]). Most *adaptable* user interfaces require complex configuration dialogues which can cause problems for the users [3] - especially for those who would benefit most from personalization. Therefore, *adaptive* user interfaces can be more effective in increasing accessibility. They include self-learning mechanisms to identify individual needs and system-initiated user interface adaptations to avoid and overcome barriers of use automatically and directly when they occur.

However, self-adaptive user interfaces also pose significant challenges in terms of usability, trust-worthiness and acceptability. The most prominent issues include confusing inconsistencies and the feeling of losing control over automatic changes in

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 623–640, 2013.

the user interface [4] [1]. Hence, mechanisms to improve the *transparency* and *controllability* of automatic adaptations will be a necessary feature of successful adaptive user interfaces. In this work, *transparency* means that users will be able to recognize and understand the automatic changes in the user interface and that confusion and disorientation will be avoided by making clear what has been changed and by supporting a "feeling of continuity between the user interface before and after adaptation" [5]. Moreover, it can be useful to explain the reasons for adaptations to the user in order to increase the predictability of the system behavior [4]. Adaptive user interfaces are said to be *controllable* if the user can control the intelligent and automatic adaptations. This can include opportunities for the user to prevent or actively accept adaptations, and to undo or override adaptations (cf. [6]). Many authors (e.g. [7] [8]) consider full user control over automatic adaptations as a major requirement of acceptable adaptive systems. For adaptive systems which aim at improving the accessibility, however, this claim might require reinterpretation. Impaired users sometimes need automatic adaptations in order to be able to access certain functionality and to control the interface at all. Trewin [9] argues that these users might experience an even increased rather than decreased level of overall control with automatic adaptations which require no extra interaction effort.

Adaptation mechanisms are error prone. Especially, incorrect adaptations challenge the transparency and controllability of an adaptive system. They lead to undesired user interface changes and make it difficult for the user to predict and understand the system behavior (cf. [10] [4]). Weld et al. [1] point out that incorrect adaptations are associated with costs for the user which have to be taken into account together with the potential advantages of an adaptive system. Systematic cost-benefit considerations cover also the costs of correct adaptations such as becoming aware of and leveraging the adaptation [11]. As a consequence, adaptive user interfaces will be successful only if the subjective benefits of the adaptations outweigh their costs [10]. Intelligent technologies will be limited in their ability to optimize individual cost-benefit ratios. Mechanisms to put the user into control of the system's adaptation behavior can be regarded as an effective means to optimize the subjective utility of automatic adaptations.

We recognize that the design of the adaptation process is a major issue for the usability and acceptability of adaptive user interfaces. In our MyUI system, we address this important field with dedicated Adaptation Patterns. They inform the users about intended or current adaptations. And they provide the users with dialogue mechanisms to optimize the costs and benefits of automatic adaptations by increasing their transparency and controllability [12]. This paper concentrates on an in-depth examination of the effectiveness and acceptability of the MyUI adaptation patterns from the users' perspective. This work makes the following contributions:

1. We introduce a systematic cost-benefit analysis of user interface adaptations for accessibility which serves as a basis for the experimental evaluation of adaptation dialogue mechanisms under different conditions.
2. The results of our study provide valuable insights into the interactions between the costs and benefits of an adaptation and the preferred level of user control. We discuss how the findings can be used for the design of effective adaptation dialogues.

2 Related Work

The preconditions and problems of *transparency* in adaptive user interfaces are well described in the literature. Findlater and Gajos [10] report that adaptive menus are more understandable and predictable if they preserve spatial stability and if the adaptations are "elective", i.e. they can be ignored by the users. Tsandilas and Schraefel [13] point out the negative effect of poor adaptation accuracy on user performance in adaptive menus. They explain their findings by a decrease of user reliance on the system and thereby a decrease of subjective predictability of the system's behavior.

However, only few approaches to increase the transparency of automatic adaptations have been published. In a first systematic attempt, Dessart et al. [5] propose animated transitions for showing the adaptation process to the user. They develop a catalogue of "adaptation operations" (e.g. resizing, relocating, etc.) and suggest suitable animated transitions to support continuity in the perception of the dynamically changing user interface. Other approaches aim at a deeper user understanding of the system's adaptations by providing detailed explanations [14] or "by supporting the user in developing an adequate mental model of the systems' adaptation mechanisms" [15]. However, it seems questionable if and how these approaches can be applied to adaptations which try to compensate for perceptual, cognitive and motor impairments in users.

The problem of *controllability* has mainly been treated in terms of theoretic taxonomies rather than specific design guidelines. In the concept of PLASTIC USER INTERFACES, a "supra-UI" [16] or "meta-UI" [17] is envisaged by which the user can control the system's adaptations. Three levels of controllability are distinguished: At the lowest level, the user can only observe but not influence the adaptation process. At the intermediate level ("approbation"), the user can accept or reject an adaptation proposed by the system. The highest control level is characterized by a full user control over the adaptation ("specification") [16]. With Jameson's [4] suggestion to collect simple user feedback after performing an automatic adaptation (e.g., "I don't like what just happened"), the above mentioned approbation strategy can be further subdivided into a confirmation before and a confirmation after the adaptation.

Lavie and Meyer [18] compare four similar adaptation strategies with different levels of user control in a user study: a "manual" condition, in which the user performs all actions; "user selection" where the user selects one of the alternative adaptations suggested by the system; "user approval" where the user can accept or reject the system-initiated action; and "fully adaptive" where the system performs all actions and the user can just abort the adaptation process. The four strategies differ in their effectiveness in supporting a secondary task during a driving simulation session. In routine situations where the system can offer correct adaptations, the fully adaptive variant is the most beneficial for the users. In unfamiliar situations with inappropriate system adaptations, the other conditions with active user involvement perform better. This cost-benefit pattern is even more apparent in their second study with older participants. They benefit more from automatic adaptations when they are correct - but experience higher costs in cases of incorrect adaptations [18]. Even if Lavie and Meyer's experiment investigates adaptivity in the sense of automating the performance of a user task, their results can be easily transferred to adaptive user interfaces for improved accessibility. For users with more difficulties in using technology, the benefits of correct adaptations seem to clearly outweigh the costs of losing control over the system's adaptation behavior.

Another valuable contribution of Lavie and Meyer's work is that they analyze the utility of adaptations in combination with different control mechanisms. Earlier research has focused on the costs and benefits of different adaptations [19] [13] [9] [10]. But Lavie and Meyer [18] investigate how different adaptation control mechanisms can influence the costs and benefits of the same adaptations.

However, for a systematic analysis of the costs and benefits of adaptations and their interactions with adaptation control mechanisms, their study has two main shortcomings. Firstly, the specific influence of costs and benefits cannot be separated since only two of the four possible cost-benefit conditions (one with high benefit and one with high cost) have been compared. Secondly, only objective performance measures have been taken. However, the predominant disadvantages of adaptive systems as reported in the literature do not relate to performance but to more subjective concepts such as experienced transparency and the feeling of control.

3 Adaptations in the MyUI System

3.1 Overview of the MyUI System

This section summarizes the main features of the MyUI system (see [12] for a more detailed description) to provide the conceptual background for the experimental user study. MyUI aims at mainstreaming accessible and highly individualised ICT products. Instead of developing dedicated user interfaces for specific user groups with certain disabilities, MyUI tackles the problem by providing a generic infrastructure to enhance mainstream products with automatically generated user interfaces which adapt to diverse user needs, different devices, and changing context conditions. MyUI adaptations include changes in the presentation formats and modalities, interaction mechanisms and navigation paths. MyUI takes a modular approach to adaptive user interfaces relying on the composition of multimodal user interface design patterns. Individual accessibility is achieved by combining patterns which provide proven solutions for specific interaction situations and characteristics of the user, environment and device [12].

MyUI strives for a smooth and natural adaptation process by avoiding the need of an initial configuration or user enrolment process. The MyUI system senses and interprets information about the user and the context of use directly during the interaction and refines the profiles accordingly. Profile changes immediately lead to automatic run-time adaptations of the user interface. This closed loop of self-learning user and context profiles and system-initiated run-time adaptations supports a continuously improved fit between the adaptive user interface and individual user needs. Thus, also altering user capabilities as typical in the course of aging and rehabilitation (cf. [20]) can be covered. Moreover, barriers of use can be overcome directly when they occur.

Profile changes that lead to user interface adaptations are triggered by a number of sensor events. Most triggering events are detected from the user's interaction behaviour with the current user interface. Examples include time-outs (the user does not react to a system prompt within a certain time frame), repeated undos (the user seems to repeatedly select wrong options by accident) and detours. Moreover, also hardware-based sensors as, for example, eye tracking in order to capture the user's attention can be used in the MyUI user profiling process (cf. [21]). In order to avoid too many changes on the user interface and to compensate for the uncertainty of the sensor-based recognition procedures, user profiling involves statistical processing mechanisms. As a result, not every single sensor event will immediately trigger an adaptation, but adaptations will be triggered only if a certain threshold in the user profile is exceeded. This supports user interface stability. But on the other hand, transparency suffers. Users can have problems in understanding and anticipating automatic adaptations. It can be difficult to establish a clear connection between the adaptations and their causes, i.e. the triggering events.

3.2 MyUI Adaptation Patterns

The MyUI design patterns repository [22] includes different types of patterns which serve distinct functions in the adaptation process. There are patterns for creating and updating a user interface profile which includes global variables to define general settings throughout the entire user interface (e.g. font size). Other patterns provide alternative user interface components and elements for current interaction situations. Finally, adaptation patterns describe the transition from one instance of the user interface to another. Their main goal is to assure high levels of transparency and controllability for the user during automatic adaptations. There are two types of adaptation patterns:

- *Adaptation rendering patterns* specify the transitions from the user interface before the adaptation to the user interface after the adaptation. They use animations as recommended by [5] to create continuity in the course of adaptations. The animated transitions draw the user's attention to the changing screen areas and make clear the relationship between the new, adapted user interface and the former user interface. Examples include animations to grow small fonts to bigger fonts or to move elements from one area to another. Adaptation rendering patterns are always part of an adaptation process as specified by the adaptation dialogue patterns.
- *Adaptation dialogue patterns* specify the dialogue around an adaptation to make sure that the user is aware of the adaptation and can control the system's adaptation behaviour. Typical elements of an adaptation dialogue include a notification and interaction options for the user to control the system's adaptation behaviour. The MyUI patterns repository includes different adaptation dialogue patterns with different levels of user control.

The main objective of this paper is to investigate the effectiveness and acceptability of the different adaptation dialogue patterns in different cost-benefit situations and to understand their contribution to optimize the subjective costs and benefits of an adaptation as experienced by the user. Before presenting the experimental design and the results of our user study, we describe the patterns and their assumed effects on the usability and acceptability of automatic run-time adaptations.

3.3 Adaptation Dialogue Patterns

Automatic Adaptation without Adaptation Dialogue (baseline). This pattern can be regarded as a baseline condition. It is currently not used in the MyUI system. The adaptation is triggered and performed automatically. Besides the animated transitions of the involved adaptation rendering patterns, there is no additional indication of the current adaptation. The user has no opportunity to control, cancel or undo the adaptation.

Automatic Adaptation with Implicit Confirmation (AI). During the automatically triggered adaptation, an animated icon notifies the user about the on-going adaptation in a dedicated adaptation area on the screen. When the adaptation is finished the end user can undo the adaptation via a button in the adaptation area. Moreover, the adaptation area offers a permanent access to the user interface profile and the user profile.

Compared to the baseline condition, this pattern is expected to slightly increase the transparency and controllability by adding the animated icon and the undo opportunity. However, both mechanisms are unobtrusive and require a certain level of system experience and attention. On the other hand, this type of adaptation control will not decrease the interaction efficiency as no additional steps are required if the user agrees with the proposed adaptation.

Explicit Confirmation before Adaptation (EB). Before the user interface is adapted, a dialogue box is displayed which requests the user to explicitly accept or reject the adaptation. A preview of the user interface after the adaptation supports the user's decision. If the user rejects the adaptation, the dialogue box is closed and the current user interface is not changed. If the user accepts the adaptation or if the system receives no user input (time-out), the dialogue box is closed and the adaptation is carried out.

This pattern is assumed to strongly increase the transparency and controllability. The dialogue box offers an obvious hint to the planned adaptation and does not start the adaptation before the user provides an explicit agreement. However, it interrupts the current interaction and requires an extra step in the dialogue. This might cause problems for some users and reduce the efficiency – especially with adaptations which are perceived beneficial by the user.

Explicit Confirmation after Adaptation (EA). When the automatically triggered adaptation has been finished, a dialogue box asks the user if the changes shall be kept or undone. In case of rejection, the adaptation is recalled. In case of acceptance or a time-out event, the adaptation is kept.

This pattern is similar to the EB pattern. We expect that due to the delayed confirmation dialogue, the transparency and controllability gains will not be as high as with the EB pattern. It also requires the described extra step.

4 Experimental User Study

4.1 Goals

Our study had two main goals. First, we wanted to validate the MyUI adaptation dialogue patterns and evaluate their effectiveness and acceptability compared to the baseline condition. Second, we wanted to gain insights in how to implement the patterns in the MyUI adaptation engine. Especially, we wanted to know if we can use a simple adaptation process with only one design pattern which is preferable under all conditions or if more complex adaptation process behaviour will be required in order to present different patterns in different cost-benefit conditions and to different users.

4.2 Hypotheses

User interface adaptations can support the accessibility by detecting and overcoming specific problems or barriers of use during the interaction. In these cases, adaptations will provide a *benefit* for the user. However, wrong system assumptions about individual user requirements can lead to adaptations which are not beneficial for the user. On the other hand, adaptations can also cause costs by raising the effort for task completion. This leads us to the distinction of adaptations with low vs. high costs and benefits:

- Adaptations with *high benefits* help to overcome an individual interaction barrier, e.g. increase the font size when the user has vision problems.
- Adaptations with *low benefits* have no positive usability effect for the user in the current situation, i.e. the user does not experience a barrier of use before the adaptation occurs. This condition represents useless adaptations as an effect of user profiling errors.
- Adaptations with *high costs* require an extra interaction effort compared to the user interface before adaptation, e.g. after increasing the font size, the user needs to scroll down the page in order to access the desired information.
- Adaptations with *low costs* do not require extra interaction effort.

The four cost-benefit conditions which result from a 2x2 combination of low and high costs and benefits are systematically produced in our experiment and serve as the experimental context conditions for testing the above described adaptation dialogue patterns. For each pair of pattern and cost-benefit condition, subjective user assessments are recorded to evaluate the perceived transparency, controllability, comfort of use and acceptance of the patterns as well as the subjective costs and benefits of the adaptations.

The following four research hypotheses are investigated:

1. The adaptation patterns AI, EB and EA perform better than the baseline in terms of subjective transparency, controllability, as well as acceptability and user preference.
2. The adaptation patterns differ in their capability to support users in optimizing the subjective utility of an adaptation. We assume that control mechanisms can help users to minimize the subjective costs of undesired adaptations.

3. In different cost-benefit conditions, users prefer different adaptation patterns and the patterns receive different levels of user acceptance. We assume that high-cost adaptations are associated with a stronger wish for control while high-benefit adaptations will lead to preferring more efficient patterns.
4. User preference and acceptance of the patterns differ for different user groups, e.g. users with higher levels of technology anxiety prefer patterns with more control.

4.3 Methods

Participants. For this experiment, twelve older adults were recruited. They were paid € 50 to participate for one test session of about 90 minutes. The participants were between the ages of 49 and 73, with an average of 63 years; there were seven females and five males. None of them reported significant accessibility problems. Three had minor reading problems with small fonts, two said they preferred bigger keys on remote controls, and four reported problems with the understandability of instructions or technical terms in interactive products. Six described their ICT literacy level as medium, three as low, and three as fairly good.

The selection of able-bodies participants for this experiment was deliberate. In accordance with ISO 9241-171 [23], accessibility does not only apply to disabled users but also to users with temporary impairments or users who experience difficulties in particular situations, etc. Common to all users who benefit from an adaptive and accessible system like MyUI is the experience of a barrier which is then removed by an adaptation – regardless of the many different possible reasons and disabilities. All our users were at ages where barriers of use can arise quite suddenly, even without disabilities. Involving only disabled users might cause a bias as these users might have found their ways to cope with accessibility issues and they might be prejudiced against intelligent solutions from prior experiences. Moreover, our able-bodied sample helped us to exclude undesired side effects of disabilities which are not covered by the design solutions presented in the experiment. Working with participants without disabilities allows us to fully control the cost-benefit conditions.

Thus, a major assumption of our study is that preference and acceptance of the adaptation patterns will not depend on a specific type or level of impairment but mainly on the experienced costs and benefits of the offered adaptation.

Apparatus. The experiment was conducted in a typical usability lab setting. Three Windows-based computers were used. The first computer ran the interactive test prototype which simulated an adaptive interactive TV system with a main menu, and an email application to receive, view, compose and send emails. This computer was connected to a 32” PHILIPS TV screen (32PFL7605H 12) and was controlled by an infrared remote control as commonly used with TV devices. The other two computers were used for accessing and manipulating the MyUI user and context management infrastructure to start adaptations by changing user profile variables in a Wizard-of-Oz manner. The participants were sitting on a sofa in front of the TV screen with a viewing distance of about 2 meters. They had to work on different tasks with the prototype system while using the remote control.

Tasks, Adaptations and Cost-Benefit Conditions. The following three test tasks were used. During each task, a system-initiated adaptation was performed:

1. *Task:* Select an application from the iTV main menu.
   Adaptation: Add numeric key labels to the menu items in order to allow for direct selection via numeric keys instead of cursor navigation. This adaptation is expected to help motor impaired users with difficulties in precise and repeated key presses.
2. *Task:* Select an email from the email inbox.
   Adaptation: Replace sender names by photos. This adaptation shall help cognitively impaired user who have difficulties in language reception or text reading.
3. *Task:* Read a received email message.
   Adaptation: Increase font size. This adaptation helps users with low vision.

With these pairs of tasks and adaptations, the four cost-benefit conditions were implemented. While the costs could be easily manipulated by the task instructions and the user interface design, the benefit conditions required special attention. The older participants were deliberately selected to have no considerable accessibility problems, i.e. none of the adaptations above would usually be beneficial for them. In order to systematically vary the benefit level we artificially induced interaction barriers (figure 1) which could be removed by user interface adaptations. In the first task (menu

Fig. 1. Gloves, glasses and nonsense font (Wingdings) to systematically vary the benefits of user interface adaptations

selection), the high-benefit condition was created by instructing the users to wear leather gloves in size XL which should pose a significant barrier to key pressing on the remote control. As a beneficial adaptation in this condition, enabling direct numeric key selection reduced the required number of key presses. In a similar manner, the Windows font type “Wingdings” was used for all text elements in the second task (inbox) so that the text output did not make any sense for the participants. In the high-benefit condition, this artificial barrier was overcome by adding sender photos to the email items so that the participants could easily recognize the searched email. In the third task (read email), glasses with dioptre values of +1.5 and +2.5 were used to pose reading problems which could be alleviated by increasing the font size adaptations in the high-benefit condition.

For the example of the task “read email”, figure 2 shows how the four cost-benefit conditions were produced by different instructions and artificial barriers of use.

It is important to explain that the artificial barriers (gloves, nonsense font type and glasses) were not intended to simulate any motor, cognitive or perceptual impairment in a realistic way. It would not be reasonable to assume that wearing gloves, for example, could create an interaction experience which is similar to the experience of a user with impaired finger precision. Moreover, we do not claim that this approach could cover the entire area of potential accessibility problems in a representative manner. However, it was our goal to systematically vary the benefits of automatic adaptations for accessibility. Therefore, it was important to produce the general underlying mechanisms of beneficial adaptations: The user experiences an interaction barrier and is not able to complete a given task. Then a system-initiated adaptation occurs and helps the user to overcome the problem. In order to prove the validity of

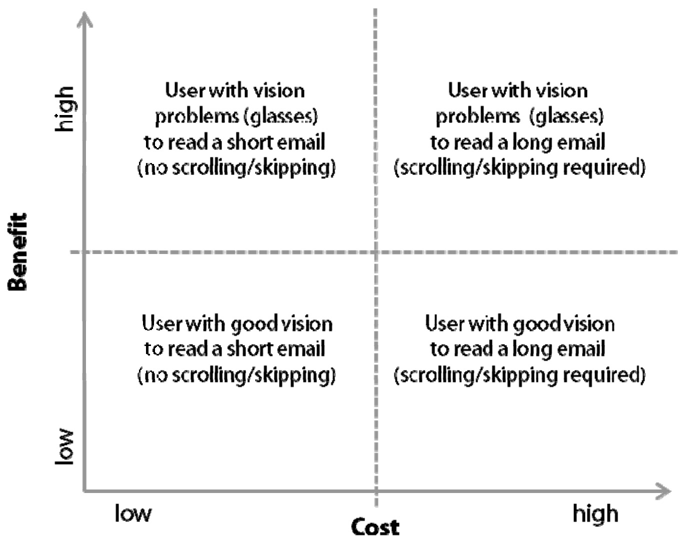


Fig. 2. Cost-benefit conditions for the adaptation of increasing the font size during the read email task (example)

our approach, we asked the participants to rate the subjective benefit of each experienced adaptation. We yielded a substantial level of agreement between the intended benefit condition and the subjective benefit ratings (Cohen's Kappa $\kappa=0.729$, n=96). The experimental variation of the costs of adaptations is supported by a fair agreement of experimental conditions and user ratings (Cohen's Kappa $\kappa=0.375$, n=96).

Experimental Design and Measures. The study was a 4 (cost-benefit condition) x 4 (adaptation dialogue patterns) within-subjects design. Task and adaptation type was not used as another independent variable but was varied systematically to make sure that our findings were not restricted to only one type of task and adaptation and that they could be generalized. Each of the three task/adaptation conditions was performed twice in all 4 x 4 test cases. The total of 96 test conditions was randomly assigned to the twelve participants so that each participant performed eight tasks. And each participant experienced two of the three task/adaptation conditions for four times, each with all four cost-benefit conditions and all four adaptation dialogue patterns – the first task/adaptation condition naïve and the second set of four runs after they were introduced into the different adaptation patterns. The restriction to two of the three task types was due to time limitations. In a pre-study we found that eight tasks were the possible maximum within the intended time frame of 90 minutes per session.

The dependent variables of the study included subjective assessments of the adaptation patterns after each of the eight test cases per participant on a four-point Likert scale for:

- *Transparency* (two items: "The design is good in communicating that a change is being processed" and "The design is good in communicating what the user can do in order to stop or undo the change")
- *Controllability* (two items: "I feel in full control of the system, its appearance and changes" and "It is easy to control the system and its appearance and changes.")
- *Comfort of use* (two items: "The design is good in supporting a comfortable system use" and "The design is good in avoiding unnecessary user input or interaction steps.")
- *General acceptance* (one item: "All in all, I like the presented style of dialogue when the system is adapting to my individual needs.")

Moreover, the participants assessed the costs and benefits of the eight experienced adaptations on four-point Likert scales. Finally, the participants indicated their preferred pattern for each of the four latter test cases and a general preference in the post-task interview at the end of the session.

Procedure. The overall procedure of a single test session was as follows:

1. After a welcome and introduction to the test situation, a questionnaire was used to obtain information on the user's ICT literacy, technology interest and technology anxiety.
2. The participants had to complete a sequence of four interaction tasks with the interactive TV application. During each interaction, an adaptation occurred with a different hypothesized cost-benefit level and a different adaptation dialogue

pattern. In the high-benefit condition, the adaptation was triggered (Wizard of Oz) after a time-out of 10 seconds when the participant was stuck with the artificial barrier. In the low benefit condition, the adaptation was triggered 20 seconds after the participant has started to work on the task with no obvious triggering event. After each task the participants assessed the transparency, controllability, comfort of use and overall acceptance of the adaptation pattern and the subjective costs and benefits of the adaptation via a questionnaire.

3. After the first four tasks, the participants were interviewed about the experienced adaptation patterns. If they did not recognize the differences, the test facilitator explained the different adaptation dialogue patterns.
4. The participants worked on another sequence of four interaction tasks with different intended cost-benefit conditions and different adaptation patterns. Again the questionnaire for the subjective assessment was filled after each task and the participants were asked to indicate which of the four different patterns they would prefer in the current situation.
5. A final interview was conducted to discuss general preferences regarding the pat-terns and to collect qualitative user feedback.

4.4 Results

Transparency, Controllability, Comfort of Use and Acceptance. The adaptation dialogue patterns AI, EB and EA were rated significantly better than the baseline adaptation in terms of transparency, controllability and acceptability (figure 3). Data analysis referred to the number of positive assessments of a criterion. For a two-item

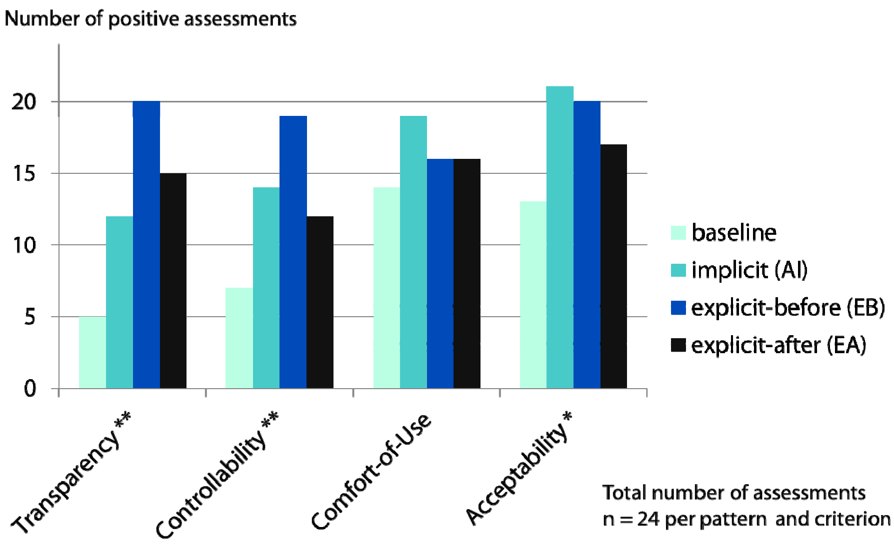


Fig. 3. Numbers of positive user assessments of transparency, controllability, comfort-of-use and acceptability for the different adaptation dialogue patterns

concept, this requires both items assessed positive (3 or 4 on the four-point Likert scale). A chi-square test against the H0 of equally distributed user ratings for all four patterns revealed significant differences for transparency (χ^2 (3, N=96) = 19.80, p<0.001**), controllability (χ^2 (3, N=96) = 12.42, p=0.006**) and acceptability (χ^2 (3, N=96) = 8.38, p=0.039*). As assumed, the dialogue pattern EB yielded the best assessments for transparency and controllability. The comfort-of-use ratings did not differ between the four patterns.

The analysis of the user preferences yielded similar results (table 1). The patterns differed significantly in their user preference for the single test tasks (χ^2 (3, N=48) = 11.00, p=0.012*). In only 3 out of 48 test cases (6.25%), the baseline condition was the preferred adaptation mechanism. The patterns AI and EB were preferred by most users.

Table 1. Preferred adaptation dialogue patterns (absolute frequencies)

	baseline	AI	EB	EA
Preference per test case (n=48) *	3	17	17	11
General preference (n=12)	1	5	5	1

Control of Subjective Costs of Adaptations. Considering all 96 test cases, the adaptation dialogue patterns did not differ in their capability to reduce the subjective costs ($Cost_{subj.}$) compared to the costs as intended by the experimental condition ($Cost_{int.}$). The number of cases where participants reported low subjective costs while being in a high-cost test condition, were equal for all four patterns (χ^2 (3, N=96) = 0.54, p=0.91). However, when users were aware of the different adaptation patterns after the review, significant differences were found (χ^2 (3, N=48) = 8.76, p=0.033*). According to our assumptions, the two patterns with explicit confirmations seemed to be most effective in reducing the costs of high-cost adaptations (table 2).

For the subjective benefit ratings, no effects of the patterns were found. It seems that even patterns which require extra effort do not reduce the advantages of adaptations which are beneficial for the users.

Table 2. Adaptation dialogue patterns and their capability to control the costs of adaptations. The table shows absolute frequencies of cases where participants report low subjective costs while being in a high-cost test condition ("$Cost_{subj.} < Cost_{int.}$").

	baseline	AI	EB	EA
All test cases (n=96)				
$Cost_{subj.} < Cost_{int.}$	6	7	8	8
$Cost_{subj.} \geq Cost_{int.}$	18	17	16	16
After review[1] (n=48) *				
$Cost_{subj.} < Cost_{int.}$	1	1	5	6
$Cost_{subj.} \geq Cost_{int.}$	11	11	7	6

[1] "After review": the users are aware of the different patterns

Preference and Acceptance Depend on Cost-Benefit Conditions. In contrast to our assumption, user preferences did not differ between different cost conditions. A reason might be that already in the low-cost conditions, the patterns with high control levels were preferred very often. A further effect into the expected direction might have been made more difficult by a ceiling effect.

However, a significant interdependency of the benefits as intended by the experimental conditions and the pattern preferences was discovered (χ^2 (3, N=48) = 8.07, p=0.045*) (table 3). Especially, the implicit confirmation pattern AI was more often preferred with highly beneficial adaptations. This effect was even stronger for the subjective benefit as reported by the users (χ^2=15.70, df=3, p=0.001**). As predicted, when the benefits of an adaptation were high enough, controllability loses its importance and the more efficient implicit confirmation (AI) was preferred.

The different cost and benefit conditions had no significant effect on the acceptance ratings of the different patterns, when regarding the intended utility. However, subjective costs and acceptance judgments were significantly associated (χ2 (7, N=96) = 22.21, p=0.002**). Especially, the baseline pattern and EA received less positive acceptance judgments when the adaptation caused higher subjective costs. This result indicates that adaptation mechanisms with insufficient user control will be disliked when adaptations with high subjective costs occur. On the other hand, acceptance judgments were also significantly associated with the subjective benefits of the adaptation (χ2 (7, N=96) = 14.38, p=0.045*). Especially, the baseline pattern gained higher acceptance with high-benefit adaptations. This finding is in line with the above reported result of less important control mechanisms for beneficial adaptations.

Table 3. Absolute frequencies of user preferences for the different adaptation dialogue patterns under the experimental conditions of low vs. high intended costs and benefits

	baseline	AI	EB	EA
Intended costs				
Low (n=24)	2	7	8	7
High (n=24)	1	10	9	4
Intended benefits *				
Low (n=24)	1	5	13	5
High (n=24)	2	12	4	6

Influence of User Characteristics. The collected measures of individual ICT literacy, technology interest and technology anxiety were not associated with the users' acceptance and assessment of the patterns, except for the comfort-of-use judgements were significant interdependencies with ICT literacy (χ^2 (7, N=96) = 24.73, p<0.001**) and technology interest (χ^2 (7, N=96) = 15.02, p=0.036*) were found. Users with higher levels of ICT literacy rated the comfort of use of EB better whereas users with low ICT literacy rated the comfort of use of EA better. Finally, users with low technology interest judge the comfort of use of EA better than users who are more interested in technology.

5 Discussion

Our work provides several interesting insights with respect to control mechanisms for automatic user interface adaptations. Some of our initial assumptions could not be confirmed by the empirical results. The rigid nonparametric statistical analyses might have been responsible for a relatively low statistical power. Our main findings can be summarized as follows:

The two adaptation dialogue patterns AI and EB have clearly shown to be best in terms of transparency and controllability. It seems that an optimal adaptive system will need both approaches: a quick and efficient adaptation process which provides only an undo for unsuccessful adaptations and a more obvious and explicit dialogue which waits for a user confirmation before changes are executed. The two preferred patterns are quite different and have their specific advantages under different conditions.

The explicit confirmation pattern proved to be more appropriate for situations in which adaptations are likely to cause costs for the user. To a certain degree, it seems that explicit control mechanism can reduce the negative impressions of automatic adaptations. In our experiment, the users still had to face the same inconveniences. But the fact that they can decide about the system's behaviour seems to alleviate the subjective disadvantages. On the other hand, the extra interaction step needed for the explicit conformation did not reduce the advantages of a helpful adaptation.

However, in situations where adaptations provide great benefits to the user, the implicit adaptation dialogue (AI) was clearly preferred. In this regard, our experiment confirms Trewin's [9] claim of a minor role for user control in adaptive user interfaces for accessibility. Even our baseline pattern with poor overall ratings for transparency and controllability was significantly better accepted when the users experienced a beneficial adaptation.

Our analysis of certain user characteristics and their influence on the subjective assessments of the different patterns brought the EA pattern into the discussion again. The three tested characteristics ICT literacy, technology interest and technology anxiety are covered by the MyUI user profile. It would be very easy to present the EA pattern to users with low levels of ICT literacy and technology interest. But the re-ported interaction effect is restricted to the EA pattern. And its absolute rating levels are – even for the most positive cases – still not better than for the other two adaptation dialogue patterns AI and EB.

Therefore, a combination of the two preferred adaptation dialogue patterns seems most appropriate. Based on these findings, we will refine the MyUI adaptation engine in a way that adaptations which cause only minor changes (i.e. changes without switching to other interaction design patterns, cf. [12]) will be associated with the implicit design pattern AI. Other adaptations which cause more substantial changes, e.g. substituting one user interface component by another, will engage the explicit adaptation dialogue pattern EB.

We have implemented a novel experimentation approach which aimed at producing generalizable results. Our cost-benefit classification aimed at generalizing from specific user interface adjustments and our procedure of systematically varying the adaptations' costs and benefits seemed to be effective in generalizing from specific impairments. If our assumptions are valid, we can generalize the findings to the design of control mechanisms for adaptive systems in general.

However, if the preference of control mechanisms depended more on specific levels and types of impairments rather than the experienced costs and benefits of an adaptation, a study with actually impaired user would be indispensable. Another limitation of our study concerns the fact that we have explained the different adaptations to the users after the first half of the experiment. This might have biased attention and awareness for the adaptation process. It is not clear how the findings apply to the real world where some users might never pay attention to the way adaptations are carried out. Finally, we were mainly interested in the preconditions of user acceptance for adaptive user interfaces. Therefore, we concentrated on aspects of the user experience and subjective ratings. In order to gain a wider view on the topic, it might be interesting to also collect objective data, e.g., about improvements in task performance for the different adaptation patterns.

In order to further validate this work we suggest a similar study with a less controlled but more realistic setting. This includes involving users with actual accessibility issues and presenting adaptations which specifically fit their individual needs. This future study shall also provide the opportunity to investigate if similar results can be achieved when the users experience the adaptation behaviour over a longer period of time without getting explicit explanations about the adaptation patterns.

6 Conclusion

The MyUI adaptation dialogue patterns mark a valuable approach to improve the transparency and controllability of automatic run-time adaptations for increased accessibility. We have presented a controlled lab experiment to validate their effectiveness and acceptance in different cost-benefit conditions. The systematic variation of cost-benefit conditions allowed us to investigate the manifold interactions between the utility of an adaptation and possible mechanisms to put the user in control of system-initiated adaptations.

Our study revealed strong support for two of the proposed adaptation patterns: "Automatic adaptation with implicit confirmation" (AI) and "Explicit confirmation before adaptation" (EB). On the basis of their specific advantages in different cost-benefit situations, we conclude to use both patterns in our MyUI adaptation frame-work. And we advise other designers to use corresponding confirmation strategies in their adaptive systems. AI is best for adaptations with great advantages for the user. It provides an unobtrusive notification of the current adaptation and offers the opportunity to undo the adaptation. In situations where high costs for the user are expected, EB should be used. This pattern has proven to reduce the subjective costs of an adaptation and offers very clear control mechanisms by requesting an explicit user confirmation before the adaptation is executed.

Acknowledgments. The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant FP7-ICT-248606. The authors acknowledge the help of other partners in the MyUI consortium in this work.

References

1. Weld, D., Anderson, C., Domingos, P., Etzioni, O., Lau, T., Gajos, K., Wolfman, S.: Automatically personalizing user interfaces. In: Proceedings IJCAI 2003, pp. 1613–1619. Morgan Kaufmann, San Francisco (2003)
2. Findlater, L., McGrenere, J.: A comparison of static, adaptive, and adaptable menus. In: Proceedings CHI 2004, pp. 89–96. ACM, New York (2004)
3. Mackay, W.E.: Triggers and barriers to customizing software. In: Proceedings CHI 1991, pp. 153–160. ACM, New York (1991)
4. Jameson, A.: Understanding and Dealing with Usability Side Effects of Intelligent Processing. AI Magazine 30(4), 23–40 (2009)
5. Dessart, C.-E., Motti, V.G., Vanderdonckt, J.: Showing user interface adaptivity by animated transitions. In: Proceedings EICS 2011, pp. 95–104. ACM, New York (2011)
6. Wobbrock, J.O., Kane, S.K., Gajos, K.Z., Harada, S., Froehlich, J.: Ability-based design: Concept, principles and examples. ACM Transactions on Accessible Computing 3(3), 1–27 (2011)
7. Norman, D.: How might people interact with agents. Communications of the ACM 37(7), 68–71 (1994)
8. Vanderdonckt, J., Coutaz, J., Calvary, G., Stanciulescu, A.: Multimodality for Plastic User Interfaces: Models, Methods, and Principles. In: Tzovaras, D. (ed.) Multimodal user Interfaces: Signals and Communication Technology, pp. 61–84. Springer, Berlin (2008)
9. Trewin, S.: Automating accessibility: the dynamic keyboard. ACM SIGACCESS Accessibility and Computing (77-78), 71–78 (2003)
10. Findlater, L., Gajos, K.Z.: Design Space and Evaluation Challenges of Adaptive Graphical User Interfaces. AI Magazine 30(4), 68–73 (2009)
11. Gajos, K.Z., Czerwinski, M., Tan, D.S., Weld, D.S.: Exploring the design space for adaptive graphical user interfaces. In: Proceedings AVI 2006, pp. 201–208. ACM, New York (2006)
12. Peissner, M., Häbe, H., Janssen, D., Sellner, T.: MyUI: Generating accessible user interfaces from multimodal design patterns. In: Proceedings EICS 2012, pp. 81–90. ACM, New York (2012)
13. Tsandilas, T., Schraefel, M.C.: An Empirical Assessment of Adaptation Techniques. In: CHI 2005 Extended Abstracts, pp. 2009–2012. ACM, New York (2005)
14. Bunt, A., Conati, C., McGrenere, J.: A Mixed-Initiative Approach to Interface Personalization. AI Magazine 30(4), 58–64 (2009)
15. Kühme, T.: A user-centered approach to adaptive interfaces. In: Gray, W.D., Hefley, W.E., Murray, D. (eds.) Proceedings IUI 1993, pp. 243–245. ACM, New York (1993)

16. Calvary, G., Serna, A., Coutaz, J., Scapin, D., Pontico, F., Winckler, M.: Envisioning Advanced User Interfaces for e-Government Applications: a Case Study. In: Assar, S., Boughzala, I., Boydens, I. (eds.) Practical Studies in E-Government: Best Practices from Around the World, pp. 205–228. Springer (2011)
17. Coutaz, J.: User interface plasticity: model driven engineering to the limit! In: Proceedings EICS 2010, pp. 1–8. ACM, New York (2010)
18. Lavie, T., Meyer, J.: Benefits and costs of adaptive user interfaces. Int. J. Hum.-Comput. Stud. 68(8), 508–524 (2010)
19. Horvitz, E.: Principles of mixed-initiative user interfaces. In: Proceedings CHI 1999, pp. 159–166. ACM, New York (1999)
20. Sloan, D., Atkinson, M.T., Machin, C., Li, Y.: The potential of adaptive interfaces as an accessibility aid for older web users. In: Proceedings W4A 2010, Article 35, 10 pages. ACM, New York (2010)
21. Hernández, J.A., Larrabeiti, D., Strnad, O., Schmidt, A.: Prototype for user context management infrastructure and user modelling. Public deliverable of the MyUI project (2011), `http://www.myui.eu/docs/MyUI_D1-2_final.pdf`
22. MyUI Design Patterns Repository, `http://myuipatterns.clevercherry.com`
23. ISO (International Organization for Standardization): ISO 9241-171:2008 Ergonomics of human-system interaction - Part 171: Guidance on software accessibility (2008)

Vibro-Tactile Enrichment Improves Blind User Interaction with Mobile Touchscreens

Maria Claudia Buzzi[1], Marina Buzzi[1], Barbara Leporini[2], and Maria Teresa Paratore[1]

[1] IIT - CNR, via Moruzzi 1, 56124 Pisa (PI), Italy
[2] ISTI - CNR, via Moruzzi 1, 56124 Pisa (PI), Italy
{Claudia.Buzzi,Marina.Buzzi}@iit.cnr.it,
Barbara.Leporini@isti.cnr.it, Maria.Paratore@iit.cnr.it

Abstract. Interaction via mobile devices is a challenge for blind users, who often encounter severe accessibility and usability problems. The main issues are due to the lack of hardware keys, making it difficult to quickly reach an area or activate functions, and to the new way of interacting via touchscreen. A touch-screen has no specific reference points detectable by feel, so a blind user cannot easily understand exactly where (s)he is positioned on the interface nor readily find a specific item/function. Alternative ways to provide content are mainly vocal and may be inadequate in some situations, e.g., noisy environments. In this study we investigate enriching the user interfaces of touchscreen mobile devices to facilitate blind users' orientation. We propose a possible solution for improving interaction based on the vibro-tactile channel. After introducing the idea behind our approach, two implemented Android Apps including the enriched user interfaces are shown and discussed.

Keywords: Accessibility, usability, mobile accessibility, haptic UIs, blind.

1 Introduction

In recent years, the development of more sophisticated smartphones with touchscreens has changed interaction modalities, while the use of hardware keys to quickly reach or activate specific functions has been decreasing. Touchscreens are completely smooth, so detecting specific user interface (UI) parts and elements is difficult for someone who cannot see the screen. Alternative ways to provide content visible on the touchscreen are mainly based on vocal channels through a voice TTS (Text-to-Speech), which may not always be enough for fully accessible and usable interaction. Vocal feedback may not work well in noisy environments; moreover, in some situations (e.g. during classes, meetings, speeches) users may prefer something more 'silent'.

Several aspects of interacting with a touchscreen can be improved to make screen exploration more satisfactory for the blind. In this paper we focus on a tactile-based solution to improve mobile interface exploration by blind users when interacting via touchscreen. The tactile channel is more immediate and direct for the blind -- cortical brain areas normally reserved for vision may be activated by tactile stimuli [11].

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 641–648, 2013.

In this perspective, a possible solution for improving interaction based on the vibro-tactile channel is presented. After an introduction to related works in the field and to the main usability issues encountered when a blind person interacts with a touchscreen, our approach will be discussed and described through examples.

2 Related Work

Several studies in the literature describe the importance of providing a user with appropriate mechanisms and techniques for better orientation on the user interface. In [4], the authors conducted a pilot study to investigate analogies between the real world and Web navigation. For the authors, organized content can benefit a reader only if (s)he is able to move around it with accuracy and agility, and is able to quickly discover and absorb its organization.

Kane et al. compared how blind and sighted people use touchscreen gestures, and proposed guidelines for designing accessible touchscreens. Blind subjects preferred gestures that use screen corners, edges, and multi-touch, enabling quicker and easier identification, and suggested new gestures in well-known spatial layouts, such as a qwerty keyboard [5]. This study observed that referencing points are particularly useful and preferred by blind users to better move around the interface. Arroba et al. proposed a methodology for making mobile touchscreen platforms accessible for visually impaired people based on a functional gesture specification, a set of guidelines to assure consistency of mobile platforms and the customization of input application [1]. These studies offered guidelines valuable for further developments.

Recently, Bonner at al. developed an accessible eyes-free text entry system that offers multi-touch input and audio output. The system, implemented on Apple's iPhone and tested with ten users, showed better performance in terms of speed, accuracy and user preference with respect to the text entry component of VoiceOver [2]. This solution focuses only on the text-entry providing audio feedback, which may not work well in some situations.

Koskinen et al. investigated the most pleasant tactile clicks, comparing piezo actuators vs a vibration motor, finding that subjectively the first was preferred [6]. In agreement with previous studies, results showed that tactile feedback improves the usability of virtual buttons pressed with the fingers, since the user is able to feel the object of interaction. Brewster and Brown proposed the use of a new type of tactile output: structured tactons, or tactile icons, i.e., abstract messages that can be used to communicate information. A tacton is characterized by parameters such as frequency, amplitude and duration of a tactile pulse, but also rhythm and location [3]. Using tactons could enhance accessibility of mobile devices for blind users as well as for sighted people in motion. Qian et al. identified the salient features of tactons when integrated with a mobile phone interface. Findings indicated that the best results use simple static rhythms, with differences in each pulse's duration. However, to ensure accurate perception, the dimensions in which paired tactons differ should be limited [10]. Yatani and Truong proposed the use of multiple vibration motors embedded in the back of the mobile touchscreen device to convey tactile feedback, providing semantic information about the touched object. They showed that users can accurately distinguish ten vibration patterns, and that the proposed system enables better interaction in an eyes-free setting than devices without tactile feedback or using only a single vibration motor [12]. However, these solutions mainly rely on hardware add-ons for providing haptic feedback while our approach offers a non-invasive/intrusive software solution, that has no impact on the usual interaction of blind users.

Other studies have investigated the use of tactile aids to enhance blind user interaction on touchscreen devices. In a preliminary study, Magnusson et al. investigated the use of haptic and speech feedback to make a digital map on a touchscreen more accessible [8], developing a prototype application that uses vibration to help blind users understand a map layout. This solution requires a time-consuming pre-processing of each map [9]. Our approach is conceptually similar to this last work since we use a mix of audio and vibration for easily detecting areas of interest on the user interface. However, despite all this research, to the best of the authors' knowledge, effective ad-hoc enrichment of general-purpose touchscreen user interfaces via software for easier orientation of blind persons has not yet been presented.

3 Interacting with a Touch Mobile Device: Usability Issues

To identify main accessibility and usability issues encountered by blind users when interacting with touchscreen mobile devices, we evaluated the interaction with Android-based and Apple mobile devices. In both cases, the inspection was carried out by all the authors (one of whom is blind) interacting with the touchscreen via screen reader (TalkBack on Android devices, VoiceOver on Apple ones) and gestures. Throughout the evaluation process, sighted authors avoided looking at the screen by activating the "screen curtain"[1] functionality. More details on the applied methodology are available in [7]. In the following we summarize the main issues observed:

- Lack of logical navigation order, to ensure the content is correctly sequentialized: the problem occurs when trying to navigate content and elements sequentially via swipe gestures ("next" - flick right - and "previous" - flick left). In this case some incongruences regarding the correct logical order occur when visiting or expanding an item.
- Unsuitable handling of focus: the problem especially occurs when editing a text field within a form composed of several control UI elements. When activating an edit field by a double tap, the system focus moves onto that field and the screen reader announces the editing modality. By exploring and clicking on the virtual keyboard letters, the focus moves on the keyboard and the user loses the editing focus point. (S)He is unable to quickly check what was edited because to do this, it is necessary to explore the UI again. This issue also arises when filling in a group of form elements. This process disorients the user and can make the action difficult and frustrating.

[1] http://www.apple.com/accessibility/iphone/vision.html

- Lack of orientation on the UI: the user can explore content on the screen by either (1) going forward and backward in a sequential order through swipe gestures, or (2) touching a point on the screen and then proceeding through the next and previous flick. This means that the user may encounter some difficulty or frustration when reading content. For instance, when reading a mail, the user has to read the message header before catching its content, unless (s)he is able to click in the exact starting point of the text of message. When clicking, the focus moves onto the clicked place and the user can start the reading from there.

4 The Proposed Solution

To address the usability issues discussed in the previous section, we propose an approach based on the use of haptic technology. Our proposal aims to support blind users as they explore and interact with content on a touchscreen. Preserving the original UI layout, this solution provides extra information and feedback for better and easy identification of the UI elements or parts.

For instance, for the Web, the W3C WAI-ARIA (Accessible Rich Internet Applications) suite[2] suggests the use of landmarks and region roles to allow developers to divide the Web page content into several parts, to create the logical partitioning of UI areas. In this way, the user is able to quickly obtain an overview of the page structure. Unfortunately this standard is still rarely applied. Moving from web to mobile devices, phone apps -- as well as any system application -- could greatly enhance applying a similar strategy to the main UI parts. In our study we decided to apply the “Logical partitioning” of UI elements to touchscreen interfaces of mobile devices to make the main parts of the UI easily and rapidly detectable. To this aim, we use reference cues (haptic mechanism) to help the user recognize those parts. Reference cues are particularly important for blind users in order to better orient themselves and move around a real and virtual space [4]. Depending on the UIs, haptic tags are added to help a blind user localize a specific part or elements on the interface.

To test our approach, we developed a prototype for an Android device, specifically a Samsung Galaxy Nexus running Android 4.2. The solution proposed herein leverages Android’s support for accessibility and aims to provide developers with a simple yet versatile tool that can be used to improve the UI usability for blind users. We will describe our approach in practice through some examples in Section 5.

4.1 Methodology

According to the Model-View-Controller (MVC) software design pattern, logic and presentation must be strongly separated. Any Android application should be 'natively' MVC-compliant, since Android development guidelines require UIs to be described entirely by means of XML files. We took advantage of this principle to develop a simple add-on that can be used to enrich our UIs. We defined a graphical object with

[2] `http://www.w3.org/WAI/intro/aria`

customizable behaviour in terms of accessibility. The goal is to offer developers a flexible tool for enriching UIs in order to improve their accessibility. XML attributes define the accessibility property of a cue and allow customizing any cue through spoken messages, sounds and vibrations. This improvement is nearly seamless since it will only require modifying an XML file to configure the cue's parameters. In the following, we refer to this kind of cue as CAC: Customizable Accessibility Cue. The XML snippet shown below represents a typical CAC, provided with a spoken message, a vibrating pattern and a sound. Once it is inserted into a layout XML file, it is rendered by Android as a 'reactive' colored ball, as shown in Fig. 1.

```
<org.cnr.iit.accessible.CustomAccessibleButton
  android:id="@+id/ballBtn01"
  android:layout_width="fill_parent"
  android:layout_height="20dp"
  android:layout_gravity="top"
  android:layout_margin="5dp"
  android:paddingTop="10dp"
  cnr:customMsg="custom spoken message"
  cnr.vibPattern="0,100,100,100"
  cnr:customSound="mySound" />
```

Fig. 1. A vibrating and speaking cue also associated with a sound, graphically rendered on a UI by a colored ball

5 Applying the Methodology to Real Cases

The proposed solution described above was tested on two real applications: a dial pad and an email client. The procedure was aimed at identifying the most suitable areas to put the cues and related feedback: a single event or a combination of events chosen from among a single vibration, vibrating pattern, speaking feedback and sound feedback, to better identify the different logical areas for each UI. We also improved the applications from other points of view, whenever their usability could be further improved. The main goal was to 'mark' the critical interaction areas by adding the customizable accessibility cues (CACs) to the layout XML files.

In the first prototype, the CACs to mark each logical session were points placed on the left side of the screen. A preliminary test with a blind user highlighted that this solution could have been misleading, since the hints were available only when exploring a narrow area of the screen. To resolve this issue, each CAC on the UI was 'stretched' (by means of proper dimension attributes) to become a horizontal strip. A pilot test was carried out with two blind users in order to assess and improve the proposed approach on two applications. The users were asked to perform five tasks using the Samsung Galaxy Nexus provided by the authors; the test was carried out in a laboratory applying the Think Aloud protocol. We first introduced users to our general idea of marking the main UI parts to support the exploration. No specific indication on how those markers were placed or implemented (audio and tactile) was given. The valuable feedback we received allowed us to refine both the haptic cue, size and vibration time, and the feedback provided, to enhance user interaction experience. All of the tests were performed after having enabled and suitably configured the Android's screen reader TalkBack.

5.1 The Dial Pad

In order to choose the better locations of the CACs, we identified the main interaction areas: the status bar (battery indicator, mobile network, time, etc.); numeric display area (where the number the user is dialing is visible), numeric keyboard (to dial the number), call and delete buttons; navigation bar (home and back buttons). Then we placed four CACs to separate these five areas. The uppermost and the lowest were associated with a sound both to highlight the boundaries of the part of the UI strictly related to the dial pad, and at the same time, marking the status -- at the top -- and the navigation -- at the bottom -- bars. These hints helped avoid accidentally pressing of one of the smartphone's built-in soft-keys that compose the status and the navigation bars. The intermediate cues were placed to highlight the other areas: the beginning and end of the numeric keyboard is marked by means of a vibrating pattern, thus making it easy to detect also the numeric display area and the call and delete buttons.

Fig. 2. The telephone app with the CACs introduced to improve its usability

Blind users performed two tasks: (1) dialing a call and (2) checking the phone number on the numeric display area. Blind users were able to perform the tasks and reach the desired areas easily and rapidly. While performing our tests, we noticed that other enhancements could be made to improve the usability when dialing a number: the app was modified in order to have each number announced, both when hovering over the relative button and when double clicking it, i.e., when the number was selected. In order to check the correctness of the dialed numbers, a "spoken message event" was added when hovering over the number display.

5.2 The Email Client

K-9 is an open-source email client based on the Email application shipped with the initial release of Android. The application is quite complex and offers a complete environment in which to organize and compose emails. Once the appropriate XML layout files were identified, we introduced the cues according to the principle of highlighting main/critical interaction areas. The application was tested by two blind users considering three tasks: (1) browsing the list of incoming emails (2) reading an email and (3) composing an email. When browsing the inbox (1) it was necessary to border the UI to highlight the email list, with one CAC on the top and one on the bottom of the list. When reading an email (2) it was necessary to mark the email text; the cues were positioned as in (1). In the ‘Compose’ UI (3) a cue was placed before the soft-keyboard to separate it from the editing area. These cues were all associated with the same sound since they share the same function of ‘UI border’.

Following the suggestions of the two blind users who tested the applications, other improvements were made to enhance the usability of text fields: a sound was associated with the event of a text field getting the focus, thus becoming suitable for filling. Moreover, to favor their detection, the text fields were announced by a 100-msec-vibration - normally -- or a 300-msec-vibration -- if the text field held the focus.

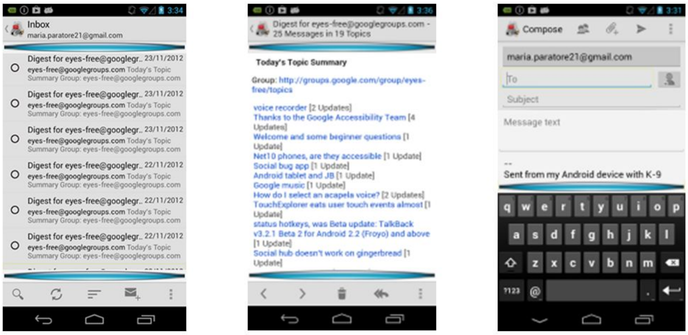


Fig. 3. Position of the CACs delimiting interaction areas in three UIs of K-9, respectively: mail browsing, single message view, mail composer

6 Conclusions

In this work we discussed how the principle of the “Logical partitioning of UI elements” can be applied to a mobile interface in order to enhance touchscreen interaction for blind users. The proposed solution is based on a customizable combination of haptic and audio feedback that can be placed programmatically on UIs with usability problems. Only vocal feedback may not work well in noisy environments or cannot be used during classes, meetings, speeches, etc. The Customizable Accessibility Cue (CAC), a flexible add-on for enriching the UIs with spoken messages, sounds and vibrations, was developed. To test our approach, we considered two Android applications, a telephone and an email client. A pilot test was conducted with two blind persons, who provided enthusiastic and positive feedback concerning the usefulness of the proposed solution, and useful feedback for refining on the UIs.

Future work will include some improvement of the methodology, e.g., expanding the number of UIs and identifying the potential best set of CACs for each UI; using different short sounds to announce useful information, such as focus shifting from one area to another, or to provide an additional confirm for number editing on the phone keypad, etc. Furthermore, we need to make a user test with a group of blind persons in order to evaluate their performance interacting with both the simple and enriched user interfaces, gathering quantitative data for evaluating and improving the proposed solution, which is potentially applicable to touchscreens of any device.

References

1. Arroba, P., Vallejo, J.C., Araujo, Á., Fraga, D., Moya, J.M.: A Methodology for Developing Accessible Mobile Platforms over Leading Devices for Visually Impaired People. In: Bravo, J., Hervás, R., Villarreal, V. (eds.) IWAAL 2011. LNCS, vol. 6693, pp. 209–215. Springer, Heidelberg (2011)
2. Bonner, M.N., Brudvik, J.T., Abowd, G.D., Edwards, W.K.: No-Look Notes: Accessible Eyes-Free Multi-touch Text Entry. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive 2010. LNCS, vol. 6030, pp. 409–426. Springer, Heidelberg (2010)
3. Brewster, S., Brown, L.M.: Tactons: structured tactile messages for non-visual information display. In: Cockburn, A. (ed.) Fifth Conf. on Australasian User Interface (AUIC 2004), vol. 28, pp. 15–23. Australian Computer Society Inc., Darlinghurst (2004)
4. Goble, C., Harper, S., Stevens, R.: The travails of visually impaired web travellers. In: Hypertext 2000, pp. 1–10 (2000)
5. Kane, S.K., Wobbrock, J.O., Ladner, R.E.: Usable gestures for blind people: understanding preference and performance. In: CHI 2011, pp. 413–422. ACM (2011)
6. Koskinen, E., Kaaresoja, T., Laitinen, P.: Feel-good touch: finding the most pleasant tactile feedback for a mobile touch screen button. In: ICMI 2008, pp. 297–304. ACM, New York (2008), `http://doi.acm.org/10.1145/1452392.1452453`
7. Leporini, B., Buzzi, M.C., Buzzi, M.: Interacting with mobile devices via VoiceOver: usability and accessibility issues. In: OzCHI 2012, pp. 339–348. ACM, New York (2012), doi:10.1145/2414536.2414591
8. Magnusson, C., Molina, M., Rassmus-Gröhn, K., Szymczak, D.: Pointing for non-visual orientation and navigation. In: NordiCHI 2010, pp. 735–738. ACM (2010)
9. Poppinga, B., Magnusson, C., Pielot, M., Rassmus-Gröhn, K.: TouchOver Map: Audio-Tactile Exploration of Interactive Maps. In: MobileHCI 2011, pp. 545–550 (2011)
10. Qian, H., Kuber, R., Sears, A.: Towards identifying distinguishable tactons for use with mobile devices. In: 11th International ASSETS 2009, pp. 257–258. ACM, New York (2009), `http://doi.acm.org/10.1145/1639642.1639703`
11. Sadato, N., Pascual-Leone, A., Grafmani, J., Ibanez, V., Deiber, M.P., Dold, G., Hallet, M.: Activation of the primary visual cortex by Braille reading in blind. Nature 380, 526–528 (1996), doi:10.1038/380526a0
12. Yatani, K., Truong, K.N.: SemFeel: a user interface with semantic tactile feed-back for mobile touch-screen devices. In: 22nd UIST 2009, pp. 111–120. ACM, New York (2009), `http://doi.acm.org/10.1145/1622176.1622198`

Designing with Dementia: Guidelines for Participatory Design together with Persons with Dementia

Niels Hendriks[1], Frederik Truyen[2], and Erik Duval[3]

[1] Social Spaces | CUO - MAD-Fac - LUCA - KU Leuven
niels.hendriks@khlim.be
[2] Cultural Studies, KU Leuven, Belgium
fred.truyen@arts.kuleuven.be
[3] Dep. Computerwetenschappen, KU Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium
erik.duval@cs.kuleuven.be

Abstract. Involving all stakeholders in the design process is often seen as a necessity from both a pragmatic and a moral point of view [1]. This is always a challenging task for designers and stakeholders and therefore many participatory design methods have been developed to facilitate such a design process. The traditional participatory design methods, however, are not fully appropriate to incorporate persons with dementia [2], [3]. They create issues as they assume that the participants are cognitively able; can make use of visual and hands-on techniques; or require a high level of abstraction ability of the person with dementia.

The aim of this paper is to present a number of guidelines which can be used as a starting point to set up participatory design projects with persons with dementia. This overarching set of guidelines provides for practical advice focusing on the role of the moderator, the preparation of a participatory session, the choice and adaptation of the method, the tools used, the role of each participant and the subsequent analysis. The basis for these guidelines stems from similar participatory projects with senior participants, persons with dementia and participants with aphasia or amnesia, two symptoms frequently co-occurring with dementia. All guidelines were evaluated and refined during four sessions with persons with dementia and a trusted family member. These participatory design sessions occurred in the course of the AToM project, a research and design project that tries to design an intelligent network of objects and people to ameliorate the life of persons with dementia.

Keywords: participatory design[1], persons with dementia, method, guidelines.

1 Participatory Design: A Necessity

Participatory Design [PD] is a set of rules, methods and theories that tries to work towards an enhanced participation of all stakeholders in the design process. Taking

[1] We choose to use the term participatory design and not collective design or the Scandinavian tradition, being the predecessors of participatory design [4]. Neither did we use the term participative design which is set more in a UK tradition of participation and is thus less focused on co-creation [5].

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 649–666, 2013.

material from cultural studies Muller [2] describes that the use of PD practices happens in or leads to an in-between region, a so-called third space. This third space gives way for dialogues between the different stakeholders (a designer or developer, an end user, a policy maker,...) of a design process and in this way becomes a space where mutual learning and collaboration (designing together) can take place.

The reasons for doing PD are quite versatile but seem to come down to either pragmatic or moral reasons [1]. On a pragmatic ground, by working in close collaboration, the chance to end up with a successful design outcome reflecting the perspectives and preferences of the future users, seems to be much higher. In the pragmatic view, knowledge of the use context is needed by non-end users such as designers and developers; end-users, on the other hand, need knowledge of, for example, technological options. The moral proposition, on the other hand, is based on the belief that the knowledge of the end-users and the designers should be bridged [6] on political and ethical grounds.

In this paper we will lend on the work that has been done in the AToM project. AToM, short for A Touch of Memory, tries to create an intelligent network of objects and people to ameliorate the life of the person with dementia, their family and professional caregivers. From the start, the AToM project had the intention to have developers, designers, caregivers, the person with dementia and their family participate in the design process. This paper mainly focuses on the PD together with persons with dementia. The aim is to list guidelines to set up a PD process with the persons with dementia. This list of guidelines is based on a literature review of participatory projects together with older people and people with dementia and participants who suffer from amnesia and aphasia. These guidelines were used and refined throughout the AToM project. The novelty of this research lies in the attempt to present an overarching set of guidelines on doing participatory design together with people with dementia.

2 Dementia and Participatory Design

Dementia is a term used to describe a decline in mental ability that will interfere severely with daily life. The most common types of dementia are Alzheimer's disease, accounting for the majority of dementia (approximately 60 to 80%) and vascular dementia, the latter occurring after a stroke [7]. Working with persons with dementia asks for a specific approach taking into account the different cognitive and psychiatric symptoms (relative to the regression of their condition and the type of dementia they are in) a person with dementia can experience. Psychiatric symptoms may include personality changes, depression, hallucinations and delusions. On a cognitive level persons with dementia (from mild to moderate) mostly suffer from a deterioration of memory (such as amnesia), difficulties in language and communication (aphasia), the inability to perform purposeful movements (apraxia) and/or orientation in time and place (agnosia) [8]. Furthermore, the large majority of the persons with dementia belongs to the group of older persons who might need to deal with the physical ailments like impaired eyesight, hearing or physical coordination [9], [10]. As indicated in the intro, doing Participator Design stems from the idea that a design will be better when a person with dementia is involved in the design process (the pragmatic motivation) and that it is an ethical or political belief to do so (the moral motivation). People with dementia have rarely been directly involved in the design process and most technology development has been done via proxies, such as the person with dementia's family, friends or the professional caregiver [11], [12]. An anecdotal example which shows that working with proxies might not always be the best choice has been described by Alm [13]. When developing a reminiscence system he notices that the vision of the professional caregiver contradicted with those of the person with dementia. The professional caregiver favoured a scrapbook metaphor while the person with dementia preferred a much simpler interface design.

To only rely on proxies stems from the vision on the person with dementia as the so-called 'uncollected corpse' [14] or as someone who no longer possesses a sense of self [15]. Kitwood [16], however, states that a person with dementia must be recognized as a person with thoughts, emotions, wishes and thus, a person who can and should actively be included in research. Letting these aged and impaired individuals participate, is thus a way to protect their previously ignored interests [17], [18]. Span et al [19] see different roles for a person with dementia in a research project: as object of study, informant or as an actual participant. Only a minority of the studies researched in their literature review has the person with dementia as an actual participant (co-creator). They stress the importance of giving the person with dementia such a role as they indicate that problems with the implementation of a design "may be due to the fact that persons with dementia were not adequately involved in the development process." Moreover they suggest that letting persons with dementia co-design can lead to better designs and may have "empowering effects on them".

Despite good reasons to do so, it seems however to be a challenge to include persons with (cognitive) impairments such as dementia, in PD processes. Muller [2] believes that the strong visual and hands-on nature of most PD methods create issues for these special needs groups. Dawe [20] states that "[t]raditional user-centered design and PD activities often ask users to describe previous usage scenarios or imagine future ones" and that this is a challenging activity for amongst others persons with cognitive impairments (such as dementia). In his work with persons with dementia, Lindsay [3] sees that traditional participatory design techniques focus on productivity and work and assume that each participant[2] is cognitively able.

3 Learning from other Participatory Design Practices

As already explained above, our working method for this paper is to investigate other collaborative and/or participatory projects designing a tool, a method, an application, a process, an environment: how did these projects set up their participatory process?;

[2] Throughout this paper we will use the term participants to denote all stakeholders in the design process, meaning the person with dementia, their family, the professional caregiver, the designer, the developer and the design researcher. By placing all stakeholders under one umbrella term we want to stress the equality of all present in the design process and stress the mutual and collaborative nature of a participatory design process.

What are the lessons learned when working with their target group?; Which tools do they use?;… We will look at participatory projects together with people with dementia and participants who suffer from amnesia and aphasia. We will also try to shed a light on participatory projects which have older people as a target group. By looking at projects that zoom in at people suffering from symptoms such as aphasia, amnesia or ageing seems to be fit as dementia must not be seen as a single disease, but as a set of signs and symptoms including memory loss, decreased communication ability,... .

We did not focus on psychiatric symptoms of dementia (such as personality changes or depression) as the cognitive symptoms seem to be more crucial to the participatory process as they have a larger effect on the sensory level, an important element for participatory design. Furthermore, we were unable to find PD projects dealing with the cognitive symptoms of apraxia or agnosia.

Out of each project we will try to abstract guidelines which form practical reflections on or a concrete help to set up participatory practices for people with dementia. Most of the time, these guidelines were not literally denominated as such, but were interpreted and abstracted from the papers. The guidelines resulting from this were used, evaluated and refined during and after the AToM project.

We are aware of three potential points of critique to our approach: most of the times we were not able to determine at which level of severity the participants were suffering from dementia, amnesia or aphasia. This was possible nor for our own work in the AToM project, nor for some of the similar participatory projects we looked at. A quantifiable comparison is thus lacking. Besides this, some of the guidelines presented can be seen as too general and not typical to working with persons with dementia. In our opinion, each individual guideline should not be treated as a single item, but only in relation to the whole set of guidelines. A single guideline can thus be seen as generic, the set of guidelines is not. A last point of critique lies in the fact that it can also be felt as being too ambitious, if not audacious to try to find an overarching set of guidelines. We must stress however that we do not see this research result -the set of guidelines- as a passe-partout for each participatory project with persons with dementia, but more as a starting point and a first toolkit for researchers and designers who work with persons with dementia. Consequently, this is only our first attempt to come to these guidelines and more case studies should lead to more refinements and thus a higher level of accuracy.

In what follows, we identified the interesting lessons learned with a de# for dementia, am# for amnesia, ap# for aphasia and el# for projects with the elderly.

3.1 Dementia

Though not working in the field of design, Allan [21] has done an intense study on how persons with dementia can participate in the evaluation of their own care. She stresses the importance of non-verbal stimuli like photos of objects (de1). To consult a person with dementia in an indirect way, using a fictive third person, turned out to be successful (de2). In general, Allan states, each chosen (set of) method(s) should be tuned towards the person's background and interest (de3). Overall, she indicates that a passe-partout method is likely not to work and promotes flexibility in the used methods. This flexibility does also mean to take into account that over the duration of your research participants might lose the capacity to take part (de4).

Hanson [22] performed a study on collaboratively designing a life book tool together with persons with early stage dementia. Hanson focuses on having small groups of persons with dementia (a maximum of 8, ideally 6) in design sessions (de5) and suggests to foresee enough time for getting to know each other, being able to be flexible and repeat the content at hand during sessions (de6). In her research Hanson also relies on the partner or care staff to play an important role (de7) and gives quite some attention to the location as she not only states that the location of the sessions should be within easy reach and accessible, but also hold the correct social status. The latter is based upon the enthusiasm of her participants caused by the fact that the sessions took place in a university building (de8). Dewing [23] adds the complexity of getting the consent to participate of people with dementia and sees getting consent as a process that runs through the whole research trajectory (de9).

3.2 Amnesia

Amnesia is a neurologic syndrome that halts the ability to create new memories. In general, there are two types of amnesia. Anterograde amnesia is the inability to process and 'store' new information; retrograde amnesia is the inability to recall old memories from long-term memory. For persons with dementia anterograde amnesia is usually the first type of amnesia they are confronted with. However, people with dementia tend to have amnesia but it does not occur in all forms of dementia [24].

Wu [25] has done extensive research on the participatory design of a memory aid with people with anterograde amnesia. Besides some general guidelines like assessing each participant (am1), understanding the cognitive deficit (am2) and adapting a chosen technique to the specificities of this deficit (am3), he sees 6 major elements to take into account when working with persons with amnesia. Wu stresses the importance of holding group sessions as a way to make key decisions with multiple persons with amnesia, instead of a designer deriving design decisions from several sessions with individuals (am4). Repetition and constant reviewing turned out to be a self-evident but necessary thing to do (am5). Planned and structured meetings help to remember specific details (am6). By creating environmental support such as name tags and the same room for different sessions one create distinctive contextual cues (am7). The use of physical artifacts like use case scenarios, option listings or a storyboard can work as a physical memory aid (am8). Wu [26] also indicates the importance of incorporating a partner, a family member or a caretaker (am8).

3.3 Aphasia

Aphasia is a cognitive disorder that besides with persons with dementia also may develop after other types of acquired brain injury such as a tumor, a stroke, brain damage or an infection. Aphasia is a language disorder that affects the capacity to speak, read, write or that has an impact on the sense making of these language utterances. Almost all forms of dementia occur together with a form of aphasia [27].

Moffat [28] describes the development of the Enhanced with Sound and Images Planner, a daily planner that enables persons with aphasia to independently manage their schedules. Throughout her research Moffat used four methods: brainstorming, low-fidelity paper prototyping, medium-fidelity software prototyping, and high fidelity software prototyping. Moffat emphasises that traditional low-fi paper prototyping is not suitable as it requires to think aloud, thus to understand and to produce verbal and textual language, which is self-evidently a hard thing to do for people with aphasia (ap1). To overcome this and other burdens, she uses non-aphasic participants to correct general design flaws not typical to the condition of aphasia (ap2). Moffat also discusses the necessity of finding a large set of participants (ap3). She ends her research with the advice to connect to existing groups and organizations (ap4). This might help to gain practical experience with the target group and grow a sensitivity, necessary to work with this special user group (ap5). In order to formalize this insight into the target group, she advises to use standardised tests to assess people's abilities (ap6).

Galliers et al. [29] have set up a series of PD workshops when developing a gesture therapy tool for people with aphasia. The research challenges they experienced deal with the difficulties in the sense-making of (abstract) words and concepts (ap7). Participants turned out to have difficulties with chains of actions or reasoning (ap8) following for example pre-defined steps to take for making a gesture. The researchers tried to minimize stimuli like other conversations in the same room or too many graphics as the participants with aphasia tend to be easily distracted (ap9). The choice of room for the workshop and the distance from the elevator were chosen to cater for people's physical disabilities (ap10). Galliers et al. end their research with two more general remarks. General practicalities (for example, to find a suitable date for a next session taking into account holiday or illness) turned out to be hard and time consuming (ap11). Finally, they end with raising the question of the representativeness of the participants as each individual's aphasia is different (ap12).

3.4 The Elderly

A large majority of the persons with dementia belong to the group of older persons. Some observations suggest that senile dementia is even the normal end-point of the ageing process [30]. Besides suffering from dementia, these older persons need to deal with the physical ailments like impaired eyesight, hearing or physical coordination [9] and cognitive impairments like diminished attention, problems with memory and decision making [31]. Incorporating older persons in the design process has not been a self-evident task as many designers hold a homogeneous view on older people or tend to fully neglect them in the design process [32]. Moreover, designers feel they lack the necessary skills and experience to work with older people [33].

Lindsay [34] sees four challenges related to participatory design with older users. Throughout his research he felt that it was hard keeping participants on focus (el1) and not let them wander onto unrelated matters (also noted by Bamford and Bruce [35]). There is also a risk in not fairly translating participants views in design (el2) by over-analysing participants utterances or giving them too much complexity. The difficulty in envisioning intangible concepts formed a next challenge (el3). Lindsay also questions the nature of traditional PD methods. As most of these methods were originally intended for work-related design they do not work well with the elderly. The methods used in participatory setups with the elderly should thus also focus on the experiential aspects of design (el4).

Massimi et. al. [36] did a study of the development of mobile phones for senior citizens. Their evaluation of doing PD with senior participants adds the importance of alternative activities taking into account the different impairments an older person is facing (el5); trying to overcome deficits by pairing persons with different deficits into one subgroup (el6); and the strict manner to control the pace and structure of a session (el7). Finally, from the UTOPIA project Eisma et al. [37] conclude that researchers should clearly explain the purpose of events and the role of the participants (el8) and this should be done in an easy to understand wording (el9).

4 Combined Guidelines

In what follows we will try to give an overview of the guidelines resulting from these previous studies in designing together with older persons, people with amnesia, aphasia and dementia. As some studies would not explicitly define guidelines or the lessons learned, we tried to abstract and interpret them from the studies found.

All guidelines in the list below are formulated in an active manner. We see a guideline as a practical reflection on or a concrete help to set up a participatory practice. If applicable, similar guidelines were merged into one and the various guidelines were grouped together in 6 subgroups: preparation, method, moderator, tools, participants and analysis. Whether a guideline stems from a study on participatory design together with a person with amnesia (am), aphasia (ap), dementia (de) or the elderly (el) is indicated between brackets behind each guideline.

4.1 Preparation

1. Search for and connect to existing groups and (patient) organisations (ap4)
2. Get to know your target group, try to understand their cognitive deficit and become sensitive to their needs and situation (ap5/am2)
3. Try to get the consent of the person with dementia on various moments throughout the research process (de9)
4. If possible, try to assess each participant in a formal way (am1/ap6)
5. Give yourself enough time for general practicalities (ap11)

4.2 Method

1. Participatory design methods should address experiential aspects (el4)
2. Each chosen (set of) method(s) should be tuned towards the persons' background, interest and specificities of the deficit (de3/am3)

3. If working in a group, modify your method taking into account the different impairments each member of the group is facing (el5)
4. Adapt your method so that it will take into account the difficulties in the comprehension and production of language, both verbal and textual (ap1)
5. Adapt your method so that it will take into account the difficulty in envisioning intangible concepts or abstract notions (el3/ap7)
6. Adapt your method so that it can overcome impairments of memory (am3)
7. Adapt your method so that it aids in following a chain of action/reasoning (ap8)

4.3 Moderator

1. Researchers should clearly explain the purpose of events and the role of the participants (el8)
2. It helps the participants to hold well planned and structured meetings (am6, el7)
3. Foresee enough time for getting to know each other, for repetition and constant reviewing of the different research/design phases (de6/am5)
4. During a participatory design session try to minimize distraction and keep participants on focus (el1/ap9)

4.4 Tools

1. The location should hold an appropriate social status (de8)
2. The choice of location should take into account the deficits of the participants and ensure easy access to the meeting room (ap10)
3. As the verbal might be a problem, make use of non-verbal elements such as visual stimuli like photos of objects or physical artifacts (notes etc.) (de1)
4. Use distinctive contextual cues (like nametags) (am7)
5. Use fictive 3rd person stories to consult a person in an indirect way (de2)
6. Use easy to understand wording (el9)

4.5 Participants

1. Give the family member or trusted caregiver an important role during each session in aiding the person with dementia in his/her participation (de7/am8)
2. Work in small groups of persons with dementia (6-8) (de5/am4)
3. Try to overcome deficits by pairing persons with different deficits into one subgroup (el6)
4. Use persons who do not suffer from a deficit to get rid of general design problems (ap2)
5. Participants might fail to stay in the research track. Make sure there is some flexibility in participants (de4/ap3)

4.6 Analysis

1. Try not to over-analyse the utterances of your participants (el2)
2. Be critical towards the representativeness of your participants (ap12)

5 Evaluation and Refinement: Putting the Guidelines to the Test

The guidelines presented here were tested during the participatory design sessions with persons with dementia within the AToM project. As already indicated this project tried to create an intelligent network of objects and persons to ameliorate the life of persons with dementia, their family and caregivers. The project has a strong participatory approach trying to involve all relevant stakeholders. During the AToM project three PD sessions were set up: one with persons with dementia and the design and technical development team and two sessions were held with the technical development team, the design team and the caregivers. As the latter two are not the focus of this paper, we will only zoom in on the PD with persons with dementia and the design team.

The evaluation of these guidelines is seen from the designer's point of view. We did not take the perspective of the person with dementia into consideration, as we did not organize a formal evaluation on how they perceived the PD sessions. This is a point of critique that can be met in future research.

The PD sessions held turned out to be useful to identify the focus of the project, to put the trusted other (a family member or care giver) in a more central position in the design and give an insight on the design requirements when creating an application for persons with dementia. These elements (the translation from the session into the actual project design) are however not the core of this paper,

5.1 The Participants Involved

All participants with dementia were recruited using the help of a memory clinic. The persons with dementia were selected on the fact whether they would have been willing to communicate about their illness and have a level of self-insight. All participants had undergone a formal diagnosis of dementia, but, legally, it was impossible to receive more information on the stage of dementia the participants were in. All participants with dementia were female, the youngest was in her 70s the oldest 95. Each person with dementia was accompanied by either a partner or one (or more) relatives, sons or daughters. The initial contact was not directly via the person with dementia, but went via a trusted person (most of the time the partner or a son/daughter). It is interesting to note that we received strict instructions not to use the terms dementia or Alzheimer in the contact with the persons with dementia (we used the more euphemistic term 'memory problems'). In total we held 4 sessions with persons with dementia.

Besides the person with dementia and their family member, one designer, who assisted the person with dementia, and one design researcher who explained the different phases of each session and took care of time-keeping were also taking part. All sessions were held at the persons with dementia home and were preceded by a visit, explaining the goals of the research, introducing the researchers and going through the informed consent some weeks before the actual participatory session.

5.2 The Outline of Each Session

The designer and the person with dementia, with the aid of the family member, used icons and basic text to map out a problem definition and possible design suggestions. The chosen method is roughly based on the Map-It project. Map-It is a mapping method, a toolkit – or MAP-(k)it –, which tries to help to guide a conversation, discussion,.... While Map-it can be used in work contexts, it is based upon the idea to have an open method with a low threshold to participate and can be used for functional or experiential aims,... The MAP-(k)it typically consists of icon stickers, maps and a scenario [38]. The first map in the PD session was an abstract drawing of the person with dementia. The person with dementia was asked to place hand drawn icons (made in advance or on the spot) that represent persons, routines, places, objects,... which were of importance to her.

A selection of these important routines, places, objects, actions and persons was placed on a sketch of the layout of a 'typical' house. On the spot, the house was more or less personalised by adding elements of the garden or the interior design, or by creating a street name plate,.... The person with dementia was then asked to indicate in which way her condition affected her routines (eg. preparing dinner for the whole family), objects (eg. operating and selecting my favourite show on the television set), places (eg. walking to the weekly market and finding my way back) or contact with persons (eg. talking to my grandchildren). The reasons for linking this to a floor plan of a house was to make the abstract notion of a 'problem' more concrete.

We then used the idea of The SuperHero, a Mr Fixit who might help to overcome the issues the person with dementia is facing. The person with dementia was asked to paste The SuperHero on the five issues she found the most important to solve (eg. operating and selecting my favourite tv show on the television set). The next step was to indicate what different steps The SuperHero had to undertake to help the person with dementia (eg. provide a warning when my favourite show is about to start, set the television automatically on the correct channel,...). Finally, the designer with the aid of the person with dementia sketches a possible technological solution focussing on integration in the daily environment of the person with dementia and aesthetics.

6 Similarities and Differences with the Guidelines

In what follows we will focus on how the guidelines were put into practice in the sessions described above. We will use the same structure (from preparation to analysis) and will end each with our refinements or additions to the proposed guidelines.

6.1 Preparation

The participatory design sessions were preceded by a series of research activities (observations, interviews with family members and caregivers, empathy exercises, house visits together with caregivers,...) trying to get an insight in the life of a person with dementia and their family and caregivers, finding help in identifying possible participants and identifying domains which could be the topic of the PD sessions. We more or less focused our PD sessions on daily routines (such as eating, watching television, making coffee,...) as these are relevant activities to all ages and to all persons with dementia (4.1.1/4.1.2). One of our PD sessions needed to be re-arranged due to a surgical operation of the partner and one dropped out after the initial talk. Her partner indicated that she became too stressed some days before the first PD session was due (4.1.5).

As each PD session consisted of only one encounter with the person with dementia, we did not need to focus on getting consent on various moments throughout the process (4.1.3). We did not assess the participants in a formal way (4.1.4) as we were not aware of any tools to use which would fit this task, without placing too hard a cognitive load on our PD session.

As the recruitment of the participants was done through a memory clinic, the aim and specificity of our research was not always made clear. It once led to the confusing situation where several sons and daughters of one person with dementia turned up expecting the designer to indicate which elements of the interior design should be adapted to fit the changed condition of their mother, thus interpreting the word design as practical interior design (and not as a phase in a research project). On another occasion the design researcher and designer repeatedly needed to stress that they were not doing any medical research (after several questions on possible medication).

Proposed guideline: Communicate about your projects' goal of without intermediaries

6.2 Method

As indicated above we worked with a derivative of the Map-It toolkit using icon stickers and maps focusing on daily routines (4.2.1). The icon stickers (pre-made or hand drawn at the spot) helped to overcome the decreased verbal competences of some of our participants. The stickers were created as simple hand drawn icons with a clear text underneath, indicating what is on the icon. This serves as clarification and as a reminder for the depicted item to the person with dementia. When trembling prohibited active mapping (tearing of an icon, slightly adapting it, pasting the icons on the map, cluster different icons,...) more attention went to the telling of stories (4.2.2/4.2.4/4.2.5). With one participant whose ability to express oneself verbally was severely decreased the family member 'guided' the words of the mother and stimulated her to reply. The different phases were hard to remember for each participant which made us cut them into smaller chunks (having the design researcher repeating what the aim of each phase was) (4.2.6/4.2.7).

Besides the difficulties in envisioning intangible concepts we see also a difficulty in making choices: making a choice, even on questions that -for the designer and design researcher- seem to be quite straightforward (eg. "What do you like to eat?") was a hard effort to do (eg. "I do not know what I like to eat?"). By using a pre-made set of icons, we tried to help the person with dementia to make a choice but it still was a heavy burden. It became even more complex when we introduced The SuperHero in our session. It was our aim to use this fictive persona so the person with dementia could imagine how technology could be integrated in their lives, without having to use the terminology related to technology. After our first session we soon found out that using this playful element didn't make the design exercise more transparent, but added a layer of complexity. The intangible technology became even more intangible by introducing a fictive element that contrasted the previous phases which all dealt with more real-life elements. In later sessions, we left out this fictive element and presented basic technology in an understandable way.

Proposed guideline: Try to avoid to make an appeal to the person with dementia's fantasy; avoid too much choice.

We didn't take 4.2.3 into consideration as we only worked in individual sessions. We will focus on the reasons for this in the 'Tools' section.

6.3 Moderator

Weeks before the actual PD sessions two design researchers thoroughly explained the goal of the research. A written version of this explanation together with an informed consent was given to each person with dementia. The roles of each of the researchers and of the person with dementia were clearly explained, focussing on why we found it important to do PD and in what way they contributed in participating (4.3.1). Before each session we communicated how long it on average would take, but foresaw enough time for repeating assignments or holding a break. Each participant knew in advance the duration of the session and they were aware that they could ask to pause or even stop the session (pausing occurred, quitting didn't). As already indicated each session was split up into small chunks and after each chunk the results were reviewed (4.3.2/4.3.3). It soon became clear that almost all participants easily drifted from the topic at hand. Conversations that occurred minutes ago became again the center of the conversation. As an example, one participant was triggered by a specific icon (a trailer) that reminded her of a warm and pleasant holiday she had with her family. Throughout the rest of the session she kept on referring to this holiday and the nicely drawn icon, causing the flow of the session to be interrupted constantly. Most of the times it was the family member who tried to keep the person with dementia on track (the distraction caused more irritation with them than with the design research team) (4.3.4).

At the start of each of the 4 sessions we tried to make it personal. As we were planning on asking the person with dementia and the family member to disclose quite some personal information, we let the design researcher and designer start off by telling about their lives as well (where do they live, married or not,...). We eventually started to bring cake to some of the sessions. The first 20 minutes of each session were thus started with drinking coffee and eating cake, chit-chatting on a variety of topics. It definitely led to a relaxed atmosphere especially with those participants who suffered from some form of aphasia.

Proposed guideline: To enclose personal info will help the participants to feel at ease and be more open

6.4 Tools

We choose the person with dementia's home as it would lead to more flexibility in finding a good date for the session as well as to have the person with dementia feel at ease (4.4.1/4.4.2). The icons and the small text underneath each one of them helped to participate in the design process in a non-verbal way (4.4.3) and in understanding what was depicted (4.4.4). On the forehand all texts/icons were checked by several experts on their comprehensibility (4.4.6). We tried to introduce a derivative of the use of 3rd person stories by introducing The SuperHero (4.4.5). This has already been analysed in the Method section).

To use the own home gave the person with dementia a feeling of security, seemed to help them to overcome the feeling of anxiety (going to a 'new' location, searching for the correct room,...). It also helped to contrast with more medical related research some of our participants were also joining. The other helpful element was that open ended questions (such as what object they liked most or what routines were pleasant to them) were easier to answer using the objects they saw surrounding them. The choice for the home as the setting for our research is of course linked to the choice of working with individual persons with dementia.

Proposed guideline: Using the person with dementia's home might help to make the participant feel at ease

6.5 Participants

Four persons with dementia participated in our PD session. Quite a few potential participants didn't want to join the sessions after the first contact with the memory clinic (who helped with the recruitment). One person with dementia became too stressed and dropped out of the research and design process after we were in contact with her (4.5.5). Before actually doing the PD, we did a test run of our set up with a person without dementia, but with insight in the person with dementia's lives. This helped us in evaluating the feasibility of the setup we proposed (4.5.4). What we however were unable to test, was the role of the partner/family member in this session. Each session the person with dementia was accompanied by a family member (partner, daughter,...). In our first session, the communication with the person with dementia turned out to be quite hard, though verbal communication was still possible. The partner of this person remained inactive, not wanting to interfere in the process and giving her all the space and freedom to participate. She seemed to be lost without his support, a support she relied on for most of her daily doings. As this trusted family member is so crucial in the lives of the person with dementia, we decided to incorporate them more into the following sessions together with persons with dementia (4.5.1). The family member could then stimulate the conversation, aid in pasting the icons, 'translating' the different goals, stimulating their loved ones to not wander off, etc. Needless to say their help turned out to be crucial. The major point of critique in using family members to help in participatory design is whether the results are still genuinely coming from the person with dementia: in what way is it not the family member who suggests a certain design choice?

The following conversation is indicative: After finding several ways to integrate technology in the activity of preparing food and eating, the conversation started to deal with the colour of an artefact. The question at hand was which colour the person with dementia would want the object to have or whether she subjected a certain colour.

Daughter: But Mum, do you remember the car daddy used to have? The one you said looked really ugly?
Person with Dementia (PwD): Uhuh.
Daughter: Now, what colour did dads ugly car have?
PwD: What car?
Daughter: The car you really said looked very, very ugly.
PwD: Uhm....(sighs).... What colour?
Daughter: That ugly black car!
PwD: Oh yes! Oh yes! Bah, black. That, I didn't like. No, I didn't like that black car. Ow, he... <stops>.
Daughter: So it shouldn't be black. No?
And, mum, that nice shirt you wear when we go out? The one with the flowers? The one you said looks really nice. What colour does that have?
PwD: The one with the flowers?
Daughter: Yes, the shirt.
PwD: Ehm. Ehm. <pauses> The shirt. Red?
Daughter: Yes, red! And you always say it's such a nice colour. So what colour should it be?
PwD: Ehm. Red.

Proposed guideline: Try to filter the research results and separate results which comes from the person with dementia and which comes from the family member

We choose to organise individual sessions and not to work in groups of people with dementia. In our contacts with a self-help group for people with dementia (as part of the ethnographic study) it was suggested to hold individual sessions as these would help us to gain a quick level of intimacy. This was also noted by Bamford and Bruce [35] who found that people with dementia sometimes showed a lack of respect to one another when participating in group sessions. A conclusion we made as well after our observations during the ethnographic field study. In care facilities we experienced a strong harshness when residents were confronted with the deficiencies of other persons with dementia leading to irate whispering on the condition of the others.

Proposed guideline: add the possibility in organizing individual participatory design sessions (refinement of guideline 4.5.2 and 4.5.3)

6.6 Analysis

The results of the participatory design sessions (the maps created together with the persons, the proposed design solutions and the conversations during the sessions) were translated in several hand drawn scenarios depicting the different possible solutions raised during the sessions. The multitude of ideas gave inspiration to create prototypes in the next phases. To overcome the problem of over-analysing a single utterance by a participant and the non-representativeness of such a small sample (4.6.1/4.6.2), we went to the (formal) caregivers and asked them to check the feasibility and transferability of the scenarios. They evaluated each scenario, asked for clarification on some choices, suggested different solutions,... We did not include family members, nor went back to the persons with dementia for this stage. We believe that the caregivers are the best persons to think beyond a single unique case, while we are unsure about the ability to do so of a person with dementia or a family member.

Proposed guideline: use caregivers to help to go beyond the single cases

7 Conclusion and Further Questions

To collaboratively design with people with dementia seems to be quite a challenge. Previous studies indicated several guidelines for working with people with dementia or suffering from aphasia, amnesia and all ailments of 'normal' ageing. We tried to cluster these guidelines and nuance them or complement to this set using our experience of the AToM-project.

As a result we see a list of guidelines that might aid in the set-up of a participatory design approach with people with dementia. We are aware of the limitations of these guidelines (no quantifiable comparison, not all of them unique to working with dementia,...) but want to stress that the proposed set is not a passe-partout for each participatory project with persons with dementia but a starting point for researchers and designers who are setting up participatory projects with persons with dementia.

Consequently, this set of guidelines is only a first attempt and the guidelines should be tested in other research and design projects. At the closing of this first research phase, we are in the midst of the second part of our research. Twelve designers (graphic designers, photographers, digital and product designers) are working together with persons with dementia to design simple objects that try to ameliorate the persons with dementia's lives. In this research and design project, the set of guidelines can be put to the test and thus evaluated and refined.

We see research necessary in four other domains. First, an attempt to link the applicability (or lack of it) to the different stages of the dementia condition is needed: what refinements are necessary when for example working with severe forms of dementia? Or, how does the way of working change within a home situation or within a day care centre? In general, we think a modular set of guidelines, taking into account the specificity of the group of persons with dementia one is working with, might be a challenge to investigate. Next, the set of guidelines is now evaluated from the perspective of the designer. Future research might look at the way the person with dementia perceives the PD session. A third step would be to deepen the methods used (4.2 in the list of guidelines): can a generalized way be found to, for example, overcome difficulties in speech or envision abstract notion? Lastly, further research is needed not on the content level of the guidelines, but on the format. We want this set of guidelines to become a toolkit that is used in the daily research and design practice. To attain this, the toolkit at hand should not be a number of lines of text, but a toolkit that integrates (more) in the way a designer and a researcher work.

Acknowledgments. Part of the research described in this article was performed in the context of the Iminds project AToM. This project is cofunded by Iminds, a research institute founded by the Flemish Government. Companies and organizations involved in the project are Televic Healthcare NV, GS1 Belgilux vzw, WZC Sint-Jozef vzw, KHLimburg, Pyxima NV, Broeders Alexianen, with project support of IWT.

References

1. Carroll, J.M., Rosson, M.B.: Participatory design in community informatics. Design Studies 28(3), 243–261 (2007)
2. Muller, M.J.: Participatory Design: the third space in HCI. The Human-Computer Interaction Handbook 4235(6), 1051–1068 (2002)
3. Lindsay, S., Brittain, K., Jackson, D., Ladha, C., Ladha, K., Olivier, P.: Empathy, participatory design and people with dementia. In: Proc. of the 2012 ACM conference on Human Factors in Computing Systems, pp. 521–530. ACM (2012)
4. Ehn, P., Badham, R.: Participatory design and the Collective Designer. In: Proc. of the Participatory Design Conference, pp. 1–10 (2002)
5. Bratteteig, T.: Participatory Design — Scandinavian tradition. Presentation given on 05.02.2003 at the University of Oslo, `http://www.ifi.uio.no/~in364/docs2/Bratteteig1.ppt`
6. Spinuzzi, C.: The methodology of participatory design. Technical Communication 52(2), 163–174 (2005)
7. Alz.org: What is dementia? `http://www.alz.org/what-is-dementia.asp`
8. American Psychiatric Association: Diagnostic and statistical manual of mental disorders: DSM-IV (1994)
9. Harman, D.: The aging process. Proc. of the National Academy of Sciences 78(11), 7124–7134 (1981)

10. Lobo, A., Launer, L.J., Fratiglioni, L., Andersen, K., Di Carlo, A., Breteler, M.M.B., et al.: Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurology 54(11), S4–S9 (2000)
11. Robinson, L., et al.: A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use. Health Technology Assessment 10(26) (2006)
12. Robinson, L., et al.: Effectiveness and acceptability of non-pharmacological interventions to reduce wandering in dementia: a systematic review. International Journal of Geriatric Psychiatry 22, 9–22 (2007)
13. Alm, N., Astell, A., Ellis, M., Dye, R., Gowans, G., Campbell, J.: A cognitive prosthesis and communication support for people with dementia. Neuropsychological Rehabilitation 14(1.2), 117–134 (2004)
14. Miller, J.: Goodbye to all this. Independent on Sunday. Sunday Review Section, 3–5 (April 1990)
15. Cohen, D., Eisdorfer, C.: The loss of self: a family resource for the care of Alzheimer's disease and related disorders. W.W. Norton, London (1986)
16. Kitwood, T.: Dementia reconsidered. Open University Press, Buckingham (1997)
17. King, S.: Co-Design: A Process of Design Participation. Van Nostrand Reinhold, New York (1989)
18. Sanoff, H.: Participatory Design: theory and techniques. Bookmasters, Raleigh NC (1990)
19. Span, M., Hettinga, M., Vernooij-Dassen, M., Eefsting, J., Smits, C.: Involving people with dementia in the development of supportive IT applications: A systematic review. Ageing Research Reviews 12(2), 535–551 (2013)
20. Dawe, M.: Reflective Design-In-Use: Co-Designing an Assistive Remote Communication System with Individuals with Cognitive Disabilities and their Families. Ph.D. Diss. Un. of Colorado, Boulder. Advisor(s) Gerhard Fischer (2007)
21. Allan, K.: Communication and consultation: Exploring ways for staff to involve people with dementia in developing services. The Policy Press, Bristol (2001)
22. Hanson, E., Magnusson, L., Arvidsson, H., Claesson, A., Keady, J., et al.: Working together with persons with early stage dementia and their family members to design a user-friendly technology-based support service. Dementia 6(3), 411–434 (2007)
23. Dewing, J.: Participatory research: A method for process consent with persons who have dementia. Dementia 6, 11–25 (2007)
24. Curran, T., Schacter, D.: Cognitive Neuropsychological Issues. In: Farah, M., Feinberg, T. (eds.) Patient-Based Approaches to Cognitive Neuroscience, pp. 291–299. The MIT Press, Cambridge (2000)
25. Wu, M., Richards, B., Baecker, R.: Participatory design with individuals who have amnesia. In: Proc. of the Eighth Conference on Participatory Design: Artful Integration: Interweaving media, Materials and Practices (PDC 2004), vol. 1, pp. 214–223. ACM, New York (2004)
26. Wu, M., Baecker, R., Richards, B.: Participatory design of an orientation aid for amnesics. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pp. 511–520. ACM, New York (2005)
27. Aphasia.org: Aphasia Frequently Asked Questions, `http://www.aphasia.org/Aphasia%20Facts/aphasia_faq.html`
28. Moffatt, K., McGrenere, J., Purves, B., Klawe, M.: The participatory design of a sound and image enhanced daily planner for people with aphasia. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), pp. 407–414. ACM, New York (2004)

29. Galliers, J., Wilson, S., Roper, A., Cocks, N., Marshall, J., Muscroft, S., Pring, T.: Words are not enough: empowering people with aphasia in the design process. In: Proc. of the 12th Participatory Design Conference (PDC 2012), vol. 1, pp. 51–60. ACM, New York (2012)
30. Ritchie, K., Kildea, D.: Is senile dementia” age-related" or" ageing-related"?—evidence from meta-analysis of dementia prevalence in the oldest old. The Lancet 346(8980), 931–934 (1995)
31. Glisky, E.: Changes in Cognitive Function in Human Aging. In: Riddle, D. (ed.) Brain Aging: Models, Methods, and Mechanisms. CRC Press, Boca Raton (2007)
32. Gregor, P., Newell, A., Zajicek, M.: Designing for dynamic diversity: interfaces for older people. In: Proc. of the Fifth International ACM Conference on Assistive Technologies (Assets 2002), pp. 151–156. ACM, New York (2002)
33. Demirbilek, O.: Involving the Elderly in the Design Process: A Participatory Design Model for Usability, Safety and Attractiveness. Ph.D. Diss., Bilkent Un., Ankara. Advisor(s) Halime Demirkan (1999)
34. Lindsay, S.: Engaging older people using participatory design. In: Proc. of the 2012 ACM Annual Conference on Human Factors in Computing Systems, CHI, p. 1199 (2012b)
35. Bamford, C., Bruce, E.: Successes and challenges in using focus groups with older people with dementia. In: Wilkinson, H. (ed.) The Perspectives of People with Dementia: Research Methods and Motivations, pp. 139–164. Jessica Kingsley, London (2002)
36. Massimi, M., Baecker, R., Wu, M.: Using participatory activities with seniors to critique, build, and evaluate mobile phones. In: Proc. of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets 2007), pp. 155–162. ACM, New York (2007)
37. Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L., Newell, A.F.: Early user involvement in the development of information technology-related products for older people. Universal Access in the Information Society 3(2), 131–140 (2004)
38. Schepers, S., Huybrechts, L., Dreessen, K.: MAP-it. On Risk, Friction and Releasing Control. Swedish Design Journal 2(11), 22–26 (2011)

Navigating, Discovering and Exploring the Web: Strategies Used by People with Print Disabilities on Interactive Websites

Christopher Power[1], Helen Petrie[1], David Swallow[1], Emma Murphy[2], Bláithín Gallagher[2], and Carlos A. Velasco[3]

[1] Human Computer Interaction Research Group, Department of Computer Science, University of York, York, YO10 5GH, United Kingdom
[2] National Council for the Blind of Ireland, Whitworth Road, Drumcondra, Dublin 9, Ireland
[3] Fraunhofer Institute for Applied Information Technology (FIT), Schloss Birlinghoven, 53757 Sankt Augustin, Germany

{christopher.power,helen.petrie,david.swallow}@york.ac.uk, {emma.murphy,blaithin.gallagher}@ncbi.ie, carlos.velasco@fit.fraunhofer.de

Abstract. The majority of research into web accessibility has focused on identifying and eliminating the problems that people with disabilities encounter when interacting with the Web. In this paper we argue that we need to move away from studying user problems to studying how people with disabilities apply interaction strategies while browsing the Web. In this paper we present a study of 19 print disabled users, including blind, partially sighted and dyslexic people, interacting with a variety of interactive Web 2.0 web applications. The participants undertook tasks using concurrent and retrospective protocols to elicit information about how they interact with web content. The result of this study was a collection of 586 strategic action sequences that were classified into seven different types of strategy. Differences in the application of strategies between the user groups are presented, as well as the most frequent strategies used by each user group. We close the paper by discussing some implications for the design of websites and assistive technologies as well as the future directions for empirical research in accessibility.

Keywords: Web accessibility, user study, user strategy, print disabled Web users, blind Web users, partially sighted Web users, dyslexic Web users.

1 Introduction

The Web provides many opportunities for users to take part in work and leisure activities online. The evolution of Web 2.0, in which websites allow users to participate in creating, contribute and share content [9], makes our connections to both other individuals and organizations stronger in the virtual world. It is important that all users be able to participate equally in these activities, including people with disabilities who use different assistive technologies and interact with the Web in very different ways

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 667–684, 2013.

from mainstream users. However, recent studies show that people with disabilities still encounter large numbers of problems on the Web [1,7,12,14]. While existing web accessibility good practice covers some of those problems encountered by disabled users, there are still a substantial number of reported problems that have no clear solution. What are the design principles that can help close that gap and better support users with disabilities?

When trying to address problems encountered by mainstream users on the Web, researchers and practitioners fall back on empirical research and design principles about user interactions with the Web and other interactive systems. However, when trying to address problems encountered by disabled users, relevant empirical studies about user interactions are rare. As a result, we are unable to go back to "first principles" when trying to either remove or, at least, reduce the impact that problems have on users with disabilities. We need more empirical work eliciting and understanding interaction strategies that disabled users apply when interacting with the Web. With that information, we can construct a framework for analysing why disabled users encounter the problems on websites that are currently beyond our understanding.

In support of this goal, we present a study with blind, partially sighted and dyslexic users, users who are collectively part of the group referred to as print-disabled users [6]. In the study, we elicit the strategies that users undertake while interacting with interactive Web 2.0 applications.

2 Related Literature

There is an increasingly large body of evidence demonstrating that the Web has not been very accessible to people with disabilities throughout its history. In 2001, Coyne and Nielsen [5] conducted a study with 20 blind, 20 partially sighted and 20 mainstream users. Each user undertook one task on each of four different websites, including one search engine. Blind users were only able to complete 12.5% of tasks; while partially sighted users fared only slightly better, with a task completion rate of 21.4%. Given that mainstream users completed 78.2% of tasks, these results show the challenge of providing an equivalent experience on the Web for disabled users. In the same paper, Coyne and Nielsen presented user-based evaluations of websites conducted with 18 blind people, 17 partially sighted people and 9 people with physical disabilities. They recorded the problems encountered by participants, and the analysis of those problems produced 75 guidelines for building accessible web applications.

The Coyne and Nielsen study was in the early days of web accessibility, and one would expect that the number of problems encountered by disabled users on the Web would decrease as developers gained awareness and knowledge of web accessibility. In 2004, the Disability Rights Commission (DRC) of Great Britain undertook a Formal Investigation into the state of Web accessibility [1]. This investigation included the largest user study of web accessibility to date. In the DRC study, users with a range of different disabilities undertook tasks on 100 websites. From 913 tasks that were undertaken on the websites, 76% of tasks were successfully completed across all disability groups. However, blind users only succeeded 54% of the time, while partially sighted users achieved a success rate of 76%. So, while web accessibility did improve between Coyne and Nielsen's study in 2001 and the DRC study in 2004, many problems still remained. The DRC report provided a comprehensive list of the types of problems that were most commonly encountered by different disabled users and demonstrated there was a large amount of overlap between the different groups. For example, even though blind and partially sighted users navigate websites in very different ways, and explore information on pages using their own distinct strategies, both groups struggled with page structure and unclear navigation mechanisms.

Lazar et al. [7] conducted a diary study with 100 blind users regarding the frustrations they encountered during their day-to-day interactions with the Web. The results echoed many of the DRC findings, with poorly designed forms, navigation structures, misleading links and page layout problems all being prominent frustrations listed by participants.

In 2007, Petrie and Kheir [12] conducted a study with six blind and six mainstream users. These users undertook seven tasks on each of two mobile phone websites. Users were asked to undertake a concurrent verbal protocol during the tasks in which they identified problems as they occurred and rated those problems for their severity. The results showed that despite having very different means of interacting with websites, the two user groups shared a number of problems.

A recent study by Power et al. had 35 blind Web users undertake two to three tasks on 16 websites, with 10 users testing each website [14]. Users encountered nearly 1400 problems, ranging from lack of feedback, through to alternatives not being provided for inaccessible content. Notably, a large number of the problems related to trying to understand the layout of the content in webpages, how navigation structures were organized and finding information within the website.

When discussing Web accessibility, the problems that disabled users encounter have been central. Certainly, the examination of those problems has raised awareness about the importance of web accessibility and provided some solutions [2,10,11]. However, we argue that the current dominant approach of encouraging developers to implement websites that simply avoid those problems is not an effective way to create websites that people with disabilities can use. While some problems, such as conflicts between assistive technologies and web browsers, may be addressed by informing developers on how to avoid them, the range and complexity of problems experienced by people with disabilities is too high to be addressed by a problem-based approach.

Given the recurring problems encountered by disabled users of exploring web content, discovering the layout of pages and understanding the navigation structures in a website, it is intractable to list of all of the problems that can occur, what their causes are, and how to address them through specific implementation techniques. We must move to an approach where we can accomplish good interaction design for websites that support the interaction strategies of disabled users.

However, information regarding how disabled users interact with the Web, as opposed to what problems they encounter, is surprisingly thin.

As early as 1995 the ACCESS project elicited requirements from 150 blind users about how to improve accessibility. Their report emphasized the need for flexibility and support in getting overviews of material on webpages, allowing blind users to jump to key landmarks and providing signposting to key paths through a website [13].

Berry [2] conducted one of the earliest in-depth interview studies about blind users' interactions with the Web. He presented a number of themes that emerged from the analysis of the interviews and he makes a key observation comparing novice and competent disabled Web users. He states that competent Web users had:

"…a strong mental image of what a webpage is and how it works, combined with a structured and confident approach to information location and retrieval."

This shows that having a good mental model of a webpage is important, and therefore we need designs that support disabled users in building those models.

In 2003, Theofanos and Redish [16] presented an observational study of 16 blind screen reader users undertaking a variety of tasks on government websites. From this study the authors developed a more concrete understanding of how people use their screen readers to understand a webpage. Among the key findings was that users skip quickly through content and listen to only the first few words. Such skipping was often sequential, but users also used lists of links and headings to skip through content in a more structured way. At that time, users were reluctant to use 'skip navigation' links, as they largely did not understand where the link would take them. Users also did not typically use functionality in the screen reader to find forms, and once forms were found, users were reluctant to move between reading mode and form mode.

A follow-up study by Theofanos and Redish [17] investigated the interactions of 10 partially sighted screen magnifier users when interacting with the Web. 5 of their participants changed from high to low magnification repeatedly, while 3 others changed font sizes in the page. Others compromised between how much they could read in a lens, versus how much of the page they could see.

Takagi et al. [15] conducted an investigation of blind people using online shopping websites. They had users undertake tasks on real online shopping sites and recorded their behaviour in regards to navigation within a page. A set of search pages for the online shop had a set of accessibility enhancements added to them including: alternative texts, properly marked up headings and 'skip navigation' links. Results from 5 participants revealed that some users used links to explore a page while others moved sequentially through content on the page. On the search pages with improved accessibility, there were no major changes in the browsing behaviour of users, with the users employing their own idiosyncratic strategies as opposed to adopting new ones based on the content. Of particular interest to this work, the authors say that users "[Stopped scanning] and then crawled around to check the content" when they found information that was relevant to their task, possibly indicating a strategy being applied. While interesting, there is no information on how often this happens, or why users were undertaking that activity. The authors conclude their paper with a statement that increasing the number of landmarks users can apply in their navigation will improve their success in navigating within webpages. Work in 2010 by Trewin et al [18] with 3 screen reader users also had indications that both established and ad hoc landmarks were important.

Wantanabe [19] examined how blind and partially sighted participants used headings on websites. He had 16 sighted users and 4 blind users undertake tasks on two different sets of recipe websites: one set had heading markup and presentation for key sections of webpages, while the other set had no structured headings. Both sets of participants had shorter task completion times when working with headings present than without. While interesting, it is not clear what actual role the headings played in the overall strategies of the users, or whether they were instructed to use the headings in the study.

Bigham et al. [3] used a proxy to record the browsing behaviours of 10 blind and 10 sighted users on their home computers. Similar to findings from previous studies, users seldom used skip navigation links, and interacted almost exclusively with links that were labelled with descriptive text. Blind users also employed probing more than their sighted counterparts, in which a webpage is visited through a link and then followed by a quick return to the previous page. What is unclear is why users were employing this probing interaction? Bigham et al. propose that users are trying to cope with inaccessible Web content; however, with the data available it is impossible to tell why users are applying those accelerators.

Borodin et al. [4] provided a high level overview of the ways that blind and partially sighted users interact with the Web with screen reader technologies. Their survey of strategies was based on "… general browsing strategies that were observed in the course of several user studies." It is difficult to determine from the information provided how often the strategies discussed were used, or how well the strategies represented the overall interactions of people with visual disabilities on the Web.

Finally, Lunn et al. [8] observed nine blind and partially sighted users browsing the Web for between 2 and 4 hours in a naturalistic setting. During this study, researchers recorded occasions that users encountered problems and identified strategies for overcoming those problems. 7 different strategies were identified for dealing with accessibility problems. Importantly, this work described in detail the strategic behaviour of users. However, the strategies identified did not include many of the common behaviours discovered in previous research. There are a number of reasons why this may have happened. Lunn et al. describe using a peripheral membership role typical to ethnographic studies to avoid influencing the observations. However, in such short sessions, with no opportunity for the researchers to integrate themselves into the setting, it is possible that there was a Hawthorne effect that influenced how participants behaved. In addition to this, the researchers collected data in a very short time while trying to observe many individuals concurrently in one classroom, and so it is unlikely that this set of strategies is complete. Finally, the approach used by Lunn et al. still involves studying the problems that users have, albeit with a different lens. In order to understand the broader contributory factors that impact on users, we must study their broader interactions on the Web in the absence of problems.

Given that there has been almost two decades of research in web accessibility, it is surprising how little empirical research there is into how disabled users interact with the Web. In comparison to the hundreds of papers that have been published on how and why mainstream users interact with the web, the participant numbers on which we form the foundations of design for web accessibility is very small. Further, the most comprehensive work on how disabled users apply strategies to navigate, discover and explore the web is now 10 years old [16,17]. While useful, there has been no recent work that demonstrates that what we knew at the very beginning of the web still holds true now. Given the evolution of assistive technologies and web application design in that ten-year period, we would expect that the strategies of users would have changed and become more sophisticated; however, we have no evidence to support that supposition. Finally, and most notably, we have no framework for describing the strategies that disabled users use when working with the Web. Almost all of the empirical studies examine user interactions only at the level of sequences of key presses and other low-level actions in the interface. We must understand how these action sequences relate to what the user is trying to accomplish in the interface. By linking the action sequences to higher-level user strategies, we will begin to have a foundation on which we can solve the continuing problems disabled users are encountering on the Web

In this paper we present a study of print disabled users, including blind, partially sighted and dyslexic users, regarding the strategies they apply when interacting with web applications. We seek to answer the following questions:

1. What are the high-level strategies used by people with print disabilities when interacting with the Web?
2. What are the most frequent strategies used by each of blind, partially sighted and dyslexic user groups?
3. What are the similarities and differences in how these different user groups apply strategies?

3 Method

3.1 Design

Participants from 3 print disability groups (blind, partially sighted and dyslexic) were asked to undertake a number of tasks on Web 2.0 applications. Participants were asked to conduct a concurrent verbal protocol while undertaking each task, concentrating on the strategies they were using in doing the task. After each task they were asked to provide a retrospective verbal protocol, providing any further information about their strategies. Information about the strategies was extracted from transcripts of the protocols for detailed analysis.

3.2 Participants

19 people took part in the study, 10 men and 9 women. Ages ranged from 18 to 60 years. Participants had between 7 and 40 years experience of using computers, and each of them used a computer on a daily basis. 5 participants were blind: 3 were congenitally blind, and the other two lost their sight more than 20 years ago. All 5 blind participants used the JAWS screen reader. All participants considered their familiarity with JAWS to be good (4) or very good (5) on a 5-point Likert scale. The participants' experience of using computers ranged from 7 to 40 years, with a mean experience of 23 years. Each of the participants used their computer very frequently

and indicated their familiarity with computers on a 5-point Likert scale from very poor (1) to very good (5) to be above the midpoint (3).

7 participants were partially sighted: 5 since birth and 2 since early childhood. The partially sighted participants used various screen reader/magnification technologies: 1 participant used JAWS; 2 used SuperNova; 2 used ZoomText (specifically for the magnification and colour contrast functionality); and 2 used the zoom settings of their browser or operating system. The participants' experience of using computers ranged from 10 to 30 years, with a mean experience of 20 years. Each of the participants used their computer very frequently and indicated their familiarity with computers to be good (4) or very good (5) on a 5-point Likert scale.

7 participants were dyslexic. Their dyslexia had been diagnosed between 1 and 6 years previously. None of the participants used any assistive technologies for using the Web. The dyslexic participants' experience with computers ranged from 9 to 14 years, with a mean experience of 11 years. Each of the participants used their computer very frequently. All participants rated their familiarity with computers to be at or above the mid-point on a 5-point Likert scale.

3.3 Equipment and Software

The study was run on a number of personal computers, each with a screen size of 17 inches. Each computer ran Windows XP and had the Morae® screen recording software installed. Each computer had the following assistive technology installed for users: JAWS (v10) and WindowEyes (v7) and Supernova (Access Suite v12.08).

3.4 Websites and Tasks

In order to elicit a wide variety of strategies from the participants, we surveyed web applications that covered the breadth of Web 2.0. We surveyed 2-5 highly interactive, participatory web applications from the following domains: social networking (e.g. Twitter, Facebook), blogging (e.g. Blogger, Wordpress), e-commerce (e.g. online banking and shopping), video sharing and viewing (e.g. YouTube, online television), e-government (e.g. government portals, participation portals).

For each web application we collected tasks that allowed users to create, retrieve, update and delete content in the web application. For each domain, we identified the most frequently occurring tasks that were shared across the web applications contained within it. These frequently occurring tasks were then decomposed through hierarchical task analysis to provide a comprehensive list of subtasks.

We selected 6 of the surveyed websites for use in the study, with a minimum of one for each domain, each with 6-10 tasks that could be completed by users. These websites were: 2 e-banking sites (Smile Online and Egg), 1 social networking site (Facebook), 1 e-participation site (Citizenscape), 1 blogging site (Wordpress) and 1 internet television site (BBC iPlayer).

The complete list of tasks along with the number of subtasks identified in the decomposition (indicated in brackets) were as follows:

- Egg/Smile Banking (e-banking): Login (2), Check Balance (3), Transfer Money to a friend (4), Manage money in accounts (9), Manage contact details (3), Arrange an overdraft (4)
- Facebook (Social networking): Sign-up (6), Add a friend (3), Update status (3), Leave a message on a friend's wall (5), Create an event (5), Upload and tag a photo (3), Chat with a friend (3), Write a note (2), Find places nearby current location (1), Update privacy settings (4)
- Citizenscape (e-participation): Find a webcast (5), Watch and navigate a webcast (5), Post a comment on a webcast (2), Send a tweet about the webcast (3)
- BBC iPlayer (Internet Television): Find last watched programme (3), Find a previous episode (6), Watch a television programme (5), Listen to a radio programme (6), Manage favourite programmes (4), Manage parental guidance features (3)
- Wordpress (Blogging): Login and view a blog (3), Create a new blog post (6), Edit an existing blog post (6), Share a video on a blog (7), Liking and comment on a post (4), Delete a blog post (3), Adjust blog settings (3)

For each of the tasks, a realistic scenario was written that would provide the participants with the contextual information about the website and the task they were undertaking and to increase the ecological validity of the study, making each task as natural as possible. An example of one such scenario from the BBC iPlayer protocol was as follows:

Due to having a busy week last week, you missed an episode of your favourite TV programme, Top Gear. You are aware that older episodes of Top Gear are available on BBC iPlayer and you would like to see if you can find the episode you missed. Please locate the previous episode of Top Gear.

As we were trying to draw conclusions about overall types and numbers of strategies applied by each user group, and not a direct comparison of strategies applied between the individual websites, users did not undertake all tasks on all websites. Further, due to the fact that we were using websites in the wild, and not a predefined website under our control, it was impossible for us to predict exactly what users would do within a given task. As a consequence of this, the comparison of user strategies between particular websites is not a research question we have pursued in this work.

3.5 Procedure

Due to the level of control that was required and the recording software needed for data analysis, observations were undertaken in controlled environments. They took place in the HomeLab at the University of York, the offices of the National Council for the Blind of Ireland (NCBI) in Dublin, and the offices of the Foundation for Assistive Technology (FAST) in London. In order to preserve ecological validity, and avoid participants acting in ways that would be inconsistent with their home use, participants used the same brands of assistive technology and user agents that they used in their home environments. Participants were given an opportunity to set up the computer display, sound and related software to preferences that approximated their home use. Great care was taken to ensure that the experience was as close to their home computing experience as possible in a controlled laboratory study.

Participants were asked to undertake each task with a concurrent verbal protocol. They were instructed to concentrate on explaining what they were doing on the website to achieve the tasks given to them. The evaluator stopped participants when they reached the end of a sub-task. Participants were asked to retrospectively describe the strategies they were using, and in particular what they hoped to achieve by using that strategy. When the user had described the strategies supporting the previous sub-task, they were asked to continue from that point to complete their task. The videos for these two stages were analysed together for the key strategies that people use to interact with the Web. Evaluation sessions lasted a maximum of 2 hours. Within that time participants performed tasks on 3 to 8 websites each, undertaking between 4 and 19 tasks.

3.6 Data Analysis

The video recordings of each participant were reviewed to identify strategic action sequences. A *strategic action sequence* is a series of operations in the Web application interface that the user applies to achieve a goal. For each recording, we noted all utterances that could be interpreted as representing an action sequence. Each utterance was transcribed and time-coded along with an indication as to whether the utterance was made during the concurrent or retrospective verbal protocol The distinction was made to avoid the possibility of double-counting a single occurrence of a particular strategic action sequence that had been mentioned in both protocols.

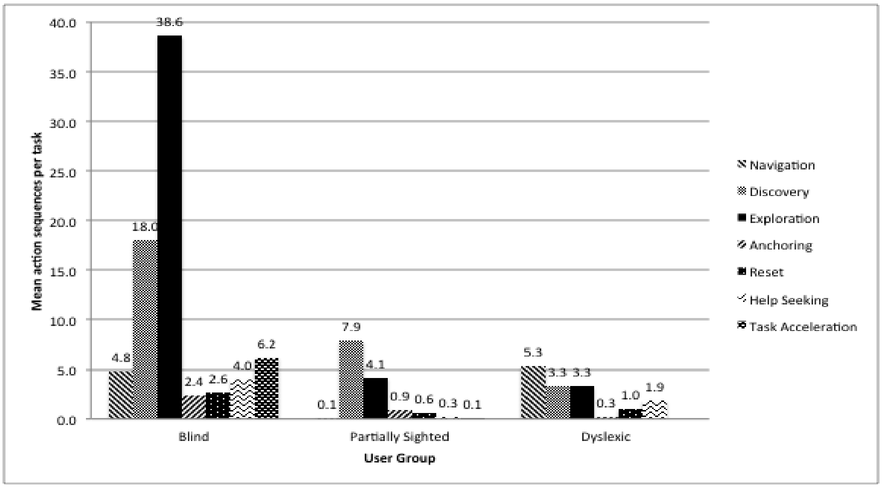


Fig. 1. Mean instances of strategy types per task for each user group

The action sequences were classified, using an emergent-grounded theory approach. Each action sequence was assigned a description and label regarding its composition. For example, an action sequence where a blind user traverses a set of headings using their screen reader was labelled *Heading Traversal*, whereas a partially sighted user scanning the area above the navigation bar and then jumping to the bottom of the page to scan the footer was labelled *Top and Bottom Scanning*. Action sequences with the same composition were assigned the same label.

When an action sequence composition was identified, the participant's comments about the action sequence and the strategy they were applying when using it were recorded. This process built up a collection of action sequences related to particular strategies. These comments were also analysed to produce a set of broad interaction strategies that were applied by users in the web applications.

4 Results

A total of 586 instances of strategic action sequences were recorded across 89 different compositions of action sequences. Blind users accounted for 383 action sequences of 62 compositions. Partially sighted users had 98 sequences of 35 compositions. Dyslexic users had 105 action sequences of 25 compositions.

From the comments of users regarding how the action strategies were applied in the web application, 7 types of strategies were identified:

- Navigation: supports users moving from page to page within a website. This includes identifying where navigation bars and menu items lead and understanding the overall site map. Examples of action sequences for this strategy include probing navigation items and then returning to the calling page, or examining tool tips for more information about where a navigation item leads.
- Discovery: is users gaining and overview of the overall structure of webpage content. An example of action sequences for this strategy would be a user reviewing the headings of a webpage to understand what are the key sections.
- Exploration: is users extracting the meaning or context of a particular piece of content or the functionality of an interactive component. An example action sequences for this strategy would be reading a heading and its related content.
- Anchoring: is users limiting the area of the webpage in which they interact. Actions sequences for this strategy includes users avoiding particular areas of the page (e.g. right hand side) or focusing on a specific area (e.g. top of page).
- Help seeking: is users actively seeking help from online documentation or from another people.
- Reset: is where a user abandons what they are doing and starts again from a safe point. This strategy could refer to restarting the whole computer or a piece of software.
- Task Acceleration: is where users are trying to speed up tasks. This includes things like keyboard shortcuts in assistive technology and web browsers.

Figure 1 presents the mean number of instances of each strategy type per task for each user group. To investigate whether the three user groups applied the strategies at different frequencies, a 2-way mixed ANOVA was conducted on the rates of strategy types per task for the strategy types Navigation, Discovery and Exploration, (the rates of the other strategy types were too low for quantitative analysis). There was a significant main effect for user group ($F = 16.25$, $df = 2, 16$, $p < 0.001$), a significant main effect for strategy type ($F = 16.90$, $df = 2, 16$, $p < 0.001$) and most interestingly, a significant interaction between the user group and the strategy type ($F = 15.20$, $df = 2,32$, $p < 0.001$).

The user group main effect showed that there was a significant difference between the overall number of strategies per task, with blind participants using more strategies per task (5.47, SD=2.36) than either partially sighted users (2.50, SD=0.82; Bonferroni post hoc: $p<0.005$) or dyslexic participants (1.53, SD=0.59; Bonferroni post hoc: $p<0.001$).

The user group by strategy type interaction showed that blind participants applied significantly more strategies for Exploration than for Navigation (Bonferroni post hoc: $p < 0.002$) and for Discovery ($p < 0.002$). Further, partially sighted people used more strategies for Discovery than for either Navigation ($p < 0.001$) or Exploration ($p < 0.01$). Dyslexics did not show any significant differences in how they used their strategies.

4.1 Frequently Occurring Action Sequence Forms

The classification of action sequences related to strategies produced a total of 89 different action sequence compositions used by participants. Table 1 presents the 12 most common compositions applied by blind users.

Blind participants spent a substantial amount of time traversing different types of content in the order it occurred on a webpage. Links and headings were commonly traversed (sequence B1 and B2 in Table 1); however, participants seldom expressed that they were trying to get an overview of the page using these sequences. Only 15

Table 1. Most frequent action sequence compositions for blind participants

Action Sequence Description	Instances N (%)
B1) Traversal of links on a page	51 (13.3)
B2) Traversal of headings on a page	43 (11.2)
B3) Keyword search within page or site	26 (6.8)
B4) Sequential traversal through page content	24 (6.3)
B5) Task accelerator in assistive technology or user agent	19 (5.0)
B6) Traversal of form controls within one form to understand form	14 (3.7)
B7) Fishing for controls with the cursor in a localized area of the page	11 (2.9)
B8) Search for form controls with assistive technology	11 (2.9)
B9) Probing controls to identify what they do	10 (2.6)
B10) Jump to known page control	9 (2.3)

of the 94 strategy instances in these two most frequently used action sequence compositions were classified as being used in support of the Discovery strategy. Instead, participants were usually exploring content in a local area on the webpage. In the case of headings, participants checked headings to see if they related to a specific piece of content they were looking for on the page. In the case of links, participants were often looking for a cue as to what the next action would be to complete their tasks.

There were two sequence compositions (B3 and B10) where blind participants would try to jump to a specific piece of content. When blind participants felt that there should be a form to aid them in their task on a webpage (e.g. login, money transfer, status update), they would use their screen reader to jump to the first form. Similarly, when specific content was expected on a page (e.g. a television programme), participants would do a keyword search within the page to try to jump to that content.

Blind participants also applied an action sequence composition for Exploration that we refer to as *fishing for controls*. In this sequence, participants would suspect that a control for a form was in the area they had been exploring but they could not locate it. Participants would activate the screen reader cursor to try to trigger any control in the area (B7). If they were unsuccessful, they would move the cursor a small amount, and then try again. This behaviour would continue until they activated a control or gave up.

The most frequent action sequence compositions for partially sighted users are presented in Table 2. The majority of the action sequences were used for Discovery.

Table 2. Most frequent action sequence compositions for partially sighted participants

Strategy	Instances N (%)
P1) Checking tooltips for further information about link purpose	9 (9.2)
P2) Scrolling to the bottom of the page looking for content	8 (8.2)
P3) Moving viewport vertically on visible edge to understand hierarchical structure of content	8 (8.2)
P4) Checking for expected content in locations consistent with web conventions	7 (7.1)
P5) Moving viewport vertically while highlighting content to understand relationships	6 (6.1)
P6) Moving viewport horizontally to locate content or read information	6 (6.1)
P7) Zooming out for overview and then zoom in for detail	6 (6.1)
P8) Moving viewport horizontally to scan links in top navigation bar	5 (5.1)
P9) Moving viewport vertically to scan links in left hand navigation bar	4 (4.1)
P10) Moving viewport to explore upper limit of a region of content, and then moving viewport to explore lower limit	4 (4.1)

Partially sighted participants would often align their screen magnifier viewport to an edge of a content area (e.g. top navigation bar) and then move along that edge, either vertically or horizontally. Sometimes, this would be very easy, such as situations where they investigated the left navigation bar where the edge of the webpage would keep the magnifier aligned (sequence P9 in Table 2). In other cases, participants manually tried to follow an edge of an arbitrary region of the web page to investigate content on the page (P3 and P6). We observed that at the beginning of tasks, partially sighted participants quickly moved their screen magnifier down the page all the way to the bottom in order to get an overview of the whole page structure (P2).

Interestingly, the action sequence composition with the largest proportion for partially sighted participants was the use of tool-tips for Exploration. They used this form extensively in order to understand the purpose of a link or control, sometimes in the absence of any other contextual information in the viewport of the screen magnifier.

Partially sighted participants would often check areas of websites where they find content on other websites (P4). For example, participants would check the upper right hand corner of a site for a search form, or look for navigation links on the left hand side of the page.

The most frequently occurring action sequence forms elicited from dyslexic participants involved navigation around the website and are presented in Table 3.

Table 3. Most frequent action sequence compositions for dyslexic participants

Strategy	Instances N (%)
D1) Carefully selecting links based on link label and link probing	25 (23.8)
D2) Returning to the website homepage and beginning task again	14 (13.3)
D3) Checking web convention locations for expected content	12 (11.4)
D4) Keyword searching within page, or within site	7 (6.7)
D5) Random probing of links	6 (5.7)
D6) Looking for feedback that actions have had desired effect or information on what errors have occurred	6 (5.7)
D7) Looking for familiar icons	4 (3.8)
D8) Scrolling to bottom of the page looking for content	4 (3.8)
D9) Checking the top of the heading area of the page above the navigation bar, and then the footer of the page	4 (3.8)
D10) Scanning links in left navigation bar	3 (2.9)

Of particular interest in the results for dyslexic participants was that they employed a variety of action sequences in support of Navigation. Dyslexic participants were very selective about what links they would follow for completing their task (sequence D1 in Table 3). Often this was through careful reading of a set of link labels and occasional, but purposeful, probing of links to investigate where they led. In other cases, when dyslexic participants became unsure of how to proceed with their task, they would select links at random to probe in the hopes of finding something useful (D5). Dyslexic participants also readily returned to the home page of a website when they became lost during a task (D2).

A number of action sequences used by dyslexic participants related to familiar or expected webpage content (D3). For example, dyslexic participants looked for content in locations that adhere to web conventions (e.g. a link to the user's account being in the top-right hand corner of the page), which was quite similar to partially sighted participants. Where dyslexic participants had an idea of what they were looking for, they would often use keyword search (D4), either within a webpage using the browser search controls or across webpages using the website search facilities.

5 Discussion

The results indicate that blind, partially sighted and dyslexic Web users all apply a wide variety of different action sequences that map onto seven different types of strategy to support their interaction on the Web. The strategies that were most frequently observed in the study were those that supported finding content within a website (Navigation), obtaining an overview of the contents of a webpage (Discovery) and understanding the meaning of individual pieces of content (Exploration). Interestingly, there were cases where action sequences, such as heading traversal, could be used for more than one strategy type: Exploration or Discovery, depending on the situation.

Many of the top strategies applied by blind and partially sighted participants match those that were elicited in the Theofanos and Redish studies [14, 15] nearly a decade ago. In some ways, this is positive as it gives us confidence in the methodology used in this study. On the other hand, it is very odd that after a decade of experience of Web development, during which both assistive technologies and the Web itself have evolved and supposedly improved, these two user groups are still largely doing the same things they were when using much less sophisticated technology. This could be interpreted as users sticking with strategies that they know work. Alternatively, it could be argued that interaction design for blind and partially sighted people has stagnated because we do not fully understand how they interact with the web and how they understand webpage content and structure.

The emphasis by blind participants on understanding individual pieces of content through Exploration strategies suggests that many blind Web users may work from a bottom-up approach in building their mental models of websites. The information gathered through use of Exploration strategies is then supplemented by Discovery strategies to understand an individual webpage in a more top-down manner. The intensive use of links and headings has important implications for the design of both websites and assistive technologies.

Current screen readers provide functionality to access a list of headings on the webpage. This functionality presents users with the headings and their level (e.g. h1, h2) in the order they occur on the webpage. Users who are building a mental model bottom-up will work from the individual headings and then integrate them into an overall model of how the webpage is structured. The current screen reader functionality supports, to some extent, both Exploration and Discovery strategies. However, if the structure of the headings in a website is poor, there will be a mismatch between a user's mental model that is constructed from the individual headings and the intentions of the designer. This shows the importance of emphasizing to web developers the use of good heading structures. It also lends support to helping users through the application webpage transcoding to add or correct headings on webpages [8, 15].

However, in our results traversing links was as common as traversing headings. Yet, in current screen readers, the support for forming an understanding of how the links relate to the rest of a page is almost non-existent. When users interact with links in screen readers it is usually through functionality that presents all of the links on the page sequentially from the beginning to the end. All structure and context for the links is lost. This functionality supports Exploration, but not Discovery. This has implications for future research and the development of screen readers. Specifically, how can we support Discovery strategies for blind users and help them build an appropriate mental model of a webpage? It is possible that new assistive technology features could be designed to do this, but there are also likely opportunities for changing the design of web content. The web is becoming more flexible in how it is implemented, especially with the advent of responsive design, and taking advantage of that to better support the strategies found in this study needs to be investigated urgently.

When looking at the action sequences of partially sighted users, the use of Discovery strategies by screen magnifier users is certainly understandable. Enlarging text and other content comes at the cost of how much of the screen a user can see, and therefore they must reconstruct the overall page. However, the action sequences used by partially sighted users provide insights into future designs. Going beyond previously reported results, we observed users working with edges and borders of content to constrain where they moved their magnifier viewport. Moving vertically on the edge of a screen was relatively easy, while moving horizontally proved to be more difficult. In order to improve the effectiveness of those action sequences, we may be able to transform, or enhance, websites to better support following those edges. One solution may be to employ current responsive design approaches to websites, where the horizontal dimension of a website can be shrunk, moving the content into a single long narrow column. This might result in a very long webpage, which would need to be augmented further with links that allow partially sighted users to move directly to sections of content that are of interest, or back to the top of the page.

Dyslexic participants employ a range of strategies in reasonably equal proportions. One aspect of the most frequently applied strategies by dyslexic participants is that many relate to moving from one page to another. Very selective Navigation strategies, returning to home pages when lost and the use of keyword searches to find content in a site are all indicators the importance of good navigational support on websites to these users.

One interesting aspect of all user groups is how many of their strategies relied on having designs that are externally consistent between websites. Both blind and dyslexic participants used keyword searches to find words that regularly appear in websites (e.g. login, search), or words commonly used in the website domain (e.g. a bank sort code). Further, partially sighted and dyslexic participants used strategies that relied on controls or content being in “typical places” on websites. The data supports the idea that good general design practices related to consistency for usable websites will benefit all users, including users with print disabilities.

6 Conclusions

In this paper we have presented a study of the interactions that people with print disabilities, specifically blind, partially sighted and dyslexic users, have with websites. This study has produced a number of important findings that add to the very thin body of empirical literature that exists about users with disabilities interacting with the Web.

The key result that comes from this study is a set of 7 key strategy types that are applied by print disabled users when interacting with web applications. These strategies provide a framework, in which we can analyse the interactions of users, above the level of individual operations in their user agent, or assistive technology, which was the focus of most previous research. Using these strategies, we have identified a number of interesting qualities about the differences in how users from different user groups approach websites.

In general, blind users appear to exhibit far more instances of applying strategies than their partially sighted or dyslexic counterparts. They have a heavy reliance on action sequences supporting the Exploration strategy. This type of interaction implies that blind users are not actively seeking out information about webpage structure, but instead letting the structure emerge.

In comparison, partially sighted users applied action sequences that were for active Discovery of webpage structure. Further, while dyslexic users have no one strategy that they prefer over others, they have a wide range of strategies that they apply.

There are very clear, distinguishing features for each of these user groups regarding how they interact with the Web. However, they all share a large number of web accessibility problems [1]. The implication of our findings on strategies is that: it is very unlikely that there is one solution that can be prescribed to solve any given problem that is encountered by all user groups. As a result, the current approach of avoiding accessibility problems in web design is not going to be sustainable in the future. As we begin to understand more deeply the interactions of users, we need to define design principles and user-validated design patterns that support the strategies of different user groups. Once we have such a set of design principles, we can begin to personalize web applications for people with disabilities in order to provide them with a truly equal experience.

However, before we can do that, we need a great deal more empirical research about disabled users and their interactions on the Web. Our future research will examine how different action sequences and their strategies are combined within the interface during user interactions. It is hoped that these strategy combinations will reveal important insights into the contributory causes of the problems users encounter on the Web, and point the way towards solutions for different user groups.

Acknowledgments. The research was conducted as part of the Inclusive Future - Internet Web Services project. We are grateful to European Commission (FP7; Grant: 257623) for part funding the project and all the project partners for their hard work and support. We thank all the participants in the study for their time and effort.

References

1. The Web: Access and Inclusion for Disabled People. Disability Rights Commission (2004)
2. Berry, J.: Apart or apart?: access to the Internet by visually impaired and blind people, with particular emphasis on assistive enabling technology and user perceptions. Information Technology and Disabilities 6(3), 1–16 (1999)
3. Bigham, J.P., Cavender, A.C., Brudvik, J.T., Wobbrock, J.O., Lander, R.E.: WebinSitu: a comparative analysis of blind and sighted browsing behavior. In: 9th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 51–58. ACM, New York (2007)
4. Borodin, Y., Bigham, J.P., Dausch, G., Ramakrishnan, I.V.: More than meets the eye: a survey of screen-reader browsing strategies. In: 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), pp. 1–10. ACM, New York (2010)
5. Coyne, K.P., Nielsen, J.: Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities - Design Guidelines for Websites and Intranets Based on Usability Studies with People Using Assistive Technology. Nielsen Norman Group (2001)
6. Engelen, J.J., Bauwens, B.: Large Scale Text Distribution Services for the Print Disabled: The Harmonisation and Standardisation Efforts of the TIDE-CAPS Consortium. In: Rehabilitation Technology: Strategies for the European Union: Proceedings of the 1st Tide Congress, 22 pages. IOS Press (1993)
7. Lazar, J., Allen, A., Kleinman, J., Malarkey, C.: What frustrates screen reader users on the web: a study of 100 blind users. International Journal of Human-Computer Interaction 22(3), 247–269 (2007)
8. Lunn, D., Harper, S., Bechhofer, S.: Identifying Behavioral Strategies of Visually Impaired Users to Improve Access to Web Content. ACM Trans. Access. Comput. 3(4), 1–35 (2011)
9. O'Reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software. Communications & Strategies 1, 17 pages (2007)
10. Paciello, M.: Making the World Wide Web Accessible for the Blind and Visually Impaired. Florida Libraries 39(4) (1996)
11. Paciello, M.: Making the Web Accessible for the Deaf, Hearing, and Mobility Impaired. Florida Libraries 39(5), 83–91 (1996)
12. Petrie, H., Kheir, O.: The relationship between accessibility and usability of websites. In: 2007 SIGCHI Conference on Human Factors in Computing Systems, pp. 397–406. ACM, New York (2007)
13. Petrie, H., Morley, S., McNally, P., Graziani, P.: Authoring hypermedia systems for blind people. In: IEEE Colloquium on Authoring and Application of Hypermedia-Based User-Interfaces, pp. 6/1–6/3 (1995)
14. Power, C., Freire, A., Petrie, H., Swallow, D.: Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In: 2012 SIGCHI Conference on Human Factors in Computing Systems, pp. 433–442. ACM, New York (2012)
15. Takagi, H., Saito, S., Fukuda, K., Asakawa, C.: Analysis of navigability of Web applications for improving blind usability. ACM Trans. Comput.-Hum. Interact. 14(3), 13/1–13/36 (2007)

16. Theofanos, M.F., Redish, J.: Bridging the gap: between accessibility and usability. Interactions 10(6), 36–51 (2003)
17. Theofanos, M.F., Redish, J.: Helping Low-Vision and Other Users with Websites That Meet Their Needs: Is One Site For All Feasible? Technical Communication 52(1), 9–20 (2005)
18. Trewin, S., Richards, J., Bellamy, R., John, B.E., Thomas, J., Swart, C., Brezin, J.: Toward Modeling Auditory Information Seeking Strategies on the Web. In: 2010 Extended Abstracts on Human Factors in Computing Systems, pp. 3973–3978. ACM, New York (2010)
19. Watanabe, T.: Experimental evaluation of usability and accessibility of heading elements. In: 2007 International Cross Disciplinary Conference on Web Accessibility (W4A), pp. 157–164. ACM, New York (2007)

Participatory Design with Blind Users: A Scenario-Based Approach

Nuzhah Gooda Sahib, Tony Stockman, Anastasios Tombros, and Oussama Metatla

School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom
{nuzhah,tony.stockman,tassos,oussama}@eecs.qmul.ac.uk

Abstract. Through out the design process, designers have to consider the needs of potential users. This is particularly important, but rather harder, when the designers interact with the artefact to-be-designed using different senses or devices than the users, for example, when sighted designers are designing an artefact for use by blind users. In such cases, designers have to ensure that the methods used to engage users in the design process and to communicate design ideas are accessible. In this paper, we describe a participatory approach with blind users based on the use of a scenario and the use of dialogue-simulated interaction during the development of a search interface. We achieved user engagement in two ways: firstly, we involved a blind user with knowledge of assistive technologies in the design team and secondly, we used a scenario as the basis of a dialogue between the designers and blind users to simulate interaction with the proposed search interface. Through this approach, we were able to verify requirements for the proposed search interface and blind searchers were able to provide formative feedback, to critique design plans and to propose new design ideas based on their experience and expertise with assistive technologies. In this paper, we describe the proposed scenario-based approach and examine the types of feedback gathered from its evaluation with blind users. We also critically reflect on the benefits and limitations of the approach, and discuss practical considerations in its application.

Keywords: scenario, participatory design, visually impaired users.

1 Introduction

When designing accessible interfaces, it is crucial for the designers to make sure that their understanding of the problem is aligned to the users' experience of their interactions with the interface. When designers interact with systems using different senses, devices and interface widgets than to the target population, it can be difficult for them to exclusively depend on their expertise to correctly imagine the needs of the users and to conceptualise the users' interactions with the system. Thus, designers have to be particularly sensitive as to how the users perceive technology [1].

As a result, design paradigms such as "Inclusive Design" [2][3], "Design for All" [4] and "User Sensitive Inclusive Design" [5] have been proposed to encourage

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 685–701, 2013.

designers to include non-standard populations such as older and disabled users in the design process. These approaches aim to give an effective voice to users in the design process and enable designers to develop real empathy towards users to ensure they communicate design ideas in an accessible form.

In this paper, we propose the use of scenarios for participatory design with blind users. We use a scenario, expressed as a textual narrative, as a basis for dialogue between designers and users in the design of a search interface. There are two levels to our approach to participatory design: firstly, we include a blind user with knowledge of assistive technologies as a full member of the design team and secondly we use the scenario and a dialogue-based interaction to gather formative feedback from 4 blind users.

The details of the finished search interface and its evaluation are reported in [6] and are not detailed in this paper. Instead, we address the question of how to successfully engage blind users in design, given that the majority of tools used for early stage prototyping by developers, such as wireframes and paper prototypes, contain barriers to participation by blind users. We focus on describing the steps in creating the scenario and its evaluation as a means of engaging blind users in the design process. Therefore, the contributions of this paper are three-fold:

(i) We propose a participatory approach based on a textual narrative scenario and a dialogue-based interaction to engage blind users in the design process.
(ii) We evaluate our approach with blind users and describe the types of feedback that we gathered.
(iii) We reflect on our approach outlining its benefits, challenges and the practical experiences that we gained from applying it so that the approach can be reused or further developed.

The rest of this paper is structured as follows: in Section 2, we describe the use of scenarios in user-centred design and discuss approaches for engaging blind users in the design process. In Section 3, we outline the steps involved in developing the scenario-based approach and we discuss the user evaluation of the proposed approach in Section 4. We reflect on the approach in Section 5, outlining the benefits, challenges and practical experiences.

2 Related Work

To the best of our knowledge, in the context of interface design, there is no reported research on the use of scenarios for participatory design with blind users. Thus, in this section, we discuss how scenarios have been used in usability engineering and we describe approaches to participatory design with blind users.

2.1 Using Scenarios in Design

The use of scenarios in the early stages of the design cycle involves designers using a description of people (actors) and their activities (tasks) to help potential users to envision an interface that will be developed in the future [7, p. 46]. Scenarios consist of a plot, including a sequence of actions and events, which help to emphasise and explore the goals that a user might adopt and pursue.

Scenarios enable rapid communication among different stakeholders and thus, scenario-based design approaches are iterative and lightweight for envisioning future use possibilities [8]. As a result, designers can work through ideas rapidly, obtaining feedback and refining their ideas to make quick progress. Scenarios focus the design efforts on use, that is, what people will use the interface for and how they will use it [9]. This compels designers to maintain a consideration for people and their needs, as opposed to focusing only on the technology.

Apart from their use in framing the design rationale, scenarios have been used in HCI for other purposes, namely, for planning and evaluating test tasks and to specify usability goals. In [10], Bodker highlighted how scenarios can be used at different times with different purposes and described three ways of using scenarios in usability work, namely to generate ideas during field studies, as a starting point in design workshops and for usability testing of prototypes. Scenarios also have a natural and inherent ability to support participatory design as they allow users to identify themselves as the actors in the scenario and to reflect on their own ideas and their implications in the context of design. In this way, scenario-based approaches provide a common language for discussions among users and designers [9].

Newell and McGregor [11] suggested a story-telling approach with older and disabled users to gather information and data about accessibility issues by using scenarios in the narrative form. In this respect, [12] used scenario-based drama to elicit user requirements in the design of a fall detector for elderly people. Four scenarios were developed which were performed by a theatre group and filmed. These videos were then used to engage elderly people in the design process by provoking discussions about the use of the system. Other examples of the use of scenarios to engage users in the design process are: [13] used scenarios to understand user requirements in the design of a location-based feedback notification system for users with mobility impairments and in the design of digital technologies for older users, [14] used video prompts of a scenario about the problem domain for participatory design with users who were in the 65+ age group.

2.2 Participatory Design with Blind Users

Participatory design with blind users can be challenging as designers have to ensure their methods of communicating design ideas with users are appropriate and effective to gather useful feedback. In [15], Okamoto reported about a workshop where scenarios were used by visually impaired and sighted students to discuss products being designed to enhance day-to-day activities for visually impaired users. However, no details of the implementation of the scenario-based method were given, so it is difficult to understand how scenarios were used in that context. In another setting, [16] conducted a workshop including round table discussions and demonstrations of early prototypes to engage visually impaired users in the design of a system to represent diagrams in sound.

Prototyping is also a common way of brainstorming design ideas with users, but for obvious reasons, visual prototyping techniques are not appropriate for blind users and therefore, alternatives have been proposed: [17] describes haptic paper prototypes (using cardboard mockups) while a tactile paper prototyping approach (with Braille and tactile graphics mockups) was discussed in [18]. Also, [19] proposed 2 types of haptic mock-ups for visually impaired children consisting of cardboard models and Braille-labelled plastic artefacts.

However, these techniques are time consuming to set up and changes are not easy to make in response to feedback. Both methods proposed in [18] and [19] exclude the significant proportion of the blind population who are not Braille readers and are also only suitable to prototype haptic interaction as opposed to speech-based screen reader interaction. Also, the cardboard and plastic abstract models such as those used in [19] have a possible drawback of not allowing users to fully conceptualise the application as a whole, since users only interact with individual artefacts at a time.

In the following section, we describe our approach to engaging blind users in the design of a search interface. As well as including a blind user in the development team, we used a textual narrative scenario in a participatory design setting as a basis for dialogue to simulate interaction, in order to discuss design ideas with potential users.

3 Using Scenarios for Participatory Design with Blind Users

In this section, we describe the development of the proposed scenario-based approach. In Section 3.1, we explain the rationale for our overall approach to participatory design and in Section 3.2, we outline the steps included in developing the proposed scenario-based approach to verify requirements with blind users in the design of a search interface.

3.1 Rationale for the Participatory Design Approach Taken

To access the Web, blind users depend on a screen reader, a software application that by default reads web pages linearly from left to right, top to bottom, rendering the content in computer synthesised speech or Braille. Blind users also use the keyboard to navigate web pages and position the focus of the screen reader to read parts of the page of interest. Typical commands supported by screen readers include web page navigation forward/backward by headings (at different levels), forms, frames, edit fields, buttons and links. The linear rendition of text by screen readers plus the fact that they do not represent the spatial layout of web pages, such as columnised format, means that the mental models of blind users can vary significantly from those of sighted users [20].

There is a parallel to be drawn here between web navigation and navigation of real world spaces. Given due consideration, it is unlikely that when giving directions to a pedestrian, the way in which one would describe those directions would be the same for a sighted pedestrian as for a blind pedestrian. Instructions to the sighted pedestrian are likely to exploit visual cues, to be given at a granularity level appropriate to someone who can take in their surroundings at a glance. On the other hand, directions to a blind pedestrian, if they are to be useful, should be in terms of landmarks that are detectable by them, and at a level of granularity related to the way in which they interact with their surroundings, given whichever mobility aid they might employ, be it a dog or a white cane etc. Similarly, within human-computer interaction, in order to be useful, the way in which interactions are articulated need to take into account the senses and tools at the disposal of the user, as well as the level of granularity at which they interact with the system.

Based on this need to embed an understanding of how end users interact with the system at a deep level and the fact that other members of the development team can not easily share that experience (using a screen reader with a covered screen is not a realistic surrogate for a blind user with thousands of hours of screen reader experience [21]), it was decided that participatory design should be addressed at two levels.

Firstly, we included a blind user with knowledge of assistive technology as a full member of the design team. This provided the development team with immediate feedback in discussions about the development of appropriate inter- face artefacts, for example, properly labelled controls, the types of interactions supported by screen readers (the use of screen reader commands for web page navigation) and the appropriate vocabulary with which to describe interactions to blind users, for example, keystrokes rather than mouse clicks.

This understanding of how screen reader interaction works led to the development of a scenario and a dialogue about it being pitched at an appropriate level to make sense to a screen reader user. For example, the interface comprised several different components such as a search box, to which the user would frequently want to navigate. In this case, knowledge of screen reader interaction suggested that the appropriate way for this to be achieved should be through a keyboard shortcut and that an appropriate means of confirming that the action has been executed could be through playing a non-speech sound.

The second level at which participatory design was achieved was through the recruitment of 4 blind participants who took part in prototyping sessions to provide formative feedback to the design team. In these sessions, the overall scenario was used as the basis of dialogue about how users would interact with the system using a screen reader and the usefulness and usability of proposed interface features.

3.2 Creating the Scenario-Based Approach

During the requirements verification stage of the design process, the requirements of a system are analysed and validated to ensure that the designers and the users share the same understanding of the problems that were identified during the requirements gathering stage. For participatory design approaches, at this stage, designers communicate early design ideas to users to gather feedback [22].

In [23], we identified user requirements for a search interface for blind users through an observational study and in this paper, we verify requirements by using a scenario expressed as a textual narrative which then formed the basis of dialogue between the designers and the users. Basing this dialogue on a narrative scenario evoked a form of role play which worked well because the human mind is adept at overloading meaning in narrative structures [7, p. 54] which are meant to stimulate the imagination [9] and to provoke new ideas [10]. Therefore, they are well suited for use in participatory approaches to engage users early in design.

Our approach is a hybrid one involving a combination of participatory design [24] and the use of a detailed scenario to discuss ideas with target users [7]. The participation of a blind user as a member of the design team was invaluable when developing the scenario and its associated textual narrative as it helped us to conceptualise how potential users will interact with the system, given their use of screen readers. It also allowed us to establish the level of detail at which the scenario should be discussed with end-users. In Figure 1, we provide a broad overview of the framework we followed to implement the scenario-based approach and in the following we describe how we implemented each step of the framework in the design of a search interface for blind users. For each step, we also highlight the contributions of the blind member of the design team.

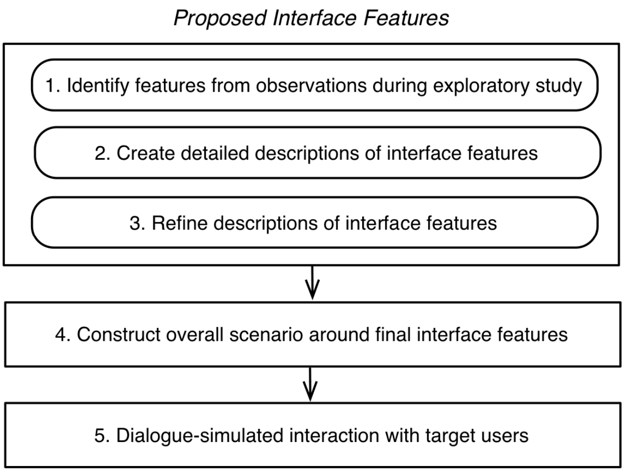


Fig. 1. Framework for the scenario-based approach

Step 1: Identify set of interface features. From our observations with blind users [23], we identified a set of search interface features that could support blind searchers during information seeking on the Web. These interface components were chosen to address the difficulties observed in [23] and were influenced by the design team's intuitions and knowledge of search user interface components. During this process, the blind team member contributed significantly from his knowledge and experience of using both graphical interfaces (via screen readers) and self-voicing auditory interfaces. This, to some extent, allowed the sighted designers to conceptualise the mental model they had to follow when designing interface components.

Step 2: Create detailed description of features. To communicate ideas for the search interface, we created detailed descriptions for all interface features. As we were using a textual narrative scenario, we had to ensure that our ideas were being conveyed correctly to the users and therefore, we focussed significantly on describing each interface component. For example, we proposed an integrated note-taking feature within the interface as described in the following:

> *Searchers can create a note and the system asks them where they would like to save this note and to give it a name. The note is divided in two parts: The first part of the note is editable by the searcher, that is, they can type ideas, copy and paste things from web pages etc. The second part cannot be edited and is used to save search results automatically by the system.*

In this step, the contributions of the blind team member were significant in discussing the functionality of suggested support features as well as how searchers would interact with them. Through such discussions, we could ensure that interaction components were appropriate, and that the correct vocabulary was being used to describe interface components.

Step 3: Refine description iteratively. To ensure that the users shared the same understanding of the proposed features as the designers, we iteratively refined the description of the features through several informal conversations with the blind team member. For example, one idea was that a context menu might be useful to provide access to a set of options that become available when exploring individual search results. The idea of doing this through a context menu was contributed by the blind team member, who highlighted that context menus were familiar interaction artefacts to most screen reader users.

The blind team member further contributed that the way to initiate the interaction with end-users about the context menu should be by telling them to use a key combination (Shift+F10), rather than right clicking, as the keystroke is the usual way a blind user will initiate the interaction compared to the right mouse click familiar to sighted users. Therefore the blind team member gave us both an appropriate interaction artefact, and the most fitting means of describing the interaction to end- users. The options to be made available through the context menu were then identified and the best ways of implementing and describing the interactions to end-users were then refined through discussions between the sighted designers and the blind team member. The following was the final description used for the context menu:

> *You are aware that this new interface has a menu associated with each search result so that you can open, save, email and copy results. You hit the menu key and you find the following options in this particular order (Save Result, Copy, Email, Open). This is rather like the context menu you have in Windows that you bring up using Shift+F10.*

To enhance our textual description of some search support features, we referred to examples from other popular interfaces such as Google Search (results presentation with title, short description and web address) and Windows (context menu) that the users would be familiar with. These familiar points of reference helped the blind users to better envision the proposed search interface.

Step 4: Construct scenario around interface features. Once we finalised which search support features to include on the interface, we created an overall scenario, like a story, with a specific setting whereby the user was using a search interface for the first time after hearing about it from a friend. As we were also evaluating a new history mechanism and interface features to support searchers in resuming search tasks [25], the scenario included a stage where the user had to leave the task midway to attend an important appointment. When constructing the overall story, it was also essential to ensure the story included all suggested interface features in the correct order and in a reasonable sequence. For example, we would not describe a feature for managing search results before the users were asked to submit their first query.

Step 5: Dialogue-simulated interaction with the scenario. After the overall scenario was constructed, we used it as the basis of a formative evaluation with potential users (described in Section 4). The evaluator conducted a dialogue with each potential user to simulate the interaction with the interface components proposed in the scenario. At each step of the interaction, the evaluator would describe the interface feature to the user and explain how to interact with it. Then, the evaluator would ask the user for feedback on the feature and they would discuss the alternative interaction paths resulting from multiple design ideas.

4 Evaluating the Use of Scenarios for Requirements Verification

We evaluated the approach through a dialogue with potential users using the scenario to simulate interaction between the users and the yet-to-be constructed interface. Our goals for the evaluation were two-fold. Firstly, we wanted to verify the requirements for a new information-rich search interface for blind users, to communicate and discuss design ideas with potential users early in the design process. Secondly, our aim was to evaluate the use of scenarios in a participatory design setting for engaging blind users in the design process. Hence, in Section 5, we discuss the benefits, challenges and practical experiences of using the approach proposed in this paper.

4.1 Participants

We recruited 4 blind users through word of mouth and via online email lists. The participants were experienced searchers who rated their proficiency with assistive technologies from intermediate to advanced. Three of the participants were educated to a postgraduate level while one had professional qualifications in IT. In Table 1, we provide additional demographic information about these participants.

Table 1. Demographics for all participants

Age	37 years
Gender	M *(3)* F *(1)*
Search Experience	12 years
Screen Reader	JAWS *(3)* VoiceOver *(1)*
Frequency of Computer Use	Daily *(3)* Weekly *(1)*
Use of Online Search Engine	Daily *(3)* Weekly *(1)*

4.2 Procedure

For each evaluation session, we used a standard script of the final scenario (described in step 4 of Section 3.2) to ensure that the users and ourselves shared the same understanding of the requirements for a new search interface. To begin with, the evaluator asked the user to think of a search task to complete. We left the choice of task open to elicit greater participation and user engagement with the scenario. The choice of search task did not affect the use of the script as it was built so that its primary focus was on interaction with individual interface components and thus could be adapted to any search task.

During the session, the evaluator who was the one in charge of the script, started the conversation with the participant by conducting a walk-through of the scenario in line with the script (The beginning of the dialogue-simulated interaction between the user and the evaluator is illustrated in Table 2). At each step, the evaluator provided the user with complete descriptions of the search interface feature and the user was prompted for feedback. The evaluator and the user also discussed how each interaction would work, including alternative interaction paths in case of multiple design ideas as shown in Table 3.

Table 2. Excerpt on query specification

Evaluator: Your friend has told you about a new search system and you would like to try it out for yourself to see how good it really is. Think of something you would like to search on this new system. Once you have chosen your search task, you type the address of this new search inter- face in your web browser and you reach the page with the cursor in the search edit box. “What do you type as a query?”
User: digital rights accessibility
Evaluator: You type this query and hit enter. If you misspell a word in your query, the system will specify which term you misspelt and allow you to submit a corrected version of your query.

4.3 User Feedback

In this section, we describe the feedback gathered from 4 blind users during the evaluation of the scenario-based approach. The dialogue-simulated interaction between the designers and the users allowed us to gather feedback from users through their comments and critiques of suggested design plans. Additionally, we also use our own observations to categorise the user feedback as presented in the following:

Verifying Requirements. The scenario-based approach allowed us to verify requirements for a search interface for complex search tasks. We identified the requirements for a search interface during an observational study with 15 visually impaired searchers [23]. Through the proposed approach, we were able to ensure that the design team and the target users shared the same understanding of the difficulties faced by blind users when using current search interfaces.

In this respect, we were able to, for example, ascertain that spelling suggestions were a source of difficulties for searchers as the way misspelt words are rendered on current interfaces is not intuitive for screen reader users. One of the users said: *"we hardly notice which term is misspelt. It would be good if the system clearly said which term is wrongly spelt"*.

This is because when spelling suggestions are presented to the user, the way in which the screen reader pronounces the suggestion for the correctly spelt word is often not detectably different from the pronunciation of the original misspelling, and so it is not clear what error is being corrected. In this case, the blind searcher can navigate to the suggestion and cursor character by character through it to find the difference, but this process loses all the immediacy of the visual representation. This was a difficulty that we had observed in [23].

Likewise, we were able to verify user requirements for a new history mechanism. In our scenario, we proposed a search history mechanism that would keep track of the queries submitted and the search results visited by the searcher. Participants in the evaluation commented on the need for such a history feature saying *"I do not like the history in IE, this is more powerful than history. It allows you to call it up and instantly be back to where you were, in the same context"* and *"It is nice to pick up from where we left because sometimes we use keywords which are useful and then forget the right combination"*.

Identifying Issues with Current Design Ideas. By discussing design plans with potential users early in the design process, we were also able to identify issues with proposed design ideas from the users' perspective. Such discussions proved to be beneficial; for example, one of the design ideas for search results presentation included limiting the display of individual results to only one line per result on the search results page. Our reasoning for this idea was that it would reduce the amount of text that screen reader users would have to go through. However, we found that participants in the evaluation did not welcome this idea, as they would rather have some context about the search results retrieved by the search engine. They felt that if there was only one line per result, there would not be enough context to decide about the relevance of a specific result among those that the search engine had retrieved.

During the dialogue-simulated interaction, the evaluator also had the opportunity to probe users on factors like keyboard navigation, which plays a central part in the user experience of blind searchers. Keyboard navigation is significantly different from visual navigation and hence, the design team have to ensure that all proposed interface features were intuitive to access via the keyboard.

In describing how they would interact with the interface, participants would often refer to how they would use the screen reader to access the proposed features. About the grouped approach for results presentation, one participant said *"Along the lines of how VoiceOver works, this grouping on the page would be good"* and another questioned how they would navigate back to a previous page *"Would I need to use the screen reader key for this or would there be a special key combination?"*.

New Design Ideas Proposed by Users. Using the scenario-based approach allowed us to engage users in the development process via their interaction with a yet-to-be constructed search interface. Through this process, users came up with ideas of their own to enhance the design of some of the features that were being proposed. For example, in the scenario, we described a note-taking feature, which could be used by searchers to automatically save search results or to make notes of their own.

The initial idea was to allow users to then download or email the note in a text format. However, one of the participants highlighted that the benefits of having an integrated note facility could be enhanced by structuring the note and by including HTML tags to allow users to easily get back to any previously accessed web pages. About the same note-taking feature, another participant augmented our basic definition of the feature with his own design ideas, suggesting, *"I can see where you are going with this, it could be in two panes, your browser and your search notes"*. The user was in fact proposing that there should be two separate areas on the interface, one for regular browser-related activities such as submitting queries and viewing web pages and the other area should be dedicated for note-taking and other search management activities. When users suggested such design ideas of their own, we discussed them with the design team including the blind co-designer to ensure that such an approach would be feasible and would enhance the users' experience.

4.4 Discussion

Engaging non-standard populations such as the elderly and disabled users in the design process is challenging as traditional user methodologies are not always effective at capturing the real user requirements. Therefore designers often have to explore different methods or adapt existing ones to ensure that such users can be successfully included in the design process [4][14].

In this paper, we described an approach, which included 2 levels of participatory design: we included a blind user in the design team and also carried out prototyping sessions with 4 blind users. Involving a blind team member who can combine a good knowledge of assistive technologies with an end-user perspective enabled us to create a scenario that was better matched to the vocabulary and interactions familiar to blind users. Thus, we could successfully engage blind users to solicit their feedback in the design of the search interface.

Table 3. Excerpt on alternative search results presentation

Evaluator: There are alternative ways of presenting the search results retrieved:

1. **Standard approach:** Results are presented in a list with each result described using a title, a short summary and a web address. Each of these items is on a separate single line.
2. **Simplified standard approach:** Results are presented in a list, but each result is described in one line, with a title and a short summary.
3. **New approach:** Similar results are grouped together and you are presented with an overview of each group of search results. For example, results that deal with similar topics will be grouped together. If you are doing a travel task, web pages describing things to do at your destination will be grouped together and another group of pages could be about possible places to stay. If you would like to explore one of these groups, you can select the group and it will open in a different window and will contain all search results in that group described with title and a short summary. You can always return to the first window to browse through other result groups.

"What are your thoughts on these results presentation alternatives? Which one would you prefer and why?"

Search interfaces are highly interactive and to progress in their search task, searchers are required to perform activities such as formulate queries and view search results etc. The scenario-based approach described in this paper allowed us, to some extent, to simulate this interaction through a dialogue between the user and the designer. During the dialogue-based interaction, users were involved in the scenario and were constantly informed about their evolving interaction, for example, how search results are being handled and the alternative paths available to them. This approach to interaction elicited a high level of participation and engagement from the users, as evidenced by the feedback received. Therefore, such use of dialogue was beneficial as a model of engagement [26] and a model for effective communication and collaboration [27] between the designers and the potential users.

Overall, the findings gathered from the dialogue-simulated interaction showed that users had no problems in 'imagining' the interface proposed in the scenario [7]. The narrative was successful in evoking the search experience in users and therefore, they were able to discuss the proposed features for inclusion in the interface in the context of their use within the scenario, and to discuss alternative interaction sequences where they arose. The fact that users were able to go beyond the described interface features to question how they would interact on a relatively low level (screen reader keystroke level) is evidence that they were able to successfully form a mental model of the search interface that was yet-to-be constructed.

In addition, involving blind users at such an early stage allowed designers to identify limitations with their own design ideas. Participants would often question the practicality of the proposed interface features, requiring detailed explanations of how these interface components would be accessed in a realistically usable way with screen readers. Identifying these limitations at that stage ensured that no further development effort was put into interface features that would not meet the needs of the users, or would raise difficult usability issues.

The benefit of an inclusive approach, such as the one proposed in this paper, is that it enables users, especially those with disabilities, to become involved in the process of design and formative evaluation. This involvement in the development process encourages users to speak about their experiences with search interfaces and to contribute to design ideas and hence, the user truly becomes the centre of the design process. User-generated ideas during the scenario walk-through resulted in valuable contributions to our design plans. This is so because the participants in our study were experts at navigating the Web through screen readers and given their experience, they had better insights into how the overall search interface and the individual components would be perceived.

5 Reflections on the Use of Scenarios for Participatory Design with Blind Users

In this section, we reflect on the scenario-based approach and its evaluation with potential users. We discuss the benefits, challenges and practical experiences of using scenarios to engage blind users in the design process.

5.1 Benefits

Scenarios are flexible and adaptable and thus they can be customised according to the needs and abilities of the user group, for example, as a scenario-based drama for the elderly [12]. For our project, we created a textual narrative scenario for a dialogue-based interaction with blind users. The value of the scenario was that it allowed blind users to envision the proposed interface and form a mental model of how they would interact with it. This was important to correctly verify user requirements with blind users and also to rapidly communicate design ideas.

In addition, scenarios are adaptable in the level of detail that they convey to the user group, which can assist in enabling them to envision the proposed artefacts. For our approach, given our focus on requirements verification, we provided detailed descriptions for the proposed interface features and less detail about the interaction or the way certain tasks could be completed when using the search interface. For example, when describing a new search history mechanism, we fully described the items such as the queries and visited results that would be recorded as history, but we did not explicitly tell users how they would navigate the trail at a keystroke level. Instead, during the sessions, the users themselves wondered and discussed how they would interact with this history mechanism for different types of tasks.

In this way, we were able to achieve the comparable ‘unfinished look’ of handwritten mock-ups that Snyder [28] claims encourages creativity during low- fidelity paper prototyping. However, depending on the users’ needs and the stage of the design process, a scenario-based approach could be used for more high-fidelity prototyping to evaluate how users would interact with the proposed artefacts. Our discussions with participants regarding how some interface features could be accessed through screen readers show that the use of scenarios is likely to be effective for such high fidelity prototyping.

In the absence of visual aids to communicate design ideas, sighted designers are likely to describe graphical user interfaces in a way that makes references to visual aspects of the interface, such as layout, structure etc. For the blind user, these descriptions would not be useful and would not convey a helpful representation of the interface features. For this reason, the involvement of a blind user was crucial to ensure that we used the right vocabulary and context to describe interactions at an appropriate level from the user's perspective.

Scenarios, especially when expressed as narratives, have an inherent ability to support participatory design [9][29] and thus complemented the level of participatory design reported in this paper. In such settings, scenarios furthered the communication between the users and the designers to enable successful collaboration [27]. We expressed the scenario in the form of a textual narrative (which was then used as the basis of a conversation between designers and users) and this enabled blind users to comment on the proposed design ideas in the context of screen reader access, as well as to suggest their own ideas for new or modified interface components.

5.2 Challenges and Practical Experiences

In the absence of visual aids, the designers in this approach relied on the textual descriptions of the interface features to communicate design ideas to the users. Therefore, the detailed descriptions played a significant role in shaping the mental model that users created of the interface. Using a standard script for the scenario ensured that variations in the way the interface was conceptualised were limited.

Our approach focussed entirely on the functionality of interface components and the way to interact with them. No efforts were directed towards conveying spatial information, which despite not necessarily being of primary importance to blind users, plays a role in how screen reader users perceive an interface, and very importantly, their collaborative use of the interface with sighted peers [20]. As an extension to this work, it will be interesting to examine the benefits and drawbacks of incorporating screen reader technology within the prototyping process, rather than the purely conversation-based approach taken here. It is unclear whether the incorporation of screen reader technology will enhance the realism of the interactions, and/or whether it may detract from the free flow of the dialogue about the interactions and their possible alternatives by overburdening the audio channel [30].

The approach proposed in this paper was a first attempt at using scenarios to engage blind users in the design process and hence, we identified some important points to consider for any future implementation or extension of this approach. Firstly, we expressed the scenario in a textual medium, with a dialogue-simulated interaction between the user and the designer. This audio-based approach works well with blind users, but as is common with audio interfaces, there is a lack of persistence. Therefore, any artefacts that are part of the scenario should be described in significant detail to ensure that users can conceptualise and "picture" the proposed design. Visual aids such as paper mock ups convey significant contextual information even in their most early versions and any attempt at replicating these types of approaches for blind users should be constructed using low-level details in the textual descriptions. Detailed descriptions can also be complemented with references to similar existing artefacts to convey as much contextual information as possible.

From the user evaluation, we concluded that scenarios, especially those expressed as a narrative, should be highly interactive to include the user as much as possible. Given that scenarios are stories about people and their activities, it is essential for users to feel part of the scenario to maximise their ability to envision the proposed interface. In the scenario, we regularly prompted users for feedback by asking them to think of a search task, by asking them for their query terms and by allowing them to choose the next step of their interaction etc. When scenarios are textual narratives and interaction with the user is dialogue-based, the designer will be speaking for relatively long periods to describe different parts of the interface. Therefore, to replicate an interactive search experience, users should be active 'actors' in the scenario activities to further user engagement.

Involving a blind person on the design team helped in many ways, but it is important to be aware of the dangers of over-relying on one person as a representative of a population. For example, the blind co-designer in our team was congenitally blind and had a lot of experience using JAWS with Windows and Internet Explorer to perform searches using Google, but only a passing knowledge of other screen readers, browsers and search engine combinations. Therefore, it is important to try to ensure relevant diversity [14], that is, users involved in the prototyping process, together with members of the design team, should provide a wide coverage of the range of tools and assistive technologies that might be used with the system being designed. It is also important to include users with less experience, as they will also be representative of members of the target population.

6 Conclusion

In this paper, we proposed a participatory approach based on a textual narrative scenario, tailored to the abilities of blind users to engage them in the design process. We evaluated the proposed approach with blind users and described the types of feedback we gathered in a participatory design setting through the use of a scenario as a basis for dialogue. The dialogue-simulated interaction between designers and users was effective in evoking the search experience in users and thus they could envision the yet-to-be constructed interface.

We achieved two levels of participatory design, namely by including a blind user in the design team and by carrying out prototyping sessions with 4 blind users. The contributions of the blind team member were invaluable to ensure that, in constructing the scenario, we used the right vocabulary and context to describe interactions at an appropriate level for screen reader users. In this paper, we also reflected on the benefits and challenges of our proposed approach and the practical experiences we gained in applying it so that it can be reused or further developed.

We believe that the proposed approach opens an interesting discussion on the ways to adapt current tools and techniques in user-centred design when designing for non-standard populations such as the elderly or users with disabilities. In this respect, future work could also focus on the comparison between different techniques to investigate the best ways of engaging users with disabilities in design depending on the stage of the design process and the type of feedback required.

References

1. Newell, A., Carmichael, A., Morgan, M., Dickinson, A.: The use of theatre in requirements gathering and usability studies. Interacting with Computers 18(5), 996–1011 (2006)
2. Clarkson, J.: Inclusive Design: Design for the Whole Population. Springer (2003)
3. Imrie, R., Hall, P.: Inclusive design: designing and developing accessible environments. Taylor & Francis (2001)
4. Newell, A., Gregor, P.: User sensitive inclusive design – in search of a new paradigm. In: Proc. of the Conference on Universal Usability, pp. 39–44 (2000)
5. Newell, A., Gregor, P., Morgan, M., Pullin, G., Macaulay, C.: User-sensitive inclusive design. Universal Access in the Information Society 10(3), 235–243 (2010)
6. Gooda Sahib, N., Tombros, A., Stockman, T.: Evaluating a search interface for visually impaired users. Submission to the Journal of the American Society for Information Science and Technology (2013)
7. Carroll, J.M.: Making use: scenario-based design of human-computer interactions. The MIT Press (2000)
8. Rosson, M.B., Carroll, J.M.: Scenario-Based design. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 145–165. Taylor & Francis (2007)
9. Jarke, M., Tung Bui, X., Carroll, J.M.: Scenario management: An interdisciplinary approach. Requirements Engineering 3(3), 155–173 (1998)
10. Bodker, S.: Scenarios in User-Centred design - setting the stage for reflection and action. In: Proc. of the Annual Hawaii International Conference on System, pp. 3053–3064 (1999)
11. Newell, A., Gregor, P.: Design for older and disabled people - where do we go from here? Universal Access in the Information Society 2(1), 3–7 (2002)
12. Marquis-Faulkes, F., McKenna, S., Gregor, P., Newell, A.: Scenario-based drama as a tool for investigating user requirements with application to home monitoring for elderly people. In: HCI International, Crete, pp. 512–516 (2003)
13. Bhatia, S., Dahn, C., Lee, J.C., Sampat, M., McCrickard, D.S.: VTAssist: a location-based feedback notification system for the disabled. In: Proc. of the Annual Southeast Regional Conference, pp. 512–517 (2006)
14. Lindsay, S., Jackson, D., Schofield, G., Olivier, P.: Engaging older people using participatory design. In: Proc. of CHI, pp. 1199–1208 (2012)
15. Okamoto, M.: Possibility of participatory design. In: Proceedings of the International Conference on Human Centered Design, pp. 888–893 (2009)
16. Metatla, O., Bryan-Kinns, N., Stockman, T., Martin, F.: Designing for collaborative cross-modal interaction. In: Proc. of Digital Futures (2011)
17. Brewster, S., Magnusson, C.: Guidelines for haptic Lo-Fi prototyping. In: Proc. of the NordiCHI, vol. 102, pp. 1483–1491 (2008)
18. Miao, M., Köhlmann, W., Schiewe, M., Weber, G.: Tactile paper prototyping with blind subjects. In: Altinsoy, M.E., Jekosch, U., Brewster, S. (eds.) HAID 2009. LNCS, vol. 5763, pp. 81–90. Springer, Heidelberg (2009)
19. Tanhua-Piiroinen, E., Raisamo, R.: Tangible models in prototyping and testing of haptic interfaces with visually impaired children. Guidelines for Haptic Lo-Fi Prototyping (2008)

20. Stockman, T., Metatla, O.: The influence of Screen-Readers on web cognition. In: Proc. of the Accessible Design in the Digital World Conference (2008)
21. Petrie, H., Morley, S.: The use of non-speech sounds in non-visual interfaces to the MS-Windows GUI for blind computer users. In: Proc. of the ICAD (1998)
22. Pekkola, S., Kaarilahti, N., Pohjola, P.: Towards formalised end-user participation in information systems development process. In: Proc. of the Conference on Participatory Design, pp. 21–30 (2006)
23. Gooda Sahib, N., Tombros, A., Stockman, T.: A Comparative Analysis of the Information Seeking Behaviour of Visually Impaired and Sighted Searchers. Journal of the American Society for Information Science and Technology 63(2), 377–391 (2012)
24. Muller, M.J.: Participatory design: The third space in HCI. In: Handbook of HCI 2nd edn., USA (2007)
25. Morris, D., Ringel Morris, M., Venolia, G.: SearchBar: a search-centric web history for task resumption and information re-finding. In: Proc. of CHI, pp. 1207–1216 (2008)
26. Wright, P., McCarthy, J.: Experience-centered Design. Morgan and Claypool Publishers (2010)
27. Anderson, R., Baxter, L.A., Cissna, K.N.: Dialogue: Theorizing Difference in Communication Studies. Sage Publications (2004)
28. Snyder, C.: Paper prototyping: the fast and easy way to design and refine user interfaces. Morgan Kaufmann (2003)
29. Luck, R.: Dialogue in participatory design. Design Studies 24(6), 523–535 (2003)
30. Chandrashekar, S., Stockman, T., Fels, D., Benedyk, R.: Using think aloud protocol with blind users. In: Proc. of ASSETS, pp. 251–252 (2006)

An Evaluation of Stacking and Tiling Features within the Traditional Desktop Metaphor

Clemens Zeidler, Christof Lutteroth, and Gerald Weber

University of Auckland, 38 Princes Street, Auckland 1010, New Zealand
czei002@aucklanduni.ac.nz, {lutteroth,g.weber}@cs.auckland.ac.nz

Abstract. Having many open windows on the desktop can lead to various usability problems. Window content may get occluded by other windows and working with multiple windows may get cumbersome. In this paper, we evaluate the idea to integrate stacking and tiling features into the traditional desktop metaphor. For this purpose we introduce the Stack & Tile window manager, which allows users to stack and tile arbitrary windows into groups that can be moved and resized similar to single windows. To evaluate if stacking and tiling can improve productivity, we conducted an experimental evaluation. We found that participants were able to perform various multi-window tasks and switch between tasks significantly faster using Stack & Tile. Furthermore, we found that the time to set up a Stack & Tile window group is reasonably low. Stack & Tile is open-source and has been used for over two years now. To evaluate its usefulness in practice, we conducted a web-based survey that reveals how people are actually using the new stacking and tiling features.

Keywords: window manager, tabbing, usability, evaluation.

1 Introduction

In the traditional desktop metaphor, windows can float freely on the desktop and are allowed to overlap each other. One problem that arises when having multiple windows open at the same time is that the content of some windows may get partially or fully occluded. To make the content visible again the user has to bring the occluded window to the front or move other windows aside. This can make the management of overlapping windows tedious and time consuming [7] and task management can become another task [18].

An alternative approach to overlapping windows is a tiling window manager where windows are tiled beside of each other. This approach can be superior to the overlapping approach [6, 16]. There is also the idea of stacking windows on top of each other [3]. Tabbed interfaces became very popular in web browsers, but are widely unused in window management. Despite these alternatives and extensions, all main desktop operating systems are still primarily using the traditional overlapping window approach.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 702–719, 2013.

In this paper we investigate the benefits of stacking and tiling features in a traditional overlapping window manager. To analyze such features, we first present Stack & Tile, which is an extension of a traditional window manager. Stack & Tile allows users to stack windows on top of one another and tile windows beside of each other (Figure 1). In this way the user is able to create window groups (Stack & Tile groups) consisting of arbitrary windows from different applications, which can be controlled similar to single windows. Window groups can still overlap other windows or window groups, so Stack & Tile integrates seamlessly into the traditional desktop.

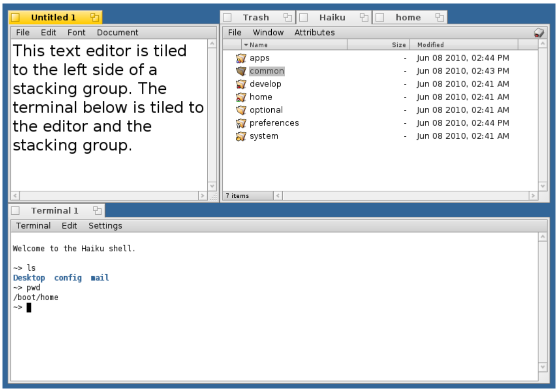


Fig. 1. Stack & Tile group. On the right side three windows are stacked into a stacking group. Tiled to this group are on the left side a text editor and at the bottom a terminal.

To answer the main question in this study, whether stacking and tiling within an overlapping window manager bring any benefits to the user, we conducted a controlled experiment to determine in which use-cases Stack & Tile performs better than a traditional window manager. We looked at use-cases where the user is working with documents of the same or of different applications, and a use-case where data is exchanged between documents of the same application, and designed experimental tasks accordingly. Another task measured the time needed to switch between different groups of windows. We found that for all tasks Stack & Tile performed significantly faster. Furthermore, the setup time used to create a Stack & Tile group is acceptably low. Most participants stated that Stack & Tile makes window management easier and more enjoyable.

Stack & Tile is already integrated in the open-source operating system Haiku[1], and thus is already exposed to a large group of developers and users. This allowed us to target another interesting question: How are stacking and tiling features used and accepted by real users? In a web-based survey we asked the Haiku community about their opinions and experiences of Stack & Tile. From 146 responses we got a detailed insight into how, how often and for what applications Stack & Tile is used.

[1] Haiku Operating System, `www.haiku-os.org`

1.1 Contribution

In this work, we analyze the potential of stacking and tiling features in the traditional desktop metaphor. In particular, we present:

1. A concept for integrating stacking and tiling unobtrusively into a traditional window manager.
2. A controlled experiment that shows how Stack & Tile performs for different use-cases.
3. A web-based survey that gives insight into how Stack & Tile is used by real users.

Section 2 introduces the Stack & Tile window manager. Section 3 describes how Stack & Tile is different from other work on window managers. In Section 4 the experimental evaluation is presented. Section 5 presents the web survey of how Stack & Tile is used. The paper closes with a conclusion in Section 6.

2 The Stack and Tile Window Manager

In Stack & Tile, users can stack and tile windows together to create window groups according to their needs. Groups behave like single windows, and can be used together with other windows as usual. Stack & Tile combines the advantages of stacked and tiled windows with the freedom of overlapping windows.

This combination can help users to manage their windows more effectively. Tiled windows in an active Stack & Tile group are always visible and not occluded by other windows. As shown later, this makes it easier to exchange data between windows. A window stack can be used to group windows together whose content does not need to be visible at the same time. Grouping windows used for a certain task together in a Stack & Tile group can result in a cleaner desktop and facilitate switching between tasks involving multiple windows.

Internally linear constraints are used to describe a Stack & Tile group, employing the tabstop and area system of the Auckland Layout Model (ALM) [17]. Areas are simply the tiles where windows can be placed, and tabstops are their borders. For example, two stacked windows are sharing the same tabstops and are therefore always getting the same size. The minimum size of a window is specified using a hard constraint while the maximum and current window sizes are specified using soft constraints. Window operations modify the constraint specifications that describe the Stack & Tile groups, and the window manager solves the specifications and re-renders the windows.

Currently, there is a fully working Stack & Tile implementation available, which is well-known and appreciated in the Haiku OS community. It is showcased regularly at large open-source conferences, and has been integrated into the default Haiku OS user interface.

2.1 Stack and Tile Operations

Stack & Tile offers two simple operations that can be used to connect windows: First, stacking, which makes use of the tab-like appearance of the Haiku OS window title bars; secondly, tiling of windows, which means that windows are arranged beside each other. A Stack & Tile operation can be triggered by holding down the Stack & Tile key, which is by default the Windows key, and dragging a window near to another window (see Figure 2). Releasing the Stack & Tile key or dropping the window finally executes a Stack & Tile operation. The dragged window is called the *candidate window*, and the window that it is dragged to is called the *parent window*. In this manner, *Stack & Tile groups* can be created.

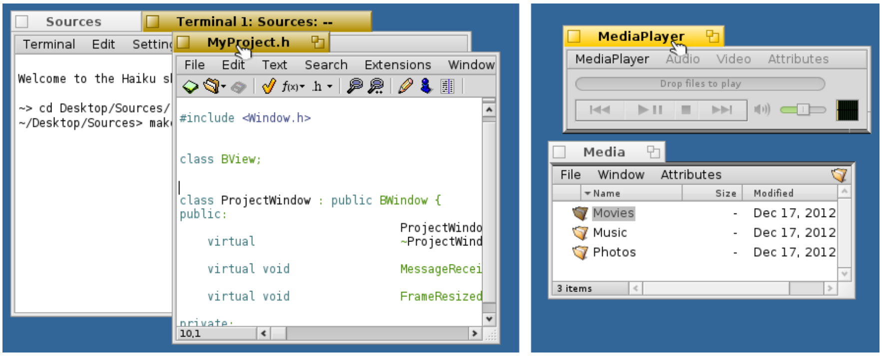


Fig. 2. Moving a window while holding the Stack & Tile key initiates stacking or tiling. Affected window tabs and borders are highlighted in gray. Left: The editor window is going to be stacked on top of the Terminal. Right: MediaPlayer is going to be tiled to the Media folder.

Stacking: Briefly, a stacking operation is triggered by moving the title tab of the candidate window onto the tile tab of a parent window while holding the Stack & Tile key. To be more precise, the candidate window has to be dragged by the title tab so that the upper edge of the candidate window tab is on the parent window tab, and the x-position of the mouse cursor is in the x-range of the parent title tab. When a valid stacking candidate-parent pair is found, the window title tabs of both windows are highlighted (left of Figure 2).

After stacking windows on top of one another, the stacked windows have the same position and size. The title tabs are automatically arranged beside each other, so that the stacked windows are accessible over a tab interface at the top of the stack. The result is comparable with a tab bar, e.g. in a tabbed web browser, with a similar functionality for reordering tabs.

Tiling: If the Stack & Tile key is pressed, dragging and dropping a candidate window border close to one or more parent window borders triggers the tiling operation (right of Figure 2). Tiled windows always share a border position. For example, when tiling two windows horizontally, the right border of the left window always has the same position as the left border of the right window. Furthermore, the top and bottom of both windows are aligned with each other. Windows can be tiled to any free rectangular region of an existing group, and can span the width and height of several other windows in the group. Figure 1 shows an example of a tiled window group.

Removing from a Group: A window can be removed from a Stack & Tile group by holding down the Stack & Tile key and dragging the window away from the group. After removing the window from the group, the window behaves just like an ordinary window in the desktop metaphor. In case the removed window was the only connection between other windows in the group, the group is split into several independent groups.

2.2 Traditional Window Management Operations

When interacting with a Stack & Tile group, the semantics of the traditional window management operations change slightly. Stack & Tile applies window management operations to multiple windows, which has already been considered to be helpful in Elastic Windows [16].

Activating one window in a Stack & Tile group raises all windows in the Stack & Tile group. Only the window that triggered the group operation gets the input focus.

Moving a window by a certain offset also moves all other group windows by the same offset. This means windows in a Stack & Tile group keep their relative position to each other.

Resizing one window in a Stack & Tile group leaves all windows in the group aligned to each other. For example, windows that are tiled to a resized window are moved or resized accordingly. This is done by temporarily setting a high priority for the size constraint of the resized window. In this way, the resized window gets its new size and the other windows adapt according to the solution of the constraint system.

Hiding or **showing** a window also hides or shows all other windows in the group. Thus, all windows in a Stack & Tile group are either hidden or shown.

3 Related Work

Novel techniques for overlapping windows such as snapping windows to a master window or organizing them in tabs have already been proposed in 2001 [3]. However, they have not been evaluated for traditional window managers.

A comparison between overlapping, stacked and “piled” windows found that tabbed interfaces are doing well when it comes to finding a document [14]. The study only looked at windows of the same application, while our study also looks at tasks involving windows of different applications (in contrast to windows of the same application, they are not grouped in the taskbar). Furthermore, they did not consider the presence of windows that were not part of the task at hand.

The Google Chrome[2] browser can stack web pages running in different processes, and extends the usage of the tab interface to stack different instances of the same application. Stack & Tile is more general and is working not only with windows of a particular application, but with arbitrary windows of arbitrary applications.

[2] Google Chrome, `www.google.com/chrome`

Roughly one year after the first release of Stack & Tile, a similar stacking feature was implemented in the KDE desktop environment[3]. Although KDE has an optional Notion-like tiling window manager (see below), a combination of stacking and tiling in the standard KDE desktop is not possible.

Tiling window managers allow users to tile windows beside each other. Windows are arranged so that they do not overlap each other and are aligned without gaps and fragmentation. In general, windows are arranged automatically by the window manager, using the whole screen, but the user has the ability to rearrange the window tiling using different layouts [4, 10, 11, 16, 19]. For example, the tiling window manager Notion[4] allows the user to rearrange the tile layout manually, and multiple windows can be stacked into one tile. By comparison, Stack & Tile integrates seamlessly with the traditional desktop metaphor. In Windows 7, two windows can be tiled horizontally using the whole screen space. However, more complex tiling layouts with more than two windows, as in Stack & Tile, are not possible.

Tiling windows can help the user to organize their windows better. Already in old studies it has been shown that tiled windows can be superior to overlapping windows in certain use-cases. For example, if all window content fits into the allocated tiles, the tiling approach leads to shorter task completion times [6]. Another study found that the completion time is lower for tiled windows when comparing information sources in multiple windows and scanning through windows of a certain window group [16]. Stack & Tile leverages the advantages of tiled windows, but integrates tiling with the traditional overlapping window management.

It is quite common that users work on different task in parallel and switch back and forth between different windows [20]. Typically windows from different tasks are overlapping each other, and task switching involves the manipulation of many windows. One approach to address this is to let the user group windows by task [16, 21]. Another approach is to analyze the user activities and assign windows to tasks automatically [5, 13, 18, 21, 23]. For example, WindowScape [21] creates a timeline of previous window configurations, and also allows users to assign windows to tasks manually. This is similar to an activity centered desktop [2] where applications can be assigned to activities. Also Stack & Tile can be used to group windows by task. Additionally, Stack & Tile helps users to optimize the layout of window groups.

To reduce the probability of overlapping existing windows on the screen and to group windows by task, the concept of multiple virtual workspaces can be used [15]. Another option is to increase the physical workspace by attaching multiple monitors to a computer [12]. However, often windows overlap other windows, and some approaches use alternative methods to make overlapped windows temporary available without losing the focus on the active window [8, 9]. For example, an occluded window can be made visible by folding the overlapping window back like a piece of paper [9].

The Scwm window manager [1] gives the user the opportunity to specify relations and constraints for each window. For example, it is possible to set a relative distance

[3] KDE Desktop Environment, `www.kde.org`

[4] Notion Window Manager, `http://notion.sourceforge.net`

between two windows or set the minimum or maximum size of a window. With this approach it is also possible to tile windows beside each other. However, because constraints in Scwm are on a lower level of abstraction, tiling is not as easily accessible as in other tiling window managers and Stack & Tile.

4 Experimental Evaluation

In this section we compare Stack & Tile with a traditional window manager in a controlled experiment, using tasks that are representative of certain use-cases that involve multiple windows (*multi-window tasks*). Stack & Tile makes it possible to group windows with direct manipulation operations, and offers window management operations that affect all the windows in a group. This can be used to manage occlusion and quickly switch between groups of windows related to different tasks. An interesting question is how much time can be saved using Stack & Tile, and if the time taken to set up a Stack & Tile group is reasonable compared to the time saved. We formulate the following hypotheses about Stack & Tile for this experiment:

- **H1** The task completion time for multi-window tasks is lower with Stack & Tile than with a traditional window manager.
- **H2** Switching between tasks that each involve multiple windows is faster with Stack & Tile than with a traditional window manager.
- **H3** The time necessary for setting up Stack & Tile groups is acceptable compared to the time that can be saved by using Stack & Tile.
- **H4** Users prefer Stack & Tile over a traditional window manager.

4.1 Methodology

We conducted a within-subjects study, i.e. each participant performed two runs of tasks: one run with Stack & Tile and one run without Stack & Tile. To avoid order bias, the order of the runs was alternated between participants. Furthermore, there were two different sets of similar tasks for each run. Also the order of the sets of tasks was permuted. Each run included a window setup phase followed by three question tasks and a group switching task.

To start with, each participant got an introduction to the Haiku window manager, its taskbar and Stack & Tile. After that the participant had time to get familiar with the system. Then each participant performed a training run, which was similar to the main runs but with shorter tasks. Participants were allowed to use different methods to switch between windows, e.g. Alt+tab or the taskbar.

Each run was guided by instructions shown in a small instruction window that was placed at the bottom-right screen corner on top of all other windows (see Figure 3). The size of the instruction window was chosen to be as small as possible to not interfere with other windows on the screen. At the beginning of each task, the instruction window was in fullscreen mode, so that the participant was not distracted by other windows on the screen. After the participant had internalized the instruction, she had to press the start button to start a timer. This resized the instruction window back to its smaller size at the bottom-right. For the question tasks, the instruction window also contained a question with a three-point multiple choice answer field. When the participant had completed the task, she had to press a finish button to stop the timer. Pressing the finish button adds a small and fairly constant offset to the task completion time, which does not affect the outcome of our comparison. Afterwards, the instruction window was shown in fullscreen mode again with the next instructions.

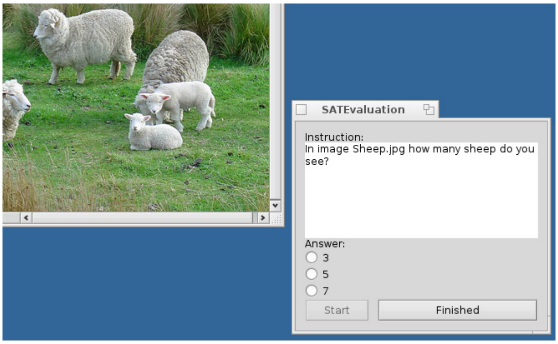


Fig. 3. The instruction window (right) with a question about a picture window (left)

Each task involved multiple windows, and all windows of all tasks (in total 12 windows) were open all the time. We think this is a realistic number of windows, but it is not unusual that users have even more windows open at a time [22]. We were interested in the efficiency of managing multiple windows and tasks, so we tried to minimize the time spent on activities unrelated to window management. Thus, the question tasks were designed to be very easy, but the answers were not guessable. The three question tasks were targeting H1, with each task addressing a different use-case.

Setup Phase. The first step of a run was to open all windows for all tasks. For each task, the participant got a description in the instruction window about what windows were involved in the task. Then the windows were opened automatically, so that the participant could get familiar with them before starting the task. When using Stack & Tile, there was an extra step where the participant had to stack or tile the windows, depending on the task. Here the *setup time* taken to create a certain Stack & Tile group was measured.

Task I. The first question task was about finding a picture in a set of five pictures and answering an easy question about it. Each picture was opened in a single window and there was one question for each picture (e.g. Figure 3). This targeted the use-case of working with multiple windows of the same application, where no data needs to be exchanged between the windows. It simulates typical lookup tasks, where a user needs to activate the window that contains a particular piece of information.

When it comes to finding a certain window, it can be helpful if all windows of interest for the task have been stacked previously. In this case, activating one window of the stack brings the whole group to the front, and the right window can be chosen through the tab interface. When using Stack & Tile, the participant was asked to stack all five pictures on top of one another to take advantage of this behavior. Without

Stack & Tile, the participant had to search for the right window on the desktop, or use the taskbar. Because the taskbar groups windows by application, it provides a similar grouping facility as a window stack, and it is unclear if Stack & Tile is advantageous.

Task II. The second question task was similar to Task I. Here, three windows were involved: a web page, a PDF document and a text document. This targeted the use-case of working with windows of different applications, where no data needs to be exchanged between the windows. As before, there was one question for each document, e.g. "What is the first word of the second paragraph in the web page?" In the Stack & Tile condition, the three windows had to be stacked first.

There are two reasons why it is expected that navigation in this more heterogeneous group is more difficult than in Task I. First, because windows are opened with different applications, they are not grouped together in the taskbar anymore. Secondly, because documents of different types also differ visually, users cannot rely as much on visual similarity as in Task I when associating a window with the task.

Task III. The third question task targeted the use-case of working with windows of the same application, where data needs to be exchanged between the windows. To simulate this use-case, a simple coordinate treasure map game was chosen. A treasure map is a 7x5 table, with each cell containing either a coordinate pointing to a cell in another map or a treasure. For example, the coordinate M2(D,5) points to map 2, cell (D,5). Starting from an initial coordinate, the participant had to visually follow the path through four different maps to find a "treasure". The treasure was always reached after three steps, and the path crossed all four maps in random order. When the treasure was found, the kind of treasure had to be selected in the instruction window. This had to be repeated three times, each time starting from a different coordinate in the first treasure map.

While this task may seem artificial, it does simulate real tasks where related information has to be collated from multiple sources. For example, a real task of that kind would be looking up the location of an appointment mentioned in an email from a calendar, and then checking a booking for that location. Many such tasks involve tabular information, as simulated by the treasure maps.

The four treasure maps were each opened in a text editor window. When using Stack & Tile, the participant had to tile the four editor windows beside each other in a 2x2 layout. In this way, it was possible to display all four maps without occlusion on the screen, i.e. all maps were completely visible.

Task IV: Group Switching. This task evaluated the efficiency of switching between different tasks (H2). All windows from the previous tasks were used, and the user was asked to bring the windows of each task to the front, one task at a time. Without Stack & Tile, this can be done by directly clicking the windows on the desktop, or by activating them from the taskbar. When using Stack & Tile, only one window has to be activated to bring the whole group to the front. We expected Stack & Tile to be clearly faster here; this task was included to shed light on how much faster it actually is. For the non-Stack & Tile condition , we expected a correlation between group size (between 3 and 5 windows per group) and activation time.

Questionnaire. After finishing both runs, the participant was asked to fill in a Likert-scale questionnaire and make some general comments about what they liked and disliked about Stack & Tile. Furthermore, the participant had to estimate their usage in percentages of the following window management techniques: taskbar, short cuts, direct window access, virtual desktops and other.

4.2 Results and Discussion

There were 30 participants with an average age of 31 ($\sigma = 6$). Six of them were female. Most of them (~ 70%) were software developers or students of Computer Science or Software Engineering. 15 of the participants were from the Haiku community and 7 of them had used Stack & Tile before. We found that users who had used Stack & Tile before performed only slightly better than those who had not, therefore we do not differentiate between these groups in the following.

For the analysis of the task completion times, the difference between the task completion time with Stack & Tile $t_{S\&T}$ and the task completion time without Stack & Tile $t_{noS\&T}$ was calculated: $\Delta t = t_{noS\&T} - t_{S\&T}$. A pairwise Wilcoxon rank-sum test was used to calculate the probability p_{wrs} for the null-hypothesis. The Wilcoxon rank-sum test was chosen because the task completion times were only roughly Gaussian-distributed.

For each question task, three answer options were given, with only one correct option. In only ~ 3% of all cases a questions was answered wrongly. Furthermore, there was no difference between the error rates with and without Stack & Tile. Thus we can say that it was easy for the participants to answer the questions. From our observations, the participants followed the instructions carefully. Even when they chose the wrong answer, they applied the necessary window operations. Because we are not primarily interested in how accurate a task was performed but in how Stack & Tile affects window management, wrong answers were not removed from the result set.

Task I Results (5 Pictures). Table 1 shows the average task completion time $t_{S\&T}$ under the Stack & Tile condition, the average time difference Δt, their standard deviation σ and the Wilcoxon rank-sum test p-value. For all five questions, Stack & Tile had significantly shorter task completion times. Thus we can accept H1 for Task I.

Answering the first question had the longest completion time. This can be explained by an additional step: first the Stack & Tile picture group had to be found and activated. After that participants were able to just select the next picture from the tab bar and the task completion time stayed roughly constant. The time difference Δt decreases from Picture 1 to Picture 5. A possible explanation is that for the non-Stack & Tile conditions, all pictures were moved to the front after some time and it became easier to find the right window. Another possible explanation is that the participants got into the routine of selecting the right picture from the taskbar.

Task II Results (3 Documents in 3 Apps). The results are shown in Table 2. For the web page question there was no significant completion time difference. However, for the two following questions, Stack & Tile had a significantly better performance.

Table 1. Task I completion times in [s]

	$t_{S\&T}$	σ	Δt	σ	p_{wrs}
Picture 1	12	5.7	5	11.8	0.01**
Picture 2	8	2.8	5	5.7	< 0.01**
Picture 3	8	4.3	3	6.7	< 0.01**
Picture 4	8	6.5	3	8.8	< 0.01**
Picture 5	7	3.3	2	4.5	< 0.01**

Table 2. Task II completion times in [s]

	$t_{S\&T}$	σ	Δt	σ	p_{wrs}
Web Page	17	7.7	-1	11.5	0.65
PDF	10	13.9	5	18.2	< 0.01**
Text file	9	4.9	8	10.9	< 0.01**

So why did participants have problems with the web page question while they were fine with the first picture question in Task I? A possible explanation is that the window with the web page was harder to activate than the windows with the pictures. When using direct activation by clicking on a window, all a participant had to do to select the Stack & Tile group of picture windows was to click any window with a picture on the screen, since there was only one group with pictures in it. Once the group was activated, the tab interface could be used to raise the right picture. However, there were two window groups with textual documents in it, so activating the window group containing the web page was not a as simple as clicking any window containing text. When using the taskbar to activate a window, for the window group with the pictures there was exactly one group of windows listed in the taskbar, as all the pictures were shown using the same application. However, the documents for Task II were all opened with different applications and hence there was a different entry in the taskbar for each of the windows. The taskbar gave no indication that the windows were grouped together, as for the five pictures. In fact, the taskbar obfuscated the Stack & Tile grouping of the windows, by grouping them by application together with other windows belonging to different Stack & Tile groups.

Hypothesis H1 cannot be accepted for the initial question of Task II. It is reasonable to assume that once a desired window group is activated, the advantages of Stack & Tile show more clearly. To facilitate this, as a future work, the taskbar could be changed to group windows by their Stack & Tile groups.

Task III Results (Treasure Maps). The results of Task III are shown in Table 3. When using Stack & Tile, finding the first treasure required a lot less time (21s). This time difference decreased to 7s and 5s for the second and third treasures. We can clearly accept hypothesis H1 for Task III.

This can be explained from the observations of the participants during the task. Without Stack & Tile, many participants first tried to position all treasure maps in a 2x2 grid to make them visible at the same time. Thus, finding the following treasures became easier for them.

However, finding the second and third treasures was still significantly faster using Stack & Tile. A reason for that is that many participants did not manage to align the treasure maps precisely without occlusion, and thus had to rearrange the windows later on. Moreover, some participants accidentally activated unrelated windows, which made more window operations necessary. This is consistent with the findings from Elastic Windows [16], were they found that setting up and working with overlapping windows becomes more difficult for an increasing number of windows.

Task IV Results (Activate Groups). As expected, activating all windows of a task is much faster using Stack & Tile (Table 4). This clearly supports hypothesis H2. However, no significant correlation between the number of windows of a task and the activation time could be detected. Here, the heterogeneous Group II (Task II) resulted in the largest $t_{S\&T}$. Activating a group containing only windows of the same application seems to be easier than activating a group containing windows of different applications. This is consistent with the results of Task II.

Table 3. Task III completion times in [s]

	$t_{S\&T}$	σ	Δt	σ	p_{wrs}
Treasure 1	23	15.6	21	22.7	< 0.01**
Treasure 2	19	8.6	7	14.5	0.01**
Treasure 3	15	5.1	5	8.1	< 0.01**

Table 4. Task IV: Time in [s] to activate the window groups of the Tasks I-III

	$t_{S\&T}$	σ	Δ	σ	p_{wrs}
Group I	6	2.5	12	5.6	< 0.01**
Group II	8	8.9	9	12	< 0.01**
Group III	4	1.3	7	4.8	< 0.01**

Stack and Tile Setup Time Results. To assess whether the time needed for setting up Stack & Tile groups is acceptable (H3), we compared the average setup time t_{setup} with the average saved time t_{saved} for each task (Table 5). t_{saved} is the sum of Δt for all questions of a particular task. The experimental tasks were artificial hence one could argue that such a comparison is not meaningful. However, all experimental tasks were quite short compared to real world tasks, so these numbers serve to indicate that t_{saved} is reasonably likely to outweigh t_{setup} when working with Stack & Tile groups a bit longer. From our observations, participants still had problems to set up Stack & Tile groups in an optimal manner after the training tasks. It is reasonable to assume that once users get more practice with Stack & Tile, the setup times will drop. A non-significant indication ($p < 0.25$) for this is that users who had used Stack & Tile before had a slightly shorter setup time (on average 2-3 seconds for each of the three tasks). Note that the setup time does not include the time users needed to decide how a Stack & Tile group should look like, because the layouts for the groups were given in the experiment.

Table 5. Setup times in [s] for the window groups of Tasks I-III

	Task I	Task II	Task III
t_{setup} (σ)	29 (15.7)	14 (11.1)	22 (10.9)
t_{saved}	18	12	33

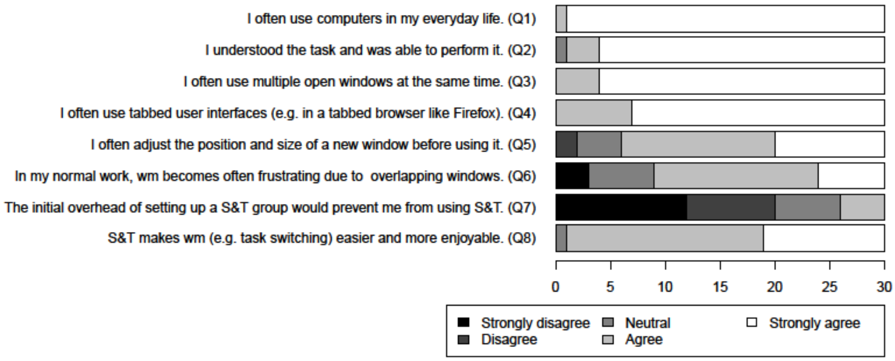


Fig. 4. Results for the Likert-scale questions

Questionnaire Results. Figure 4 shows the results from the Likert-scale questions. All participants agreed that they often use tabbed user interfaces (Q4), which shows that the stacking feature is not a fundamentally new concept to them.

There was an overall agreement that they often have multiple windows open (Q3), and over 2/3 of them agreed that window management often becomes frustrating when working with overlapping windows (Q6). This indicates that there is a need for better window management.

Only 3 of the 30 participants agreed that the initial overhead to setup a Stack & Tile group would prevent them from using Stack & Tile (Q7), and 24 of them agreed that they often adjust the position and size of a window before using it (Q5). Both supports H3 (see also Section 4.2).

Lastly, there was a general agreement that Stack & Tile makes window management easier and more enjoyable (Q8). This is also supported by many comments. As an example, one participant said: “Much easier to use. Better grouping, while still having more freedom than MDI apps”. This supports H4.

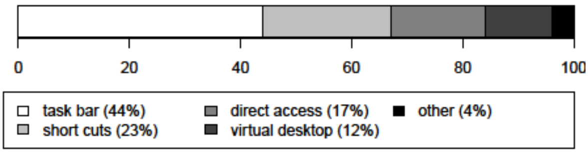


Fig. 5. Used techniques for window management

The results for the question about the techniques participants use to manage windows are depicted in Figure 5. The results are consistent with the empirical findings in [20]. Participants said that they use direct window access only 17% of the time. This is another indication that direct window activation is not perceived as optimal.

5 Web-Based Survey

Stack & Tile has been available in the Haiku OS for over two years now. An interesting question is how people are actually using Stack & Tile. To answer this question we conducted a web-based survey in the Haiku community. The questions were mainly a mix of Likert-scale questions and open questions.

The survey began with some demographic and general questions:
S1: I often use computers in my daily life.
S2: I heard about S&T before.
S3: I think S&T can be useful.
S4: I don't think there is any need for S&T.
S5: Have you ever tried S&T? (Denying this question ended the survey.)

Afterwards the participants were asked to upload screenshots of how they use S&T most frequently and describe them, followed by questions about S&T:
S6: How often do you use S&T?
S7: I think the stacking feature is more useful than the tiling feature.
S8: S&T helps with resizing window groups.
S9: Estimate the percentage of how often you use stacking and how often you use tiling.
S10: When you use S&T, how many S&T groups do you use on average at the same time?
S11: I exchange information between single windows (not in a S&T group).
S12: I exchange information within a single S&T group.
S13: I exchange information between multiple S&T groups.
S14: I exchange information between S&T groups and other single windows.
S15: Are there certain tasks where you use S&T? For example, programming or browsing.
S16: In what other situations are you using S&T?

Finally, participants had the opportunity to make suggestions and comments.

5.1 Result

During a period of two month we got 146 responses. The average age of all subjects was 32 (σ = 10). There was only one female participant. The majority of the participants was working in or had a degree in an IT-related field.

There were two groups of participants, distinguished by the general questions at the beginning: the group that had not tried Stack & Tile before (left of Figure 6) and the group that had tried Stack & Tile before (right of Figure 6). Generally both groups were heavy computer users. For the group that had not tried Stack & Tile before, almost half of them had heard about Stack & Tile. Most participants agreed that Stack & Tile is useful and disagreed that there is no need for Stack & Tile. This general opinion was much stronger for the group that had tried Stack & Tile before, indicating that Stack & Tile had made a positive impression.

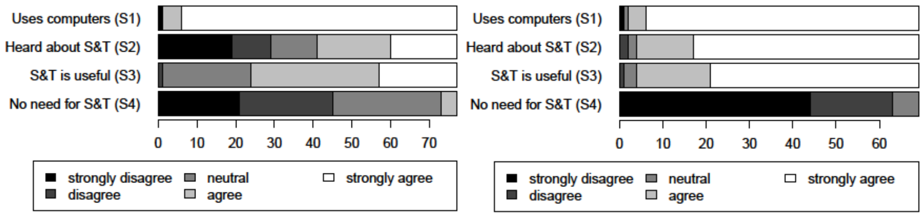


Fig. 6. Answers for S1 - S4. Left: users who had not tried Stack & Tile. Right: users who had tried Stack & Tile.

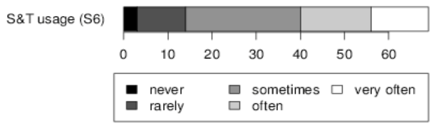


Fig. 7. Answers for S6

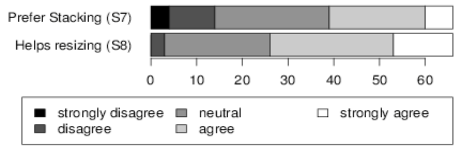


Fig. 8. Answers for S7 and S8

Figure 7 shows the results for question S6. Here we can see that most people who had tried Stack & Tile before were still using it.

Question S7 asked for preferences for the stacking or the tiling feature. There is a trend that the participants feel stacking is more useful than tiling (see Figure 8). S9 asked for the usage of the stacking and tiling features. The average estimated percentage of tiling is 45% ($\sigma = 29\%$) and for stacking 55% ($\sigma = 29\%$). Stacking is slightly more used than tiling with $p_{wrs} = 0.06$. This indicates that participants think that stacking is more useful (S7) and consequently use it more (S9). However, keeping in mind that stacking is much more common in today's applications (e.g. tabbed browsing), it is still interesting that the tiling feature is reportedly used that much.

S8 targeted the feature that windows in a group stay aligned when resizing one window in the group. Most of the participants judged this to be helpful (see Figure 8).

Figure 9 shows the average number of groups when Stack & Tile was used. Most participants used between one and four groups at a time, with the majority using more than one. This is quite interesting because it means they not only used it sporadically for a single task, but seem to have used it for different tasks in parallel.

We already analyzed the advantages of Stack & Tile when exchanging data between windows or Stack & Tile groups in the controlled experiment. S11 - S14 targeted the question in how far participants encounter such use-cases in practice (Figure 10). The results are quite similar for S11, S12 and S14. Most people are exchanging data between groups/windows, which indicates that the results from the controlled experiment are relevant for real users. The least frequent data exchange was between Stack & Tile groups (S13), which indicates that Stack & Tile groups are used to group windows of different tasks.

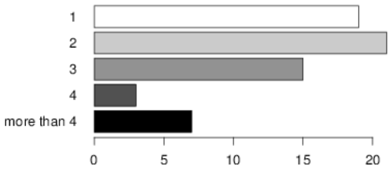


Fig. 9. Answer for S10: Number of S&T groups used at a time

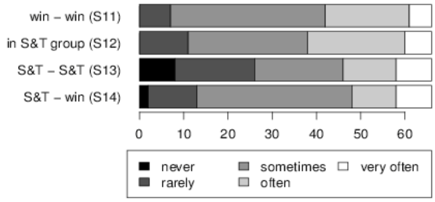


Fig. 10. Answers for S11 - S14: Exchanging data between windows

5.2 Use-Cases for Stack and Tile

We asked the participants to upload screenshots of how they use Stack & Tile most frequently, and we got 23 screenshots in response. Furthermore, one user sent a link to a video with a demonstration of how he uses Stack & Tile. S15 and S16, targeting the questions for what tasks and in what situations Stack & Tile is used, received 47 and 23 responses, respectively.

The results indicate that there are three main use-cases where the participants used Stack & Tile: programming (28 participants), web browsing (17 participants) and file management (11 participants). For example, C++ source and header files were tiled beside of each other, or all source files were stacked to one stack and header files to another stack. Many users reported that they use Stack & Tile to group windows by task, e.g. group web browser and chat window together. Other examples were grouping a music directory and a media player, a picture directory and an image viewer, or creating an ad-hoc development environment by grouping source files, source directory and terminal together. One user tiled a web browser and a text editor together to copy information from the browser to the editor. An unexpected use-case was the creation of a Stack & Tile group to move multiple windows across different virtual desktops more easily.

From the screenshots we observed that the Stack & Tile groups had mostly moderate complexity. There were only a few use-cases where more than two windows were tiled together. However, the stacked window groups had usually more than two windows in them.

There were 34 responses for the open-ended comments & suggestions field. Here participants had some ideas for improvements and better integration into the desktop. For example, there was the request to show Stack & Tile groups in the taskbar, and that it should be possible to store and restore Stack & Tile groups, especially on reboot. Applications that are using a tabbed interface should use the stacking feature instead. These are points we want to target in future work. One participant stated that he already started to replace the tabbed interface of a browser with the Stack & Tile stacking feature. 11 participants explicitly said that they like Stack & Tile, while one participant did not think that Stack & Tile can improve manual window management.

6 Conclusion

In this paper we investigated the advantages of integrating stacking and tiling features into the traditional desktop metaphor. Stacking and tiling can help to manage windows more effectively, for example by grouping windows by task. We presented Stack & Tile, a window manager which integrates stacking and tiling seamlessly with traditional window management operations.

In a controlled experiment, we found that stacking and tiling features can significantly improve completion times for tasks involving several windows (of the same application as well as of different applications). Furthermore, switching between different tasks was found to be much faster when windows were grouped by task. Setting up a Stack & Tile group is an initial overhead that may prevent users from using these features. However, the potential time savings as well as questionnaire answers indicate that the advantages outweigh this overhead.

In a web survey we investigated how often and how Stack & Tile is used in practice. There was a wide agreement that Stack & Tile can be useful, especially by participants who had used Stack & Tile. The stacking feature was perceived as being slightly more useful and also estimated to be used more than the tiling feature. We found that people were using the stacking and tiling features for a multitude of different use-cases. In a field for general comments many people wrote that they like Stack & Tile and suggested further ideas to integrate it more into the desktop. These ideas include future works such as grouping of windows by their Stack & Tile group in the taskbar and Stack & Tile group persistence.

References

1. Badros, G.J., Nichols, J., Borning, A.: Scwm: An extensible Constraint-Enabled window manager. In: Proc. of the FREENIX Track: 2001 USENIX Annual Technical Conference, pp. 225–234 (2001)
2. Bardram, J., Bunde-Pedersen, J., Soegaard, M.: Support for activity-based computing in a personal computing operating system. In: Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 211–220. ACM, New York (2006)
3. Beaudouin-Lafon, M.: Novel interaction techniques for overlapping windows. In: Proc. 14th Annual ACM Symposium on User Interface Software and Technology, p. 154 (2001)
4. Bell, B.A., Feiner, S.K.: Dynamic space management for user interfaces. In: Proc. 13th Annual ACM Symposium on User Interface Software and Technology, p. 248 (2000)
5. Bernstein, M.S., Shrager, J., Winograd, T.: Taskposé: exploring fluid boundaries in an associative window visualization. In: Proc. 21st ACM Symposium on User Interface Software and Technology, UIST 2008, pp. 231–234. ACM, New York (2008)
6. Bly, S.A., Rosenberg, J.K.: A comparison of tiled and overlapping windows. ACM SIGCHI Bulletin 17(4), 106 (1986)
7. Bury, K.F., Darnell, M.J.: Window management in interactive computer systems. ACM SIGCHI Bulletin 18(2), 65–66 (1986)
8. Chapuis, O., Roussel, N.: Metisse is not a 3D desktop! In: Proc. 18th ACM UIST, p. 22 (2005)

9. Chapuis, O., Roussel, N.: Copy-and-paste between overlapping windows. In: Proc. CHI, p. 210 (2007)
10. Funke, D.J., Neal, J.G., Paul, R.D.: An approach to intelligent automated window management. International Journal of Man-Machine Studies 38(6), 983 (1993)
11. Haraty, M., Nobarany, S., DiPaola, S., Fisher, B.: AdWiL: adaptive windows layout manager. In: Proc. 27th International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4177–4182 (2009)
12. Hutchings, D.R., Smith, G., Meyers, B., Czerwinski, M., Robertson, G.: Display space usage and window management operation comparisons between single monitor and multiple monitor users. In: Proc. of the Working Conference on Advanced Visual Interfaces, pp. 32–39 (2004)
13. Ishak, E.W., Feiner, S.: Content-aware layout. In: CHI 2007 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2007, pp. 2459–2464. ACM (2007)
14. Jakobsen, M.R., Hornbæk, K.: Piles, tabs and overlaps in navigation among documents. In: Proc. 6th Nordic Conference on HCI: Extending Boundaries, NordiCHI 2010, pp. 246–255. ACM (2010)
15. Henderson Jr., D., Card, S.: Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. ACM Transactions on Graphics (TOG) 5(3), 211–243 (1986)
16. Kandogan, E., Shneiderman, B.: Elastic windows: evaluation of multi-window operations. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pp. 250–257 (1997)
17. Lutteroth, C., Strandh, R., Weber, G.: Domain specific High-Level constraints for user interface layout. Constraints 13(3) (2008)
18. Oliver, N., Smith, G., Thakkar, C., Surendran, A.C.: SWISH: semantic analysis of window titles and switching history. In: Proc. 11th International Conference on Intelligent user Interfaces, p. 201 (2006)
19. Stewart, D., Sjanssen, S.: Xmonad. In: Proc. of the ACM SIGPLAN Workshop on Haskell Workshop, p. 119. ACM (2007)
20. Tak, S., Cockburn, A., Humm, K., Ahlström, D., Gutwin, C., Scarr, J.: Improving window switching interfaces. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 187–200. Springer, Heidelberg (2009)
21. Tashman, C., Edwards, W.K.: Windowscape: Lessons learned from a task-centric window manager. ACM Trans. Comput.-Hum. Interact. 19(1), 8:1–8:33 (2012)
22. Wagner, J., Mackay, W.E., Huot, S.: Left-over Windows Cause Window Clutter.. But What Causes Left-over Windows? In: Ergo'IHM 2012 - 24th French Speaking Conference on Human-Computer Interaction. International Conference Proceedings Series. AFIHM, ACM, Biarritz, France (October 2012)
23. Xu, Q., Casiez, G.: Push-and-pull switching: window switching based on window overlapping. In: Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI 2010, pp. 1335–1338. ACM, New York (2010)

Investigating Pointing Tasks across Angularly Coupled Display Areas

Fabian Hennecke[1], Alexander De Luca[2], Ngo Dieu Huong Nguyen[1], Sebastian Boring[3], and Andreas Butz[1]

[1] HCI Group / [2] Media Informatics Group
University of Munich (LMU)
Amalienstr. 17, 80333 Munich, Germany
[3] University of Copenhagen, DIKU
Njalsgade 128, Copenhagen S, Denmark
{fabian.hennecke,alexander.deluca,andreas.butz}@ifi.lmu.de,
nguyenn@cip.ifi.lmu.de, sebastian.boring@diku.dk

Abstract. Pointing tasks are a crucial part of today's graphical user interfaces. They are well understood for flat displays and most prominently are modeled through Fitts' Law. For novel displays (e.g., curved displays with multi-purpose areas), however, it remains unclear whether such models for predicting user performance still hold – in particular when pointing is performed across differently oriented areas. To answer this question, we conducted an experiment on an angularly coupled display – the *Curve* – with two input conditions: direct touch and indirect mouse pointer. Our findings show that the target position affects overall pointing speed and offset in both conditions. However, we also found that Fitts' Law can in fact still be used to predict performance as on flat displays. Our results help designers to optimize user interfaces on angularly coupled displays when pointing tasks are involved.

Keywords: Pointing, Fitts' law, display orientation, curved surface.

1 Introduction

Since the commercialization of the WIMP paradigm (Windows, Icons, Menus, Pointer), pointing has become the fundamental interaction technique for a variety of displays – either through pointing devices or more recently through direct touch. The abundance of different input technologies and display types turned pointing on a flat display into a widely researched field. In his original experiment, Fitts [1] studied direct pointing at physical objects. MacKenzie et al. [5] and others looked at indirect pointing and confirmed that Fitts' Law – while not intended for such scenarios – is still applicable to different input techniques. However, they could show that different input devices heavily affect a user's pointing performance.

As interactive surfaces with different sizes and orientations (e.g., tables, walls, etc.) have become more and more commonplace since the DigitalDesk [14], recent pointing experiments focused on such displays. Although these displays still allow indirect pointer input, they also provide the possibility of direct touch. For large horizontal

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 720–727, 2013.

surfaces (e.g., tabletops), several experiments revealed that a Fitts'-related formula still describes this type of interaction well [6,10]. Po et al. [9] demonstrated the predictability of input performance also on large vertical displays. While pointing tasks on horizontal or vertical screens are well understood individually, as of today it is unclear whether those results will hold for pointing across a combination of such displays as the display's orientation influences the precision of direct pointing [2]. Nacenta et al. [7] found that gaps between displays in multi-display environments influence indirect pointing performance. Recently developed displays like the Curve and BendDesk aim to seamlessly combine horizontal and vertical displays into a single, curved screen [13,15] and allow for mouse and touch input. Studies already revealed an influence on dragging and flicking across the display connection [3,12].

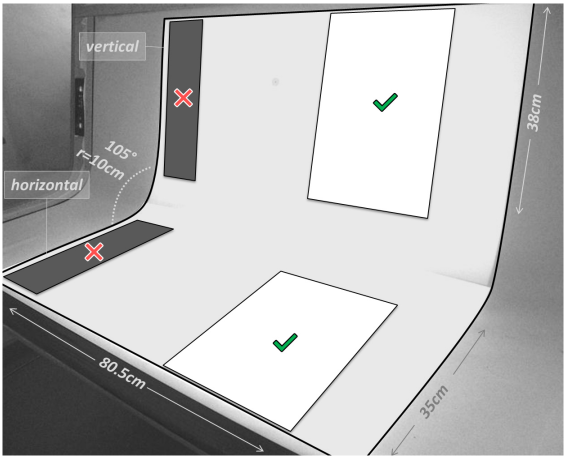


Fig. 1. Pointing performance: central target areas (white) performed best and outer areas (grey) performed worst in terms of task completion time for touch and pointer

Beside those already known effects, such novel displays introduce a series of challenges for pointing tasks: a change of pointer perception during display transition [8], oblique touch and viewing angles [2], and different finger and arm movements compared to a hypothetical planar, angular displays. They also introduce different pointing distances between targets on different display areas with regard to the input modality: while the cursor has to cross the display surface, the user's finger can use a midair shortcut (i.e., a more direct way). Since pointing is extremely important in current user interfaces, it is vital to understand how these effects influence pointing and whether or not there are different sweet spots for pointing depending on the input modality.

We conducted an experiment to identify the influence of angularly coupled display areas on generic pointing tasks. We used two input modalities – direct touch and indirect mouse pointer – in a Fitts'-like task design. In this paper, we present the study design and its results, which are a first step towards a deeper understanding of the placement of interactive elements with regards to the input modality. While the best position for a pointer-sensitive button is in a corner of the display (e.g., Start-Button in former Windows versions), it is different for touch-sensitive areas. While the handedness of a user already narrows down the choice of potentially good areas, it still remains unclear where these areas are exactly for a given input modality.

2 Evaluation

To better understand the main influences of angularly coupled display areas, we focused on three main research questions: (RQ1) does pointing performance vary for different display areas in terms of time and offset? (RQ2) Does the target's position affect the user's subjective perception of pointing performance? (RQ3) Can pointing time be predicted based on the target's position and its size?

2.1 Apparatus, Design and Participants

We conducted our experiment on the Curve display (see Figure 1), whose design is ergonomically optimized as shown by Wimmer et al. [15]. It contains two HD projectors for high-resolution output and four PointGrey FireflyMV cameras for touch input. With both the projections as well as the tracking cameras overlapping, the output resolution is 60 dpi (tracking resolution: 14 dpi). An 800 dpi optical laser mouse with standard Windows 7 cursor properties (e.g., acceleration) was used for pointer input.

MacKenzie et al. [4] described the problem of participants entering targets at an angle and thus increased the width of these two-dimensional targets. To overcome this, we used circular targets and varied their diameter. We only investigated tasks along the vertical axes to avoid effects of the crossing angle on user performance and perception [12]. In order to cover the height of both display areas we also varied the distance between the starting points and the targets.

We conducted two experiments with different input conditions: touch and pointer input. Within each experiment, we used a repeated measures design. We varied the horizontal position of the target area (six different axes, spaced 269 px (11.5 cm) apart from each other), the size of the target areas (diameter: 40 px (1.7 cm), 54 px (2.3 cm), 70 px (3.0 cm), and 91 px (3.9 cm)), the distance between start button and target area along the surface (402 px (17.2 cm), 810 px (34.6 cm), 1212 px (51.8 cm), and 1616 px (69.1 cm)), and the direction along the axis (upwards, downwards) as within-subject variables (see Figure 2). We decided to use the surface distance as this considers the midair shortcut in the touch condition as a special capability of the input technique. The order of the axes was counterbalanced using a Latin square and all other factors were randomized per participant. Each participant had to complete two blocks of 192 trials each ($6 \times 4 \times 4 \times 2$).

We recruited 30 participants per experiment, none of which participated in both experiments (touch: 22 male, 8 female; body-height: 159 cm – 194 cm; pointer: 22 male, 8 female; body-height: 155 cm – 194 cm), 27 being right-handed in the touch- and 24 in the pointer-input experiment.

2.2 Task and Procedure

Each participant was seated centrally in front of the display. Before the experiment began, participants familiarized themselves in a training phase with 20 random pointing tasks, which were similar to those used in the actual experiment. Participants were allowed to use the hand of their choice for input.

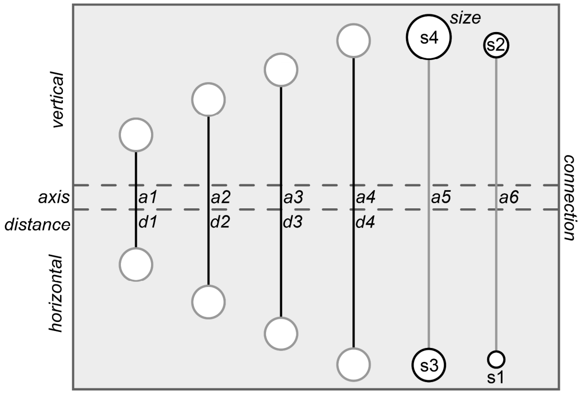


Fig. 2. Task Layout with the within-subject factors: 6 *axes*, 4 *distances* and 4 target *sizes* resulting in 192 trials as pointing was done in both *directions* (upwards and downwards)

The participants of the touch-input experiment were further equipped with a small marker (weight: 4 grams) on their input finger. A camera kept track of the marker movement to gather information on how they bridged the distance between both display areas. In both conditions, participants first had to press a start button on one end of an axis and then aim for the target area (see Figure 2) without feedback about pointing performance. Afterwards, all participants completed a questionnaire.

2.3 Measures

We measured: (1) Task completion time (TCT) as the time between the lift-off event within the start button (finger lift-off, button release) and the first recognition of an event within the target (touch recognition, mouse button down; and (2), the Pointing offset (PO) as the distance between the center of the target and the center of the participants' input (center of touch, pointer position) without correcting for touch perception or pointer movement in the display's connection.

To assess the participants' own perception of their performance and possible fatigue, we used a self-assessment questionnaire with 5-point Likert scales.

2.4 Statistical Tests and Analysis

We used the first set of each session (192 trials) for our analysis except for analyzing fatigue, as the repetition did not have any effect on the results. Additionally, we did not take any targeting errors (participant does not hit the target) into account, as they were too few to have an influence at all. If not stated otherwise, we used an Axes × Distance × Size × Direction (6 × 4 × 4 × 2) repeated measures analysis of variance (ANOVA).

3 Results

In the pointer input experiment, we had to remove two participants due to corrupted log files. In cases in which the assumption of sphericity was violated, we applied Greenhouse-Geisser corrections.

3.1 Task Completion Time

Touch. We found significant main effects for the target's Size ($F_{3,87} = 41.08$, $p < .001$), and the target's Distance ($F_{2.281,66.146} = 114.86$, $p < .001$). The factor Axes also showed a significant main effect ($F_{5,145} = 2.31$, $p < .05$) with axis 4 (M = 817ms) and axis 5 (M = 803ms) being the best, and axis 1 (M = 851ms) the worst. Looking at the participants' handedness revealed only a small influence on the TCT. Interestingly, the task's Direction also showed a significant main effect ($F_{1,29} = 8.484$, $p < .05$) with downward pointing being on average 4% (36ms) faster than pointing upwards We also found significant interaction effects for Distance × Size ($F_{6.244,181.08} = 2.18$, $p < .05$), Size × Direction ($F_{3,87} = 2.855$, $p < .05$), Axes × Distance × Size ($F_{12.926,374.851} = 1.935$, $p < .001$) and Distance × Direction ($F_{2.14,62.072} = 7.766$, $p < .001$).

Pointer. As for touch input we found a significant main effect of the factor Size ($F_{3,81} = 225.564$, $p < .001$), and Distance ($F_{3,81} = 328.514$, $p < .001$). The factor Axes also had a significant main effect ($F_{3.252,87.796} = 10.723$, $p < .05$). Similar to the results for touch tasks, axes 4 (M = 1211 ms) and 5 (M = 1215 ms) were completed fastest, and axis 1 (1337 ms) slowest. Post-Hoc tests showed significant differences between axes 1 and 4, 1 and 5, and 1 and 3 (M = 1220 ms). We neither found interactions, nor – unlike for touch input – did we find a significant effect for Direction.

3.2 Pointing Offset

Regarding the results for pointing offset during our study (RQ1), one has to keep in mind that participants were primarily asked to point as fast as possible. We acknowledge that our results can only provide an indication regarding the effects on pointing precision in terms of offset, and that this topic will require further studies. Nevertheless, we think it can help to optimize the size and position of interactive areas for common tasks like pressing a button. That said, we analyzed PO in two ways: across all target sizes (all) and only for the smallest targets (smt) to eliminate the obvious larger offset results for larger targets. We define PO as the Euclidean distance in pixels (px) of the participants' input from the target's center.

Touch. The Axes had a significant main effect (all: $F_{5,145} = 5.619$, $p < .001$; smt: $F_{5,145} = 3.741$, $p < .05$). Through a post hoc test, we found significant differences (all) between axes 3 (M = 15.105 px (0.64 cm)) and 5 (M = 13.531 px (0.57 cm); $p < .05$) and 6 (M = 13.957 px (0.59 cm), $p < .001$) and also between (smt) axes 5 (M = 11.046 px (0.47 cm)) and 1 (M = 12.492 px (0.53 cm)). We also found a significant main effect for Distance (smt: $F_{3,87} = 11.283$, $p < .001$), and a significant Direction × Distance interaction (all: $F_{3,87} = 35.992$, $p < .001$, smt: $F_{3,87} = 19.667$, $p < .001$) as well as an Axis × Distance (smt: $F_{15,435} = 2.597$, $p < .001$), and an Axis × Direction interaction (smt: $F_{5,145} = 2.49$, $p < .05$). Post-hoc tests revealed that participants' pointing offset in the lower part of the vertical display area is smaller than on the horizontal area near the display connection ($p < .05$).

Pointer. We found no significant main effects for Axes, Distances, or Directions on pointing offset for pointer input neither across all nor for smallest target sizes. Not surprisingly, Size had a significant effect on PO ($F_{3,81} = 499.663$, $p < .001$) ranging from M = 12.272 px (0.52 cm, smallest size) to M = 24.388 px (1.03 cm, largest size).

3.3 Subjective Ratings

We used 5-point Likert scales to assess our participants' subjective ratings regarding TCT and PO. They are combined into three categories for this report: 'I disagree' ('1', '2'), 'Neutral' ('3') and 'I agree' ('4', '5').

Touch. The subjective data regarding TCT is mainly in line with our objective measures. 93% of the participants stated they performed fastest on the axes in the display's center. Concerning PO 86% of the participants found that the offset near the connection on the horizontal area was small, while only 73% considered this on the vertical area near the connection. Interestingly, measured data revealed the exact opposite. Though participants reported shoulder (53%) and arm (83%) fatigue, we found no evidence that this influenced the pointing performance.

Pointer. Looking at the ratings for the targeting speed with respect to Axes, 96% stated that they could hit the target on the two most central axes fast while only 46% stated that for the four outer axes. This is in line with our measurements (RQ2). Despite the lack of objective differences, 76% of our participants stated that trials with an upward direction could be completed fast while only 66% said so for the downward trials. This indicates that it might be harder to keep track of the pointer moving it downwards onto the horizontal area than the other way around onto the vertical area.

3.4 Predictability

We calculated the general index of difficulty (ID) of our task setup and the throughput (TP) of both input styles to assess the applicability of Fitts' Law and to review an accuracy-speed trade-off. The ID in our study ranged from 2.4 to 5.4 bits, which is within the range proposed by Soukoreff et al. [11]. The TP for touch input is 5.62 bps and 3.57 bps for pointer input. Although this shows that touch input performed better than mouse input, we cannot determine whether the differently oriented display areas lead to a better performance compared to planar displays [11]. We combined the TCT of all participants for each ID and calculated regression lines resulting in these formulas for prediction of movement time:

$$MT_{Touch} = 192.96 + 129.46 * ID;\ with\ r^2 = 0.932$$
$$MT_{Pointer} = 199.08 + 230.04 * ID;\ with\ r^2 = 0.985$$

They show that Fitts' Law is able to accurately predict the pointing performance for both touch and pointer input across both display areas of our setup (RQ3).

4 Discussion and Future Work

Our results show that pointing performance with both touch and pointer input across differently-oriented display areas is influenced by both a target's position and the direction of a task. Besides this, both input styles are generally predictable using Fitts' Law for tasks including perpendicular crossings of the display connection. Despite the simple task design, this still allows to identify a first set of suitable interaction areas for touch and pointer input.

We found that pointing performance with touch input is best in the center of the screen with a tendency towards the right display area. As most of our participants were right-handed, this tendency indicates that important interface elements should be placed toward the dominant hand's side of the user. Likewise, we found that touch input close to the display connection was more accurate on the vertical display area than on the horizontal one, which should be considered by application designers. Though this happens at the cost of slightly worse interaction times in this additional area, it may be reasonable to mitigate accuracy problems based on an oblique viewing and touch angle [2].

Contrary to touch input, we found only little evidence for an influence of differently-oriented display areas on pointer input performance. While pointing offset was only influenced by the target size, we found significant differences between task completion times depending on the target's position. Although only axis 1 differed significantly from all others, both outer axes performed worse than the central axes even across both input modalities. We also noticed bulged movement trajectories on the outer axes of the horizontal display area with pointer input as described by Hennecke et al. [3]. Though the understanding of this observation definitely requires additional studies we think the different results for the display areas could be caused by perspective distortion.

Our results are only directly applicable to the display setup used in the study and tasks, which cross the display connection vertically. For this reason, we plan to investigate the influence of different angles of the vertical display area as well as the task axes. Though we did not find any statistical evidence for an influence of a user's handedness, we also see the need for an additional study investigating this matter. It will be very interesting to see which of these parameters can be taken into account leading to a general Fitts' Law formula for a curved display.

References

1. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology. General 121(3), 262–269 (1954)
2. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-touch vs. mouse input for tabletop displays. In: Proc. of CHI 2007, pp. 647–656 (2007)
3. Hennecke, F., Matzke, W., Butz, A.: How Screen Transitions Influence Touch and Pointer Interaction Across Angled Display Arrangements. In: Proc. of CHI 2012, pp. 209–212 (2012)
4. MacKenzie, I., Buxton, W.: Extending Fitts' law to two-dimensional tasks. In: Proc. of CHI 1992, pp. 219–226 (1992)
5. MacKenzie, I., Sellen, A., Buxton, W.: A Comparison of Input Devices in Element Pointing and Dragging Tasks. In: Proc. of CHI 1991, pp. 161–166 (1991)
6. Micire, M., Schedlbauer, M., Yanco, H.: Horizontal Selection: An Evaluation of a Digital Tabletop Input Device. In: Proc. of AMCIS 2007, p. 164 (2007)
7. Nacenta, M.A., Mandryk, R.L., Gutwin, C.: Targeting across displayless space. In: Proc of CHI 2008, pp. 777–786. ACM, New York (2008)
8. Nacenta, M.A., Sallam, S., Champoux, B., Subramanian, S., Gutwin, C.: Perspective cursor: perspective-based interaction for multi-display environments. In: Proc. of CHI 2006, pp. 289–298 (2006)
9. Po, B.A., Fisher, B.D., Booth, K.S.: Mouse and touchscreen selection in the upper and lower visual fields. In: Proc. of CHI 2004, pp. 359–366 (2004)
10. Sasangohar, F., MacKenzie, I.S.: Evaluation of mouse and touch input for a tabletop display using Fitts' reciprocal tapping task. In: Proc. of HFES 2009, pp. 839–843 (2009)
11. 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)
12. Voelker, S., Sutter, C., Wang, L., Borchers, J.: Understanding flicking on curved surfaces. In: Proc. of CHI 2012, pp. 189–198 (2012)
13. Weiss, M., Voelker, S., Sutter, C., Borchers, J.: BendDesk: Dragging Across the Curve. In: Proc. of ITS 2010, pp. 1–10 (2010)
14. Wellner, P.: The DigitalDesk calculator: tangible manipulation on a desk top display. In: Proc. of UIST 1991, pp. 27–33 (1991)
15. Wimmer, R., Hennecke, F., Schulz, F., Boring, S., Butz, A., Hußmann, H.: Curve: Revisiting the Digital Desk. In: Proc. of NordiCHI 2010, pp. 561–570 (2010)

Semi-supervised Learning Based Aesthetic Classifier for Short Animations Embedded in Web Pages

Dipak Bansal and Samit Bhattacharya

Dept. of Computer Science and Engineering, IIT Guwahati-781039, Assam, India
{d.bansal143,samit3k)@gmail.com

Abstract. We propose a semi-supervised learning based computational model for aesthetic classification of short animation videos, which are nowadays part of many web pages. The proposed model is expected to be useful in developing an overall aesthetic model of web pages, leading to better evaluation of web page usability. We identified two feature sets describing aesthetics of an animated video. Based on the feature sets, we developed a Naïve-Bayes classifier by applying Co-training, a semi-supervised machine learning technique. The model classifies the videos as *good*, *average* or *bad* in terms of their aesthetic quality. We designed 18 videos and got those rated by 17 participants for use as the initial training set. Another set of 24 videos were designed and labeled using Co-training. We conducted an empirical study with 16 videos and 23 participants to ascertain the efficacy of the proposed model. The study results show 75% model accuracy.

Keywords: Aesthetics, web page, short video, classification, semi-supervised learning, Co-training.

1 Introduction

Usability professionals over the years have been working extensively on developing methods and techniques to determine the usefulness of interactive systems. These activities are sought to be augmented in recent years with the studies on measuring perceived usability of the system, which relies heavily on aesthetics [7]. Postrel [9] contented that the 21st century is the “age of aesthetics”. The contention may well be true in the context of interactive systems, as the large number of recent works show [3,17,20,28-30].

Web pages are good examples to consider the importance of aesthetics in interactive system design. Most of the pages contain various types of information, put together using various design patterns. Consequently, the complexity of the interfaces in terms of information content and layout is usually high. Evidently, the aesthetics of the design determines to a great extent its acceptability (and therefore, usability) to the users [11, 12, 18, 19, 22, 23].

While the role of aesthetics on usability is clear, the problem lies in measuring it. Usability studies depend on quantifiable measurements. However, development of such measures for evaluating aesthetic quality of an interactive system is still in its

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 728–745, 2013.

infancy, primarily because of the perception that "aesthetics is subjective." While it is true to an extent, it is not impossible to develop quantitative measures of web page aesthetics, as reported in [16, 24]. Both these works actually demonstrate the possibility of computational modeling of aesthetics. The advantage of having a computational model is in the ability to evaluate aesthetic quality of an interface *automatically*, thereby making it possible to integrate the model as a tool in a web page design environment so that the designer can check their design quickly.

In this work, we propose a model to compute aesthetics of short animation videos, which are embedded in many of the web pages nowadays. This work is part of a larger goal of computational modeling of whole web page aesthetics. We base our work on the philosophy that modeling component aesthetics and then combining those models will lead to an overall web page aesthetics model.

We propose a semi-supervised learning model to compute aesthetics of short videos. The model is essentially a Naïve-Bayes classifier, which classifies a video into one of the three classes: *good*, *average* and *bad* with respect to its aesthetic quality. On the basis of a study of 18 web pages and prior work on this field, we identified two feature sets to capture short video characteristics with respect to its aesthetic quality. The feature sets are used in a co-training method to develop the classifier. We validated our model with an empirical study involving 16 videos and 23 participants. The feature sets, the co-training method and the validation study are described in this paper.

The paper is organized as follows. In Section 2, we discuss the related works reported in the literature along with their limitations that we address in this work. The development of the proposed model is detailed in Section 3. The empirical study conducted to evaluate the model performance is reported in Section 4. In Section 5, we discussed the strengths and limitations of the proposed model along with the scope for further works. The paper is concluded in Section 6.

2 Related Works

Considered as a branch of philosophy that deals with the nature of beauty, art, and taste[1], aesthetic design has been extensively studied in the field of fine and commercial arts [1, 7]. The importance of aesthetics in human affairs has been elaborated by Maquet [13]. In fact, as early as 1984, the role of aesthetics in determining usability of interactive systems was highlighted [9], where it was reported that a poorly designed computer screen can hinder communication. The positive effect of good graphic design and attractive displays on the transfer of information has been found by Aspillaga [2]. Elements of aesthetic considerations were present in other works as well [21,31-33].

Despite the presence of such early works, only the later part of the 1990s saw a spurt in activities in this area. These works included investigation of the role of aesthetics on interactive system design in general as well as on the effects of

---

[1] From Wikipedia. See `http://en.wikipedia.org/wiki/Aesthetics`

aesthetics in specific interaction domains. Researchers argued about the role of aesthetics in interactive system design [27]. Set of guidelines for screen design, keeping in mind the aesthetic aspect, were proposed [8]. In the context of e-learning, the effect of aesthetically pleasing layouts on the student's motivation to learn has been reported [26]. Szabo and Kanuka [25] found that subjects who used the lesson with *good design principles* completed the lesson in less time and had a higher completion rate than those who used the lesson with poor design principles.

A typical scenario where aesthetics play important role in the overall usability of the system is the design of web pages. Relationship between visual appeal and perceived usability of web pages was investigated in Lindgaard et al. [12]. Schmidt et al. [23] found correlation between usability and aesthetics in the context of subjective evaluation, depending on the user's background, goal, task, and application type. Several works concentrated on developing measures to assess aesthetic quality of web pages [11, 15].

Aesthetic evaluation of interfaces poses problem due to its subjective nature: an aesthetically pleasing interface may not look so to a different person. Computational aesthetic modeling attempts to overcome this problem by proposing objective measure of aesthetics [10, 16, 24].

One of the early works in this direction was by Ngo et al. [16]. In the approach, a numerical value is computed from the specification (in terms of elements, their positions, shapes and sizes) of an interface. The value signifies aesthetic of the layout. Aesthetics of two interfaces may be compared on the basis of the computed value. The model assumed a very simplified representation of the interface (i.e. each on-screen element is a rectangle). Aesthetic is determined by the geometric arrangement of the rectangles only. The content of the rectangles are not taken into consideration. Moreover, it considered only static images (i.e. the content does not change over time). Therefore, when we consider short videos embedded in a web page, it is not possible to apply the model, as we have to see "inside the box" (the content inside the rectangles) as well as consider the dynamic nature of the content.

In the context of short animation videos that are typically found embedded in web pages, some of these issues were addressed by Shyam and Bhattacharya [24]. In their work, a computational model was proposed to classify a short video into either of the classes *good, average* and *bad*, based on the aesthetic quality of the video. The model takes into consideration three factors that characterize a video, namely *symmetry*, *balance* and *color contrast*. We briefly discuss these factors in the following, as we have used them in our work.

The symmetry measure determines the extent to which the interface is symmetrical in vertical, horizontal and diagonal direction. In order to calculate symmetry (*Sym*) of an interface, Eq. 1 was proposed.

$$Sym = 1 - \frac{|S_h| + |S_v| + |S_r|}{3} \in [0,1] \tag{1}$$

In Eq. 1, S_h, S_v and S_r refers to the symmetry in horizontal, vertical and radial directions, respectively. Horizontal symmetry is calculated about a horizontal axis passing through the center of interface (Eq. 2a). Vertical symmetry is defined similarly, with respect to a vertical axis (Eq. 2b). Radial symmetry (Eq. 2c) refers to the symmetry about a diagonal passing through the center.

$$S_h = \frac{\begin{array}{l} |X'_{UL} - X'_{LL}| + |X'_{UR} - X'_{LR}| + \\ |Y'_{UL} - Y'_{LL}| + |Y'_{UR} - Y'_{LR}| + \\ |H'_{UL} - H'_{LL}| + |H'_{UR} - H'_{LR}| + \\ |B'_{UL} - B'_{LL}| + |B'_{UR} - B'_{LR}| + \\ |\theta'_{UL} - \theta'_{LL}| + |\theta'_{UR} - \theta'_{LR}| + \\ |R'_{UL} - R'_{LL}| + |R'_{UR} - R'_{LR}| \end{array}}{12} \tag{2a}$$

$$S_v = \frac{\begin{array}{l} |X'_{UL} - X'_{UR}| + |X'_{LL} - X'_{LR}| + \\ |Y'_{UL} - Y'_{UR}| + |Y'_{LL} - Y'_{LR}| + \\ |H'_{UL} - H'_{UR}| + |H'_{LL} - H'_{LR}| + \\ |B'_{UL} - B'_{UR}| + |B'_{LL} - B'_{LR}| + \\ |\theta'_{UL} - \theta'_{UR}| + |\theta'_{LL} - \theta'_{LR}| + \\ |R'_{UL} - R'_{UR}| + |R'_{LL} - R'_{LR}| \end{array}}{12} \tag{2b}$$

$$S_r = \frac{\begin{array}{l} |X'_{UL} - X'_{LR}| + |X'_{UR} - X'_{LL}| + \\ |Y'_{UL} - Y'_{LR}| + |Y'_{UR} - Y'_{LL}| + \\ |H'_{UL} - H'_{LR}| + |H'_{UR} - H'_{LL}| + \\ |B'_{UL} - B'_{LR}| + |B'_{UR} - B'_{LL}| + \\ |\theta'_{UL} - \theta'_{LR}| + |\theta'_{UR} - \theta'_{LL}| + \\ |R'_{UL} - R'_{LR}| + |R'_{UR} - R'_{LL}| \end{array}}{12} \tag{2c}$$

X'_j, Y'_j, H'_j, B'_j, θ'_j and R'_j (j = UR/UL/LR/LL) are the normalized values of the corresponding expressions shown in Eq. 3. UR, UL, LR and LL denote upper-right, upper-left, lower-right and lower-left, respectively. (x_{ij}, y_{ij}) and (x_c, y_c) in Eq. 3 refer to the center of each object *i* in quadrant *j* and the center of the interface. b_{ij} and h_{ij} are the width and height of the object. n_j is the total number of objects in the quadrant.

$$X_j = \sum_{i=1}^{n_j} | x_{ij} - x_c | \quad (3a)$$

$$Y_j = \sum_{i=1}^{n_j} | y_{ij} - y_c | \quad (3b)$$

$$H_j = \sum_{i=1}^{n_j} | h_{ij} | \quad (3c)$$

$$B_j = \sum_{i=1}^{n_j} | b_{ij} | \quad (3d)$$

$$\theta_j = \sum_{i=1}^{n_j} \frac{| y_{ij} - y_c |}{| x_{ij} - x_c |} \quad (3e)$$

$$R_j = \sum_{i=1}^{n_j} \sqrt{(x_{ij} - x_c)^2 + (y_{ij} - y_c)^2} \quad (3f)$$

The balance measure computes the difference between total optical weighting of components on each side of the horizontal and vertical axis. The optical weighting refers to the perception that some objects appear heavier than others. The expression for balance (*Bal*) is shown in Eq. 4, where B_h and B_v are the balance measured in the horizontal and vertical directions, respectively. Equation 5 shows the expressions to calculate the two components.

$$Bal = 1 - \frac{| B_h | + | B_v |}{2} \in [0,1] \quad (4)$$

$$B_h = \frac{w_T - w_B}{\max(| w_T |, | w_B |)} \quad (5a)$$

$$B_v = \frac{w_H - w_R}{\max(| w_H |, | w_R |)} \quad (5b)$$

In the above equations, $w_j = \sum_{i=1}^{n_j} a_{ij} d_{ij}$, j=L/R/T/B, L, R, T, B stand for left, right, top and bottom, respectively, a_{ij} is the area of object *i* on side *j*, d_{ij} is the distance between the central lines of the object and the interface and n_j is the total number of objects on the side.

The above formulations were for static images. The idea is extended in [24] for video that is a sequence of frames. The symmetry and balance for each frame are calculated separately and then weighted averages of these individual values are calculated to get the respective symmetry and balance for the whole video, as shown in Eq. 6, where sym_i and bal_i are the symmetry and balance values of the i^{th} frame respectively, f is the total number of frames, sd_{ij} is the symmetry difference between consecutive frames and bd_{ij} is the balance difference between consecutive frames.

$$Sym = \frac{sym_1 + \sum_{i=2}^{f} sym_i \times \frac{1}{|sd_{i,i-1}|}}{1 + \sum_{i=2}^{f} \frac{1}{|sd_{i,i-1}|}} \in [0,1]$$

$$Bal = \frac{bal_1 + \sum_{i=2}^{f} bal_i \times \frac{1}{|bd_{i,i-1}|}}{1 + \sum_{i=2}^{f} \frac{1}{|bd_{i,i-1}|}} \in [0,1] \tag{6}$$

Since the objects may change their position in a video, the above are calculated in terms of either *fixed objects* (i.e., those objects that don't change their position throughout the entire video) or the center of the frame (if there are no fixed objects).

The color contrast is the difference in visual properties that makes an object (or its representation in an image) distinguishable from other objects and the background. A three-stage approach was reported in [24] to calculate color contrast of a video. In the first stage, the video is divided into frames and then each frame is converted to gray image. Next, each gray image is converted to standard color enhanced image by histogram equalization. Finally, the original gray image is compared with the corresponding enhanced image in the third stage, to determine the color contrast of the video. Eq. 7 shows the computation of the color contrast (CC) value.

$$CC = \left| \frac{\sum_{i=1}^{p} std_i - org_i}{255 \times p} \right| \tag{7}$$

Based on these three factors, an expression shown in Eq. 8 was proposed in [24] to compute an aesthetic value (AS) for a video. On the basis of the computed value, videos are categorized as *good* (AS≥0.75), *average* (0.5≤AS<0.75) or *bad* (AS<0.5).

$$AS = \frac{Sym + Bal}{2} - CC \tag{8}$$

Although high classification accuracy (about 87%) was reported by Shyam and Bhattacharya [24], the model was developed based on several assumptions. These include, (a) the objects in the videos are of regular shapes, (b) they do not change size across frames and (c) the objects follow linear motion paths. In this work, we propose a machine-learning based approach to overcome these limitations.

3 Proposed Model

We propose a classifier that is *trained* with a set of training videos (i.e., short videos that are already classified as *good, average* or *bad*). As it was difficult to create a large training set, we used the co-training algorithm [6], which can work on small training set. Co-training is a semi-supervised learning technique that requires two views (represented by two feature sets) of the data. Ideally, the two views are conditionally independent (i.e., the two feature sets are conditionally independent given the class) and each view is sufficient (i.e., the class of an instance can be accurately predicted from each view alone).

Co-training first learns a separate classifier for each view using a small set of labeled (training) examples. The most confident prediction of each classifier for an unlabeled data is then used to iteratively construct additional labeled training data. We used the *Naive Bayes* classifier [14] to classify data in the co-training method.

3.1 Identification of Feature Sets

The first step was the development of feature sets. A feature set denotes a set of features that characterize a short video. For the proposed model, we identified two feature sets, denoted by FS_1 and FS_2.

The feature set FS_1 contains three features, namely *symmetry*, *balance* and *color contrast*. These are the factors described in [24] (discussed in the related works section), each of whose value lies within the range [0, 1].

The feature set FS_2 was determined from a survey of 18 web pages, sampled randomly from the Internet, containing short videos. In the survey, we looked for the shapes (regular/irregular) of the objects in the video, motion pattern (linear/non-linear) of the objects, presence of *fixed* objects and change in object size across frames of a video. The observations are summarized in Table 1. From the table, we can conclude that the characteristics *object shape*, *change in size*, *presence of fixed objects* and *motion path* may have an influence on the perceived beauty (aesthetics) of a video. Along with those, it is also important to take into account the *total number of objects* in a video, since too many or too few objects may not be pleasing to the eye.

On the basis of the analysis of the survey results, we propose five features that form FS_2:

1. Total number of objects (N).
2. Fixed objects measure (represented as n_f/N, where n_f is the number of fixed objects).
3. Measure of size change across frames (represented as n_s/N, where n_s is the number of objects changing size).
4. Measure of movement path (represented as n_l/N, where n_l is the number of objects with linear movement).
5. Object shape measure.

Table 1. Summary of the observations made with 18 web pages with embedded short videos sampled from the Internet

Characteristics	**Observation**
Object shape	Videos containing irregular shaped objects: 17 (94.4 %) Videos containing regular shaped objects: 1 (5.6%) Videos containing both regular and irregular shaped objects: 0 (0%)
Object size changes across frames	Videos where objects change their size across frame: 5 (27.8%) Videos where objects do not change their size: 13 (72.2%)
Fixed objects	Number of videos containing at least one fixed object: 7 (37.8%) Number of videos with no fixed objects: 11 (62.2%)
Motion paths	Videos containing objects with linear motion only: 6 (33.3%) Videos having objects with non-linear motion paths: 12 (66.7%)

In order to compute the last feature value (object shape), we used the formulation of Birkhoff [4], which works for object with polygonal shape[2]. According to the formulation, aesthetic quality (M) of any object can be computed in terms of *order* (O) and *complexity* (C) as in Eq. 9.

$$M=O/C \tag{9}$$

The Complexity C of an object is defined as the number of indefinitely extended straight lines which contain all the sides of the object (i.e., the number of distinct straight lines containing at least one side of the object). The Order O is a composition of five elements, as shown in Eq. 10.

$$O = V+E+R+HV-F \tag{10}$$

The individual terms on the right hand side of Eq. 10 are briefly described below (see [4] for more details).

- V stands for *vertical symmetry*. V=1 if the object posses symmetry about the vertical axis and V=0 otherwise.
- E stands for *equilibrium*. E=1 if V=1 or if the centre of the object is situated directly above a point P on a horizontal line segment AB supporting the object from below such that |AP| and |BP| > 1/6 of the total horizontal breadth of the object. If the center is above P but the above condition does not hold, E=0. For all other cases, E= -1.

---

[2] We can use this for any object shape in principle since any shape can be approximated with polygonal meshes. Thus, the solution is general, not specific to polygonal objects only.

- R stands for *rotational symmetry*. Let $360^{o}/q$ be the least degree of rotation which rotates the object into itself. Then, R=min{q/2,q/3} if V=1 for the object or its enclosing polygon, R=1 in any other case when q is even (i.e., in case of central symmetry) and R=0 otherwise.
- HV stands for relation of the object to a *horizontal-vertical network*. It can take the values of 0, 1 or 2 depending on the shape of the object.
- F stands for *unsatisfactory form*. F=0 if (a) the minimum distance from any vertex to any other vertex or side, or between parallel sides, is not less than $1/10^{th}$ the maximum distance between points of the polygonal object or (b) the angle between two non-parallel side is not less than 20^{o} or (c) there are at most two types of directions or (d) V and R are not both 0 or (e) there is at most one type of niche or (f) there is no unsupported re-entrant type. F=1 if the above conditions are fulfilled with only one exception. F=2 otherwise.

Let a video has *f* number of frames and M_{ij} is the Birkoff measure of the i^{th} object (i= 1,2...*N*) in the j^{th} frame (j= 1,2...f). Then, the object shape measure for the j^{th} frame (F_j) is computed as,

$$F_j = \frac{\sum_{i=1}^{N} M_{ij}}{N} \tag{11}$$

The above equation is for one frame. We calculate for each frame and take the average of all the frames. Hence, the object shape measure for the video is given as,

$$\text{Object shape measure} = \frac{\sum_{j=1}^{f} F_j}{f} \tag{12}$$

3.2 Creation of the Initial Training Set

The next step in the model development was the creation of a set of short videos that are already classified (i.e., labeled data). These labeled videos served as the initial training set. In order to create this training set, we conducted an empirical study in which we asked participants to rate a set of 18 artificially created short videos. From the participants' ratings, we labeled those 18 videos as *good*, *average* or *bad*. The details of the empirical study are discussed next.

Experimental Setup and Participants. We designed 18 videos using Adobe Flash Professional CS5™. The videos were divided into 3 sets of 6 videos each, containing regular shaped objects, irregular shaped objects and combination of both. We considered rectangular and circular shapes as regular. All other shapes were treated as irregular. Each video was displayed on a window of 320×233 resolution, had 40 frames with 2 sec duration (frame rate = 20) and were 2D, that is, the motion of all the objects were on a plane. The number of objects remained fixed in a video, that is, none of the objects were added or removed between the frames.

The total number of objects varied between 4 and 6 in each video. Two of the videos in each set contained fixed objects (1 and 3, respectively). One video in each set had 2 objects changing size across frames. The number of objects in linear motion varied between 0 and 4 in each set.

These 18 videos were shown to 17 participants on 17'' widescreen color displays attached to PCs having Intel® Core2™ Duo processor with 2.00 GHz speed, running Windows XP Professional with SP3. The participants included both male and female. All were either undergraduate or postgraduate students with average age of 21. All of them had normal or corrected-to-normal vision and were regular computer users. None were familiar with screen design concepts.

Table 2. Sequence of the videos in the playlist. Pi - playlist number.

P_1	1	2	18	3	17	4	16	5	15	6	14	7	13	8	12	9	11	10
P_2	2	3	1	4	18	5	17	6	16	7	15	8	14	9	13	10	12	11
P_3	3	4	2	5	1	6	18	7	17	8	16	9	15	10	14	11	13	12
P_4	4	5	3	6	2	7	1	8	18	9	17	10	16	11	15	12	14	13
P_5	5	6	4	7	3	8	2	9	1	10	18	11	17	12	16	13	15	14
P_6	6	7	5	8	4	9	3	10	2	11	1	12	18	13	17	14	16	15
P_7	7	8	6	9	5	10	4	11	3	12	2	13	1	14	18	15	17	16
P_8	8	9	7	10	6	11	5	12	4	13	3	14	2	15	1	16	18	17
P_9	9	10	8	11	7	12	6	13	5	14	4	15	3	16	2	17	1	18
P_{10}	10	11	9	12	8	13	7	14	6	15	5	16	4	17	3	18	2	1
P_{11}	11	12	10	13	9	14	8	15	7	16	6	17	5	18	4	1	3	2
P_{12}	12	13	11	14	10	15	9	16	8	17	7	18	6	1	5	2	4	3
P_{13}	13	14	12	15	11	16	10	17	9	18	8	1	7	2	6	3	5	4
P_{14}	14	15	13	16	12	17	11	18	10	1	9	2	8	3	7	4	6	5
P_{15}	15	16	14	17	13	18	12	1	11	2	10	3	9	4	8	5	7	6
P_{16}	16	17	15	18	14	1	13	2	12	3	11	4	10	5	9	6	8	7
P_{17}	17	18	16	1	15	2	14	3	13	4	12	5	11	6	10	7	9	8

Data Collection Procedure. We created 17 playlists for 17 participants using *balanced Latin squares* [5] (see Table 2). In Table 2, each row represents a playlist (P_i) shown to the i^{th} participant. The numbers in each cell represent one of the 18 videos. Video numbers 1-6 correspond to the set containing regular objects, 7-12 correspond to the set containing irregular objects and 13-18 correspond to the set containing both types of objects.

The videos were shown to the participants in the sequence shown in Table 2 and they were asked to rate the videos on a scale of 1 (least attractive) - 7 (most attractive) as per their perception of the attractiveness of the videos. Figure 1 shows the screenshot of the interface used by the participants to rate the videos. A play button allowed the participant to play the next video in the list, once s/he was finished with the current video. A replay button was also provided to enable the participant replay the current video. In the figure, it can be seen that the entire background screen was covered by the interface while the participant was rating the videos. This was done to avoid distraction.

Fig. 1. Screenshot of the rating interface

The ratings by the participants are shown in Table 3. We mapped the participants' rating to one of the three classes *good*, *average* and *bad*. We considered a rating of 1, 2 and 3 as *bad*, 4 and 5 as *average* and 6 and 7 as *good*. After the mapping, we took the statistical *mode* of the classes for each video, which was the final label (class) of the video. In case of a tie (i.e., more than one class occur in equal number), we take the average of the original ratings. The average value was used to assign class (between 1-3 as *bad*, 4-5 as *average* and 6-7 as *good*). The results are summarized in Table 4. From Table 4, it can be seen that three videos were labeled as *bad*, five were labeled as *good* and the remaining ten were labeled as *average*.

Table 3. Rankings of the videos by participants (in the scale of 1-7). The numbers in the top row denotes the videos and the leftmost column shows the participants.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
P_1	2	3	5	6	4	7	5	2	5	4	3	7	5	5	7	6	2	5
P_2	4	4	5	6	5	6	5	5	5	6	5	7	5	5	5	6	5	7
P_3	2	3	3	5	4	5	2	3	4	4	3	5	4	4	4	5	4	6
P_4	3	4	5	7	3	7	4	3	5	3	4	6	5	2	5	2	3	4
P_5	5	5	5	6	2	7	4	4	4	5	3	7	2	4	2	6	3	7
P_6	5	4	4	5	4	5	3	4	5	5	4	6	4	5	5	6	4	6
P_7	7	6	4	7	5	6	4	2	3	3	4	4	2	2	3	2	2	3
P_8	2	4	4	2	4	5	4	4	4	6	2	7	5	5	4	6	1	6
P_9	3	3	3	5	4	5	4	4	7	5	4	7	4	4	5	5	4	7
P_{10}	6	4	4	7	5	6	4	4	3	6	4	5	4	4	5	6	5	6
P_{11}	2	4	3	4	3	4	4	4	6	5	5	6	5	6	6	7	7	7
P_{12}	4	3	4	5	4	6	4	3	5	5	4	5	3	4	3	4	3	6
P_{13}	4	3	5	2	2	7	6	4	3	2	1	3	2	3	4	2	1	6
P_{14}	1	3	4	5	7	6	5	4	5	3	6	5	7	5	6	7	6	6
P_{15}	4	1	3	2	6	7	2	4	5	7	6	2	1	4	1	2	3	5
P_{16}	1	1	2	6	5	7	4	5	3	7	7	3	2	2	1	2	2	5
P_{17}	2	4	5	3	6	6	4	3	1	3	4	6	7	5	4	7	6	5

3.3 Unlabeled Dataset Creation

We designed another set of 24 short videos, which served as the unlabeled dataset (i.e., these were not classified from empirical data), using the same development platform as that of the labeled videos. The resolution, frame rate and duration of the videos were also the same along with the nature of the motion paths of the objects (2D).

The purpose of these unlabeled videos was to increase the training set size so as to cover a wide range of values for all the features. The videos were divided into 3 sets of 8 videos each. One set contained videos with regular objects only, one set was having videos with only irregular objects and the third set was having videos containing both regular and irregular objects.

The total number of objects in the videos varied between 2 to 7. The videos contained between 0 (5 videos) and 3 fixed objects. About 50% of the videos (13) contained objects (between 1 and 3) that changed size across frames. Two of the videos did not have any objects following linear motion path. In the remaining 22 videos, objects with linear motion path varied between 1 and 5.

Table 4. Labeling of videos from participants' rating (Table 3). The numbers inside parenthesis in the middle column show the number of participants who rated the video to belong to the corresponding class. A rating of 1, 2 or 3 was mapped to *bad*, 4 or 5 was mapped to *average* and 6 or 7 was mapped to *good* class. The final label is obtained as the statistical mode of the labels given by the participants.

Video	Participant rating	Final Label
1	Bad (9 participant), Average (6 participant), Good (2 participant)	Bad
2	Bad (8 participant), Average (8 participant), Good (1 participant)	Bad
3	Bad (5 participant), Average (13 participant), Good (0 participant)	Average
4	Bad (4 participant), Average (6 participant), Good (7 participant)	Good
5	Bad (4 participant), Average (10 participant), Good (3 participant)	Average
6	Bad (0 participant), Average (5 participant), Good (12 participant)	Good
7	Bad (3 participant), Average (13 participant), Good (1 participant)	Average
8	Bad (6 participant), Average (11 participant), Good (0 participant)	Average
9	Bad (5 participant), Average (10 participant), Good (2 participant)	Average
10	Bad (5 participant), Average (7 participant), Good (5 participant)	Average
11	Bad (5 participant), Average (9 participant), Good (3 participant)	Average
12	Bad (3 participant), Average (5 participant), Good (9 participant)	Good
13	Bad (6 participant), Average (9 participant), Good (2 participant)	Average
14	Bad (4 participant), Average (12 participant), Good (1 participant)	Average
15	Bad (5 participant), Average (9 participant), Good (3 participant)	Average
16	Bad (5 participant), Average (3 participant), Good (9 participant)	Good
17	Bad (9 participant), Average (5 participant), Good (3 participant)	Bad
18	Bad (1 participant), Average (5 participant), Good (11 participant)	Good

3.4 Implementation of the Training Method

The implementation of the Co-training method was done in MATLAB™. In order to calculate the feature values in the feature sets FS_1 and FS_2, we first divided a video into frames or sequence of images. Then, we tracked objects in each frame and found out the coordinates of the center of every tracked object. For tracking the objects in each frame, we first converted the frame to a binary image. Then, we applied the *bwmorph* function, which shrinks the objects to points. The final frame contains only points representing the number of objects in the frame. The steps are illustrated in Fig. 2(a)-(c).

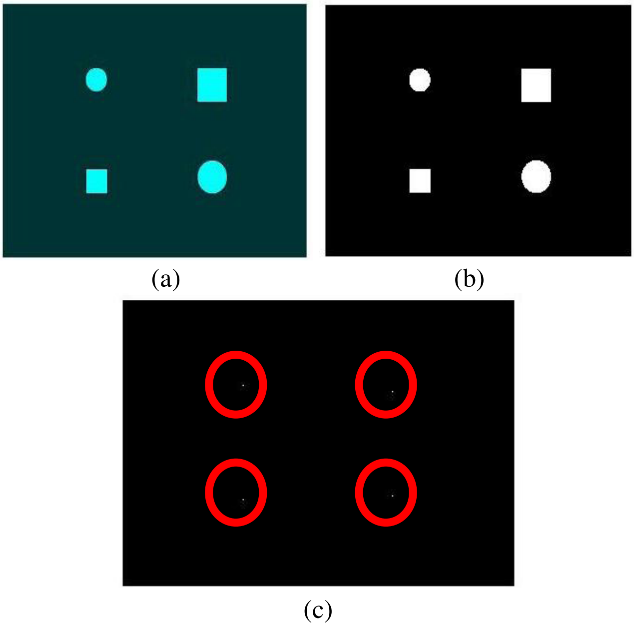


Fig. 2. Illustration of the feature value computation steps. From the frame in (a), a binary image is created (b). It is then converted to points representing objects (c) (shown inside circles).

Using the co-training algorithm, we classified the unlabeled videos. Among those videos, 7 were classified as *bad*, 11 as *average* and the remaining 6 as *good*.

4 Model Validation

We conducted an empirical study to check the accuracy of the model (classifier). In the study, we used the model to classify 16 short videos. The videos were then rated by 23 participants. From the rating, we determined the classes of the videos. The model classifications were then matched with the empirical classification to determine model accuracy. The details of the validation study are described next.

4.1 Experimental Setup

All the 16 videos were 2D (i.e., objects moved in 2D), designed using Adobe Flash Professional CS5^{TM} as before. Other characteristics, namely the frame rate, display resolution, total number of frames and duration were also the same as to that of the videos designed for training the model. None of the objects in a video was added or removed during the running of the video. The feature values were varied at random in the videos.

4.2 Participants and Procedure

Among the 23 participants, 10 took part in the previous study and 13 were new. All the participants were undergraduate or postgraduate students with regular computer exposure. Average age of the participants was 21.23 yrs. Among them, 15 were male and the rest were females. All of them had normal or corrected to normal vision. None of them had any experience with screen-design concepts before.

In order to collect data, the procedure we followed was similar to the one we used for labeling of the training videos. We created playlists of the videos for each participant following the Latin square method. The participants were asked to rate the videos using the same interface and rating scale. The ratings were then mapped to one of the classes, leading to the statistical mode based final classification of the videos.

4.3 Results

According to the participants' ratings, 4 videos were classified as *bad*, 4 as *good* and the remaining 8 as *average*. The classification we obtained using the model matched 12 of these empirical classes, resulting in 75% accuracy. The results of the study are summarized in Table 5.

Table 5. The comparison of the model prediction to that of the classification from empirical data

Video Number	Empirical Classification	Model Prediction
1	Bad	Bad
2	Bad	Bad
3	Good	Good
4	Average	Average
5	Good	Average
6	Bad	Average
7	Average	Average
8	Good	Good
9	Bad	Bad
10	Average	Average
11	Average	Good
12	Good	Good
13	Average	Average
14	Average	Bad
15	Average	Average
16	Average	Average

5 Discussion

In this work, we tried to address the limitations in the work reported by Shyam and Bhattacharya [24], by proposing a more generalizable classifier, which is trained using the co-training method. The results of the validation study show that the proposed classifier is able to classify short videos according to their aesthetic appeal, with a reasonably high accuracy rate of 75%.

The classification helps a designer decide if a video needs to be improved to increase its aesthetic appeal. For videos belonging to the *good* category, improvements may not be necessary. For *average* category videos, improvements may help while for videos classified as *bad*, it is definitely required. As is obvious, this has significant implication from the point of view of usability of web pages, when we consider web pages with embedded videos. We believe the work can be extended for the development of a more generalized aesthetic model for web pages.

An important characteristic of the videos used in the study was that the number of objects remained fixed (i.e., no addition/deletion of objects was considered). Admittedly, the constraint may not characterize some real-world embedded videos. Therefore, it may be necessary to carry out further work to determine the validity of the proposed model for videos that do not have fixed number of objects.

Although the model accuracy was reasonably satisfactory, we feel that further improvements are possible. The feature sets were developed on the basis of a survey of 18 videos sampled from the Internet. A larger sample size may reveal other characteristics, thereby enriching the feature set. Moreover, the initial training set was created with data of 17 participants for 18 videos. There are scopes to improve the initial training set by increasing the number of videos and participants and also by introducing more variations, in terms of age, gender, educational background and so on, to the participants' profile. Finally, the accuracy figure also needs to be corroborated further by considering more videos and larger number of participants with more variations in their profile.

6 Conclusions

In this paper, we reported a computational model to classify short videos based on their aesthetic quality. The model is a Naïve Bayes classifier, developed using the co-training method. The model was developed and validated using empirical data. Experiments show that the model can classify videos with 75% accuracy.

In future, we plan to work on two directions: refinement of the model and using the model to propose an overall computational model for aesthetic evaluation of a web page. We plan to refine the model by carrying out the following tasks.

- Refinement of the feature set by surveying larger number of real-world embedded videos.
- Increase the initial training set by increasing the number of videos and a larger set of participants with more varied profile to label those videos.
- Perform more extensive validation experiments with more videos and larger number of participants with more variations in their profile.

Acknowledgements. We are thankful to all the participants who volunteered for the empirical studies.

References

1. Arnheim, R.: Art and Visual Perception. University of California Press, Berkeley (1954)
2. Aspillaga, M.: Screen design: Location of information and its effects on learning. Journal of Computer-Based Instruction, 89–92 (1991)
3. Bartelsen, O.W., Petersen, M.G., Pold, S. (eds.): Aesthetic Approaches to Human-Computer Interaction, NordiCHI 2004 Workshop, Tampere, Finland (2004)
4. Birkhoff, G.D.: Aesthetic Measure, ch. 1. Harvard University Press, Cambridge (1933)
5. Blandford, A., Cox, A.L., Cairns, P.: Controlled experiments. In: Cairns, P., Cox, A.L. (eds.) Research Methods for Human-Computer Interaction, pp. 1–16. Cambridge University Press (2011)
6. Blum, A., Mitchell, T.: Combining Labeled and Unlabeled Data with Co-training. In: Proc. 11th Annual Conference on Computational Learning Theory, pp. 92–100 (1998)
7. Dondis, D.A.: A Primer of Visual Literacy. The MIT Press, Cambridge (1973)
8. Galitz, W.O.: The essential guide to user interface design: an introduction to GUI design principles and techniques. John Wiley Sons Inc., New York (1997)
9. Heines, J.: Screen Design Strategies for Computer-assisted Instruction. Digital Press, Bedford (1984)
10. Lai, C.Y., Chen, P.H., Shih, S.W., Liu, Y., Hong, J.S.: Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images. International Journal of Human-Computer Studies 68, 41–56 (2010)
11. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer Studies 60(3), 269–298 (2004)
12. Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., Noonan, L.: An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Transactions on Computer-Human Interaction 18(1) (2011)
13. Maquet, J.: The Aesthetic Experience. Yale University Press, New Haven (1986)
14. Mitchell, T.: Machine Learning, ch. 6. McGraw-Hill (1997)
15. Moshagen, M., Thielsch, M.T.: Facets of visual aesthetics. International Journal of Human-Computer Studies 68, 689–709 (2010)
16. Ngo, D.C.L., Teo, L.S., Byrne, J.G.: Modeling interface aesthetics. Information Sciences 152, 25–46 (2003)
17. Norman, D.A.: Introduction to this special section on beauty, goodness, and usability. Human-Computer Interaction 19, 311–318 (2004)
18. Pandir, M., Knight, J.: Homepage aesthetics: The search for preference factors and the challenges of subjectivity. Interacting with Computers 18(6), 1351–1370 (2006)
19. Park, S., Choi, D., Kim, J.: Critical factors for the aesthetic fidelity of web pages: empirical studies with professional web designers and users. Interacting with Computers 16(2), 127–145 (2004)
20. Petersen, M.G., Hallinas, L., Jacob, R.J.K.: Introduction to special issue on the aesthetics of interaction. ACM Transactions on Human-Computer Interaction 15(4) (2008)
21. Reilly, S., Roach, J.: Improved visual design for graphics display. IEEE Computer Graphics and Applications 4(2), 42–51 (1984)

22. Schaik, P.V., Ling, J.: Five psychometric scales for online measurement of the quality of human-computer interaction in web sites. International Journal of Human-Computer Interaction 18(3), 309–322 (2005)
23. Schmidt, K., Liu, Y., Sridvasan, S.: Webpage aesthetics, performance and usability: Design variables and their effects. Ergonomics 52(6), 631–643 (2009)
24. Shyam, D., Bhattacharya, S.: A Model to Evaluate Aesthetics of Short Videos. In: Proc. 10th Asia Pacific Conference on Computer Human Interaction (APCHI 2012), Matsue, Japan, pp. 315–324 (2012)
25. Szabo, M., Kanuka, H.: Effects of violating screen design principles of balance, unity and focus on recall learning, study time, and completion rates. In: Proc. ED-Media/ED-Telecom 1998 Conference, Charlottesville, VA, Association for the Advancement of Computing in Education (1998)
26. Toh, S.C.: Cognitive and Motivational Effects of Two Multimedia Simulation Presentation Modes on Science Learning. PhD thesis, University of Science Malaysia, Malaysia (1998)
27. Tractinsky, N.: Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In: CHI 1997 Conference Proceedings. ACM, New York (1997)
28. Tractinsky, N.: Does aesthetics matter in human computer interaction? Mensch and Computer, 29–42 (2005)
29. Tractinsky, N., Hassenzhal, M.: Arguing for aesthetics in human-computer interaction. i-com Z. Interakt. Koop. Medien, 66–68 (2005)
30. Tractinsky, N., Shoval-Katz, A., Ikar, D.: What is beautiful is usable. Interacting with Computers 13(2), 127–145 (2000)
31. Tullis, T.S.: An evaluation of alphanumeric, graphic, and colour information displays. Human Factors 23, 541–550 (1981)
32. Tullis, T.S.: Predicting the Usability of Alphanumeric Displays. PhD thesis, Rice University, Kansas (1984)
33. Tullis, T.S.: Screen design. In: Helander, M. (ed.) Handbook of Human-Computer Interaction, pp. 377–411. Elsevier Science Publishers, Amsterdam (1988)

Switchback Cursor: Mouse Cursor Operation for Overlapped Windowing

Shota Yamanaka[1] and Homei Miyashita[1,2]

[1] Meiji University
[2] JST CREST
stymnk@meiji.ac.jp, homei@homei.com

Abstract. When we perform a task that involves opening a number of windows, we cannot access the objects behind them. Thus, we are forced to switch the foreground window frequently or to move it temporarily. In this paper, we propose a Switchback Cursor technique where the cursor can move underneath windows when the user presses both the left and right mouse buttons. We also discuss some of the advantages of our method and effective situations that may be suited to the Switchback Cursor.

Keywords: Cursor, Graphical user interfaces (GUIs), Mouse, Pointer, WIMP.

1 Introduction

Numerous window manipulations are performed when we work on a PC running multiple windows, such as moving and resizing. Besides, we need to click background windows frequently to switch to foreground windows. These actions are performed simultaneously in our main task, such as application operation or file organizing, and we are sometimes required to perform many complicated window manipulations.

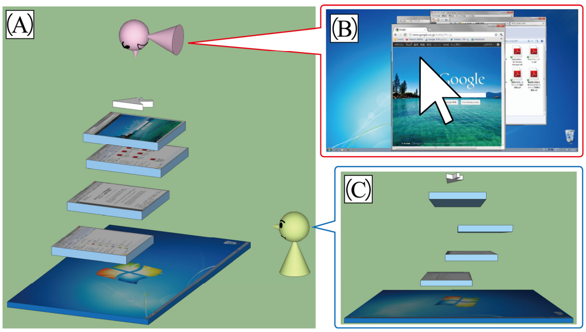


Fig. 1. (A): Observing overlapping windows from different perspectives. (B): The normal view of windows and the cursor on a PC. (C): The windows are piled up on the desktop and the cursor is always on the top.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 746–753, 2013.

These operations are required because there is a difference between the dimensions of the windows and the movement of the mouse cursor. In Fig. 1, (A) shows a 3D image of some overlapping windows, where (B) represents the picture that we normally view on the PC. The windows are distributed in 3D but the cursor can only move in 2D. Thus, the background windows can only come to the foreground if we click on them or use a keyboard command to control them.

In our novel approach, called Switchback Cursor, the mouse cursor can move underneath windows by hitting them (it does not move freely in 3D). When a user holds both the left and the right mouse buttons, the cursor moves from the edge of a window to the background, so that it can control the objects there. Fig. 2 shows how the cursor moves underneath windows using our proposed technique. A movie showing the behavior of the Switchback Cursor can be viewed at [1].

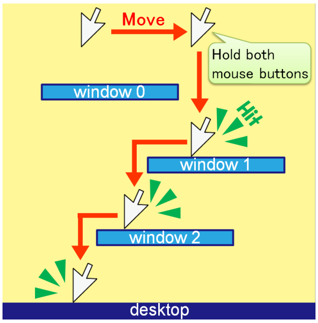


Fig. 2. The cursor moves progressively deeper after hitting successive windows

2 Related Work

Many techniques have been developed to address the problems of overlapping windows. Free-space Transparency (FST) [2] allows the free space in a window to become transparent so users can see through the objects behind the foreground window. This also allows us to perform basic interactions such as drag-and-drop and clicking icons using the white region in a window, which is similar to our proposed system. In the stack leafing technique [3], a mouse button is pressed and the foreground windows on the same layer switch to another layer by dragging. In QuickSpace [4], a user moves a window and it pushes *friend-registered* windows so they are not overlapped. Beaudouin-Lafon proposed a technique [5] that allows some window manipulations such as tabbed windows and peeled-back windows. The tabbed windows technique has now been implemented in Google Chrome and this method is used widely to solve the overlapped windows problem. Metisse [6] is a windows management system that allows users to rotate windows in 2D and 3D, so the windows can be allocated without overlapping to use the desktop space efficiently. WindowScape [7] addresses rearranging windows problem by photograph metaphor. Screenshots are stocked each time windows are miniaturized, and when the user selects one of the shots, WindowScape automatically allocates the windows state.

Some systems have been developed that allow the cursor to adopt irregular behaviors and their goal is mainly to make the mouse operation more efficient. Bubble Cursor [8] is a round-shaped cursor that extends the pointing range from one dot to a large circle, which reduces the movement distance. Dynaspot [9] enhances the size of the cursor area, depending on the cursor movement speed. When the cursor moves rapidly, the selection area becomes larger to increase the accuracy. However, *area cursor* methods such as Bubble Cursor and Dynaspot make it difficult to point when the targets lay side-by-side, but like our Switchback Cursor, these two techniques also allow movement into inaccessible areas using traditional techniques and clip manipulation. Delphian Desktop [10] moves the cursor to the target object immediately, while Drag-and-Pop [11] is a technique that allows the target object to travel to the cursor. These methods reduce the mouse manipulation time in a direct manner by allowing the targets or the cursor to warp. Ninja Cursors [12] uses multiple pointers and the user operates them all, i.e., the user only moves the cursor nearest the target object. The double mouse system [13] operates two cursors using two mice with one in each hand, which has the same effect as Ninja Cursors in reducing the manipulation time by operating a convenient cursor. Semantic Pointing [14] allows the cursor to speed up when it is further from a target by estimating the object that the user wants to select. MAGIC [15] makes the cursor jump to near the gaze-point using an eye tracker. This technique exploits the tendency to look at a target first before moving the cursor there. Fold-and-drop [16] allows items to be drag-and-dropped into the back window by turning over the windows like papers. This limits the manipulations using drag-and-drop but the user can search folders in the windows and run an icon on the desktop with the Switchback Cursor.

3 The Problem

When we perform our main task on a PC, it is often necessary to operate a number of windows, which are associated with the main task. However, a problem arises from the difference between the dimensionality of the windows and that of the cursor. Windows are set in depth layers from the front to back where the layers are structured in 3D. However, the cursor can only move in 2D. Thus, we cannot control objects hidden by windows and we have to click on a background window to perform our intended manipulation. This is a typical problem of the overlapping window system. A tiling window system does not have this problem, but it has bad visibility due to the small size of the windows so it is used less widely than the overlapping system.

There is also the problem with window focus, which refers to a window that receives inputs from the keyboard and mouse. A focused state is referred to as “the active window.” In Windows7, a clicked window is focused (focus follows click; FFC). However, the active window comes forward and the foreground window becomes hidden. This does not cause a problem when we begin another task but if we only want to access a background object briefly it is necessary to switch the foreground window a number of times. Another approach to window focusing depends on the position of the mouse cursor (focus follows mouse; FFM). However, the active window can be changed accidentally if the mouse is moved carelessly, so this approach has not been adopted by the Windows OS or the Mac OS. Thus, the general window viewing method and mouse cursor manipulation cause problems. This is a major problem because unnecessary window operations are required when controlling background objects using the overlapping window approach. Therefore, we can make mouse operation more comfortable by solving these two problems, i.e., the cursor cannot reach objects behind windows so we are forced to switch the foreground window many times in an FFC environment.

4 Switchback Cursor

We developed our system for Windows7. The system obtains windows information such as the handler, position, size, window style, and z-order and monitors these parameters. When a user presses both mouse buttons, the cursor moves to the same layer as the window beneath it. For example, Fig. 2 shows that a user moves the cursor from its initial position to a position above *window 1* and presses both buttons, so the cursor moves to the layer of *window 1*. If the user keeps pressing both buttons and moving the cursor to the left, it moves underneath *window 0*. In the same way, the cursor can move to progressively deeper layers by hitting successive windows.

When a user presses both buttons, the windows in front of the cursor are set as the *topmost window*, which become semi-transparent so the objects can be seen underneath them, and these windows are set *through mouse actions*. Subsequently, mouse actions such as click and drag are not received by the windows in front of the cursor and the actions are received by the window beneath the cursor. If the window receives a mouse signal, the order of the windows in front of the cursor is not changed because the front windows are set as the *topmost windows*. A transparency setting of 0% (where the windows are completely transparent) is discouraged because the user cannot perceive the position and shape of the windows, but this parameter can be changed by the user.

When both buttons are not pressed, the cursor aims to move to the front side but it hits the window above itself and remains in the layer behind the window. For example, in Fig. 2, a user releases the mouse button(s) when the cursor is underneath *window 0*, it remains behind *window 0*. If the cursor is returned from *window 0* without pressing at least one button, it moves to the foreground and dissolves the topmost setting of *window 0*.

When a user presses both buttons, the cursor size becomes 15% smaller as it moves to a window. When the cursor is around a window while pressing both buttons, the angle of the cursor changes so the user can see the direction where it is going as it moves beneath a window. When the cursor moves to a rear layer, a metallic sound rings to indicate that it has hit a window.

When the maximum-sized windows are open, they become slightly smaller (40 pixels less from each edge) while pressing both buttons so the cursor can move underneath them, as shown in Fig. 3. Further, when a window hides one or more windows, the background windows slide slightly to produce a small gap (30 pixels), as shown in Fig. 4.

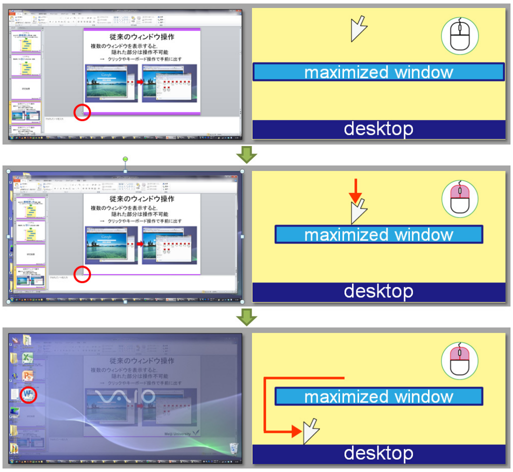


Fig. 3. A maximized window(s) becomes slightly smaller when both mouse buttons are pressed. The red circles indicate cursor positions. The left screenshot shows a maximized window while the images on the right show the left images from a location rotated to 90°, as shown in Fig. 1 (C). The cursor can reach the desktop icon while both buttons are pressed.

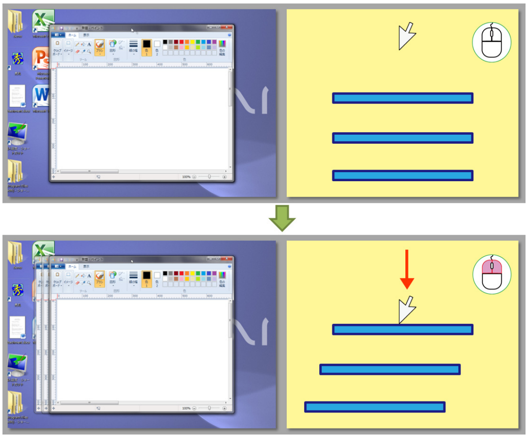


Fig. 4. The windows hidden by a front window (same size and same position, or completely covered) move to produce a gap when both buttons are pressed so that the cursor moves to their layers. The cursor cannot move under two window layers if not pressing both buttons. When pressing both buttons, the covered windows move so that the cursor can reach their layers.

5 Advantages and Effective Situations

In our proposed approach, the cursor moves from the foreground to the target window so the cursor moves in three dimensions within the display. A number of 3D mouse or 3D desktop systems have been developed in the past but in most cases the cursor moves in the Z direction freely and a special device or GUI is required for 3D input.

Unlike these systems, the cursor moves backward in our approach but it hits a window and stays there. No special devices are needed and only a normal mouse with left and right buttons is used.

We do not suggest that users employ our technique alone instead of traditional window switching. If he/she uses another application for a while, it is better to switch the foreground window; if a keyboard shortcut is suitable, it is preferable that the user performs keyboard operations. However, in situations where these manipulations become cumbersome and our technique provides a better alternative, users might prefer to use our technique. Therefore, we developed this system for use in a traditional GUI environment so the user can invoke this technique anytime by pressing both buttons, depending on the situation. Our system has high compatibility with modern overlapping windows settings and it does not block existing manipulations. In the subsections below, we describe some scenarios that may be suited to our technique.

5.1 A Window Hides other Windows When Switching the Foreground Window

Fig. 5 shows a situation where a person is browsing the web while listening to music. If we want to listen to a different piece of music, as shown in Fig. 5, we normally click on the background folder to switch to the foreground and drag-and-drop a file onto the music player (or double-click it). The browser then becomes hidden partly by the folder so we have to click the browser again to switch to the foreground to return to our browsing task. An alternative is to arrange the windows so they do not overlap one another, but the process is complicated in both cases. With the Switchback Cursor, we simply allow the cursor to reach the background of the browser and drag-and-drop the file so we do not need to click on the folder or the browser multiple times to switch to the foreground window.

Fig. 5. Web browsing while listening to music

5.2 Double-Clicking an Icon on the Desktop

In Fig. 6, the top screenshot and the illustration show a situation where several windows are open. If we want to double-click on an icon on the desktop to run an application or open a file, we would typically minimize all of the windows using the command Show Desktop shortcut, or move them out of the way and then double-click on the target icon. We would also have to reconstruct the original window layout to return to our former task, which increases the number of operation steps. By contrast, our technique requires no window manipulations because the cursor moves directly to the desktop and double-clicks on the icon, as shown at the bottom of the screenshot and in the illustration in Fig. 6.

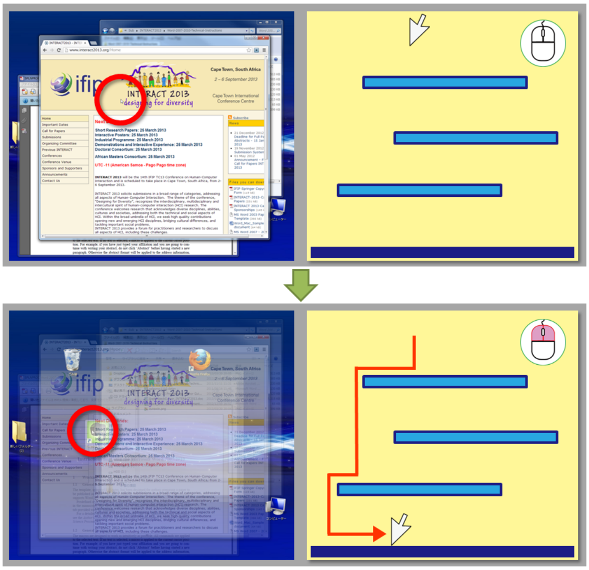


Fig. 6. The cursor moves underneath the windows and reaches the target icon on the desktop. The red circles indicate cursor positions. Some windows are pile up in the left screenshots while the images on the right show the left pictures from a location rotated by 90°. The user simply needs to press both buttons and move the cursor to the desktop before double-clicking the icon, without any window manipulations.

6 Conclusion

In this paper, we described the two problems of overlapping window systems: the difference in dimensions between the cursor and the windows, and the high number of manipulations required to switch foreground windows. To address these problems, we proposed the Switchback Cursor technique, which allows the cursor to move underneath the windows. We discussed the advantages of this methods and situations where using our method might prove effective. In our future work, we would like to test and verify the effectiveness of our technique by some tasks, such as moving icons within folders that exist in various layers. And we also plan to evaluate the limitation of Switchback Cursor; not only performance time and error rate but also the visibility of layers beneath the overlapped windows, and the window layout that makes the cursor hard to go underneath.

References

1. Switchback Cursor movie, http://www.youtube.com/watch?v=_nLt5PurgPY
2. Ishak, E.W., Feiner, S.K.: Interacting with Hidden Content Using Content-Aware Free-Space Transparency. In: Proc. of UIST 2004, pp. 189–192 (2004)
3. Faure, G., Chapuis, O., Roussel, N.: Power tools for copying and moving: useful stuff for your desktop. In: Proc. of CHI 2009, pp. 1675–1678 (2009)
4. Hutchings, D.R., Stasko, J.: QuickSpace: New operations for the desktop metaphor. In: Proc. of CHI Extended Abstracts 2002, pp. 802–803 (2002)
5. Beaudouin-Lafon, M.: Novel interaction techniques for overlapping windows. In: Proc. of UIST 2001, pp. 153–154 (2001)
6. Chapuis, O., Roussel, N.: Metisse is not a 3D desktop! In: Proc. of UIST 2005, pp. 13–22 (2005)
7. Tashman, C., Edwards, W.K.: WindowScape: Lessons learned from a task-centric window manager. ACM Transactions on Computer-Human Interaction 19(1), Article 8 (2012)
8. Grossman, T., Balakrishnan, R.: The Bubble Cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In: Proc. of CHI 2005, pp. 281–290 (2005)
9. Chapuis, O., Labrune, J.-B., Pietriga, E.: DynaSpot: Speed-Dependent Area Cursor. In: Proc. of CHI 2009, pp. 1391–1400 (2009)
10. Asano, T., Sharlin, E., Kitamura, Y., Takashima, K., Kishino, F.: Predictive interaction using the Delphian Desktop. In: Proc. of UIST 2005, pp. 133–141 (2005)
11. Baudisch, P., Cutrell, E., Robbins, D., Czerwinski, M., Tandler, P., Bederson, B., Zierlinger, A.: Drag-and-Pop and Drag-and-Pick: techniques for accessing remote screen content on touch- and pen-operated systems. In: Proc. of Interact 2003, pp. 57–64 (2003)
12. Kobayashi, M., Igarashi, T.: Ninja Cursors: Using Multiple Cursors to Assist Target Acquisition on Large Screens. In: Proc. of CHI 2008, pp. 949–958 (2008)
13. Nakamura, S., Tsukamoto, M.: Design and Implementation of the Double Mouse System for a Window Environment. In: Proc. of PACRIM 2001, pp. 204–207 (2001)
14. Blanch, R., Guiard, Y., Beaudouin-Lafon, M.: Semantic pointing: improving target acquisition with control-display ratio adaptation. In: Proc. of CHI 2004, pp. 519–526 (2004)
15. Zhai, S., Morimoto, C., Ihde, S.: Manual And Gaze Input Cascaded (MAGIC) Pointing. In: Proc. of CHI 1999, pp. 246–253 (1999)
16. Dragicevic, P.: Combining Crossing-Based and Paper-Based Interaction Paradigms for Dragging and Dropping Between Overlapping Windows. In: Proc. of UIST 2004, pp. 193–196 (2004)

A Scandinavian Approach to Designing with Children in a Developing Country - Exploring the Applicability of Participatory Methods

Nahid Wakil and Peter Dalsgaard

Participatory Information Technology Centre (PIT), Aarhus University
nahidwakil06@hotmail.com, dalsgaard@cavi.au.dk

Abstract. *Participatory Design* (PD) offers a democratic approach to design by creating a platform for active end-user participation in the design process. Since its emergence, the field of PD has been shaped by the Scandinavian context, in which many early PD projects took place. In this paper we discuss the challenges that arise from employing participatory methods in a different socio-cultural setting with participants who have had comparatively limited exposure to digital technologies. We offer a comparative study of two PD projects carried out with school classes in Scandinavia and India. While the setup for the two projects was identical, they unfolded in very different ways. We present and discuss this study, which leads us to conclude that PD can be a useful approach in both settings, but that there is a distinct difference as to which methods bring about fruitful results. The most prominent difference is the ways in which *abstract* and *manifest* participatory methods led to different outcomes in the two settings.

Keywords: Participatory Design, Developing Countries, Interaction Design, Future Workshop, Inspiration Card Workshop, Mock-ups.

1 Introduction

In this paper, we address the Interact 2013 theme of '*Designing for Diversity*' by examining how methods and techniques from Participatory Design (PD) translate from the setting in which they emerged, Scandinavia, to a rather different setting, namely an impoverished district in New Delhi, India. The aim of this work is to examine the role of PD and the challenges that arise from employing participatory methods in developing countries where access to and knowledge about technology and digital artifacts is yet relatively limited. As interaction designers increasingly work on projects outside of the settings in which participatory methods emerged and evolved, we find it pertinent to study if and how we can employ well known PD methods in new domains. Our work builds on the assumption that some methods may be employed more or less in their original format and setup, whereas other methods may need revision; some existing methods may prove of little or no use, while we may need to develop new methods to address issues at hand in new settings. Our work furthermore examines an assumption that we have encountered in various forums,

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 754–761, 2013.

namely that a relative lack of technological knowledge may limit the extent to which user can take part in shaping an interaction design process. In order to examine and challenge these assumptions, we have carried out a comparative study of two design projects with school children aged 12-19, one in India and one in Scandinavia. The setup and framing for the two projects was identical and consisted of employing three established PD methods: 1) a Future Workshop [Kensing & Halskov 1991; Vidal 2006] 2) an Inspiration Card Workshop [Halskov & Dalsgaard 2006] and 3) a Mock-up session [Ehn & Kyng 1991] to test, evaluate and develop the product. Since it is a single study, we focus on the specific findings; further studies are required to examine and develop more generalizable claims.

The contributions of this paper are two-fold. Firstly, we offer a comparative case study using the same methods within the same topic and frames in two design projects in very different domains, which in itself is rare in the field of HCI. Secondly, our analysis of the cases lead us to propose that the concept of abstract and manifest methods for user involvement can help practitioners select, frame and employ methods well-suited for the domain they work in.

2 Background and Related Work: Beyond the Scandinavian Heritage of Participatory Design

Participatory Design (PD) as a field is concerned with user involvement and decision making in the development of new technologies, in which "[…] the people destined to use the system play a critical role in designing it." [Schuler & Namioka p. xi] Originating in Scandinavia in the 1980s, PD methods and approaches have since found widespread use, most notably in Europe and North America. The participatory agenda has also to some extent inspired systems and services thatrely upon participation in use [Dalsgaard, Dindler & Eriksson 2008]. While methods and approaches inspired by PD have found an uptake outside of Scandinavia, there are few studies of how these methods translate to other settings. Recently, Zander, Georgsen & Nyvang [2011] have examined the potential contributions of PD through a case study of a participatory development project in Bangladesh, pointing out the need to further explore the potentials of PD in this region. Banaji [2012] offers a harsh critique of current IT projects in the global south by stating that many of them are "painfully ignorant of the everyday realities" in these domains. Our work can be seen as a response to these concerns, in that we study the commonalities and differences in employing PD in design projects in Scandinavia and South Asia.

Examining the origins and development of PD, Gregory [2003] identifies three characteristics particular to Scandinavian PD: a deep *commitment to democratisation*, discussions of *values in design and imagined futures*, and the use of *conflict and contraditions as resources for design*. PD seeks to create a platform for active end-user participation in the design process, although there is not a fully formed consensus as to the scope and consequences of involving users. While some argue that users can serve as *sources of inspiration* for designers [Christiansen & Kanstrup 2006], other contributions advocate a stronger role for users in design decisions. Kensing [1983] argues that there are three fundamental conditions of user participation in PD: *access to relevant information*, *the possibility for taking an independent position on the problems* and *participation in decision making* [Kensing, F., 1983]. These conditions emphasize an active engagement with end users, which have informed our studies in this paper. Muller [2002] argues that successful PD methods bring about a *'third space'*, a shared conceptual space in between designers' and users' respective domains, in which potential futures can be explored, developed and examined in collaboration. In the work presented here, we have built on these insights to establish what we label *direct user involvement*, based on the criteria that users should be involved in early stages of a project and take part in defining problems and visions as well as ongoing design decisions.

3 Case Study: Participatory Workshops in India and Scandinavia

In order to examine the applicability of participatory methods, we carried out two design projects with identical setups and framing. The first project took place in New Delhi, India, while the second project took place in Aarhus, Denmark. Here we present the setup of the design projects and the findings from India and Scandinavia, respectively.

3.1 Setting Up the Design Projects

We set up agreements with two schools to carry out a week-long design project with primary school pupils aged 12-19[1]. Each project was introduced and framed in the same way in order for us to examine the differentiation in use of participatory methods in a developing country and a Scandinavian country. The workshops were held in the children's classrooms in their respective schools and the theme for the workshops in both India and Scandinavia was *how technology can be used to improve everyday life*? We employed three established participatory methods to involve the users from the very beginning of the design project: Future Workshop [1] to define problems in the current situation, Inspiration Card Workshop [2] which was employed to develop ideas on how problems can be solved using inspiration materials and mock-up sessions [3] to test, evaluate and develop the product. In order to maintain the children's attention in the workshops, they were introduced to one phase at a time. And to maintain equal frames in both domains the teachers were present during all events. The children worked in groups and every group was asked to choose a spokesperson for the presentations after each stage. All events were documented in recordings and extensive field notes which were subsequently codified, condensated and analyzed.

[1] In the Scandinavian case, the school groups pupils in classes based on age, whereas the Indian school groups pupils on the basis of competencies, hence the spread in age.

3.2 Findings from the Workshops in India

Starting with the critique phase in the Future Workshop the children were asked to brainstorm over the problems they meet in their everyday lives. As the children only wrote down a few words within the given time, an ongoing adjustment of the estimated time became necessary. They were confused and uncomfortable with the situation. The children then were asked to write down at least five problems in the next five minutes. The time pressure made them work effectively but at the same time they needed to be confirmed that what they wrote down were “real” problems.

In the following fantasy phase each group were asked to choose 2-3 problems to discuss and solve using technology. In the presentations the children mentioned the opposites of the problems as solutions; e.g. the problem “mathematics” was solved by not having the subject at all. And the problem “being pushed in the bus” was solved by “not having many people in the buses”. The setup did not lead them to use their imagination to solve the problems, resulting in breakdowns in the ideation process. However, due to limited time we decided to move on to the next phase. In the Inspiration Card Workshop the children were introduced to the inspirational material in the form of inspiration cards and digital products such as cell phones, disc-players, iPods, calculators, CDs etc. The inspiration cards ended up in a row and were not moved any further. None of the groups combined the cards or developed their own cards. The blank cards, however, were used to write the definitive “answer” on. No new concept or idea was presented and the workshop resulted in seven already existing concepts directly copied from the inspirational material; the “mathematics” problem was solved by using a calculator and the “feeling alone” problem was solved by “talking on the phone with friends” etc.

Since some of the groups had expressed a desire for a cleaner city, we chose to work with the concept "Cleaning machine". The final concept was a trash can with a coin system, inspired by deposit systems. We developed a basic mock-up, which was first tested through plays and then evaluated and developed in a second workshop. To simulate an outdoor environment garbage was thrown on the classroom floor. The scene was presented to the children but the coin function was not mentioned, as we wanted to avoid predetermining the test. The children could buy chips and rice cakes on paper plates with paper coins in the shop. The trash can was located close to the shop and the teacher had written “trash” on the trash can in Hindi. Two children acted as salesmen and one boy was responsible for the trash can coin system. The rest of the class lined up in front of the shop and none of them noticed the garbage on the floor. Once the coin system was discovered all of the children wanted to get rid of their paper plates.

After the mock-up test the groups discussed what they liked and disliked about the trash can and how they could improve it. Although the groups still had the same spokesperson it was not easy for them to present their ideas. They were careful about saying anything “wrong” or “negative” about the trash can. Instead we asked them to draw a trash can using their imagination. Surprisingly, this gave very good results indeed. The children were creative and imaginative and used technology as a part of the solution. They improved the functionality, suggested a better way to interact with the trash can, mentioned the social aspect of how to get people wanting to use the trash can, suggested a better visual appearance, mentioned health problems that garbage can cause and how it could be solved and not least they considered the eco system to create power for robotic trash cans. Their independent input showed an understanding of the concept the mock-up has presented through a variety of creative ideas.

3.3 Findings from the Workshops in Scandinavia

In the critique phase the Danish children easily wrote down many things that they would like to change. This phase proceeded as expected and due to limited time the children presented only selected problems. In the fantasy phase the groups chose 2-3 problems to discuss and solve using their imagination and technology. The presented ideas were creative and imaginative and in the presentations the group discussed the presented concepts and ideas by asking each other question.

In the Inspiration Card Workshop the inspiration cards were used actively and the children also made their own cards. Few were inspired by the digital products and used sound as a part of the solution. In the presentations the groups defended their own ideas as the best. They had concrete scenarios in mind and were precise about the concept details. They were very specific in the choice of technology and it was obvious that they drew on their own experiences in the development of new concepts. The workshop resulted in five new creative concepts and the final product was a "memory box" based on the children's wish of an easier teenage life. The memory box had a scanning system to help the teenager remember things by placing her bag or wallet inside the box.

In the second workshop, the mock-up session, the memory box was tested, evaluated and to a certain limit improved. Three children were controlling the memory box by switching out the screens, changing the buttons to the right color and controlling the sound, which was played on an iPhone. One played the role of a busy teenager using the memory box while the rest of the class watched the play. After the mock-up test the groups were asked to answer a number of questions about how the memory box could be improved. Although the groups made a number of suggestions, the new memory boxes resembled the mock-up and had roughly the same functions and features. Many of the groups were locked into the idea of a square box with its already existing functions. One group made an egg shaped box and another came up with the idea of a camera function while the rest of the improvements referred to the box color.

4 Discussion

While both design projects led to interesting results and can be considered successful in bringing about new concepts through direct user participation, the ways in which the pupils arrived at them differed substantially. In the following, we compare the projects and introduce the notion of abstract and manifest design methods in order to account for these differences and discuss the applicability of participatory methods across different socio-cultural settings.

4.1 Comparing the Design Projects in India and Scandinavia

The first stages of the design process in India were problematic. The pupils were reluctant to engage in the design activities and were cautious in discussing events. They had trouble articulating problems and design opportunities, leading to breakdowns.

This required ongoing adjustments and interventions from the facilitator in order to keep the process going. The two workshops were characterized by limited creative output and independent proposals. However, the mock-up session really turned things around, spurring creativity and out of the box thinking that also led them to revisit aspects that had been troubling in earlier stages. It was only in this stage that a true 'third space' was established. In the Scandinavian design project, things unfolded in almost the exact opposite way. The pupils were initially open towards the project and eager to discuss both the general project and the specific events. In both workshops, they also worked effectively and imaginatively and put forward new ideas and proposals. In many ways, these workshops unfolded, as we would expect on the basis of previous experience and accounts in literature. However, the relative level of creative output waned in the mock-up session. The ideas put forward resembled the existing one, and the proposals for changes were superficial. Table 1 summarizes the projects.

Table 1. Summary of the design projects in India and Scandinavia

	India	**Scandinavia**
General attitude	Reluctant, cautious No discussions	Open, relaxed Eager to discuss
Future Workshop	Three breakdowns More time required	No breakdowns Work effectively on the tasks
Inspiration Card Workshop	Limited creativity Limited independent proposals	Good ideas Creative, imaginative
Mock-up	Creative, imaginative Active out of the box thinking	Design fixation Limited creative proposals

Comparing the two projects, the loose and open structure of the first two workshops seemed to limit the engagement of the Indian pupils, whereas it was well received by the Scandinavian pupils. If we reconsider the aforementioned criteria for PD, it proved difficult to establish a productive session in spite of providing access to relevant information and allowing for the pupils to take an independent position on the problems and participation in decision making [Kensing 1993]. The Indian pupils conceived of the facilitator as an authoritative figure possessing the 'right answers' in the design project up until the mock-up session. The Scandinavian pupils, on the other hand, had no problems taking on the roles of position as co-creators in the initial phases, however they experienced design fixation when exposed to the mock-up.

4.2 Abstract and Manifest Methods for User Participation

The contrast in how the two design projects in terms of active participation and creative output is quite striking. In light of our objective of understanding if and how traditional PD methods can be employed in different domains, we have analyzed setup of the three PD methods in more detail. If we first consider the Future Workshop, it can be construed as being relatively abstract, in that it offers a loose framing in which the pupils' point of departure was their own situation and preconceptions. The Inspiration

Card Workshop is also relatively abstract but with certain manifest properties, in that while it offers tangible representations of technologies and domain concerns but does not prescribe how they are to be combined; this relies on the pupils' preconceptions. Finally, the mock-up sessions can be defined as manifest, in that they offer a concrete artifact that brings the potential technology to life in a tangible form. In the two projects, we thus find a clear correlation between the abstract and manifest properties of the methods and the perceived creative output from the pupils (see Figure 1).

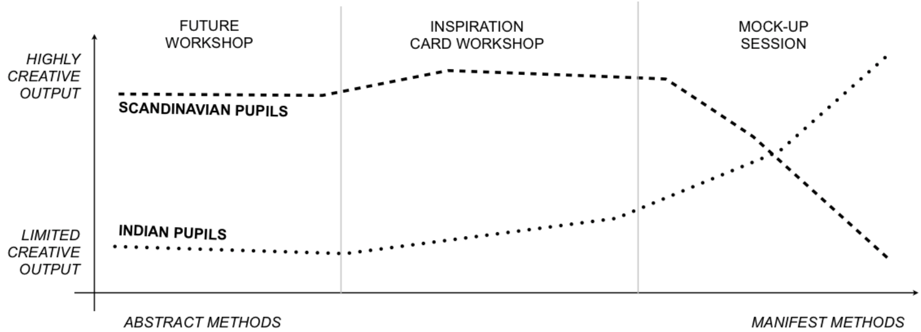


Fig. 1. Correlation between creative output and abstract/manifest properties of the PD methods

If we seek to understand why this correlation occurs, at least two prominent factors come into play: a) the Scandinavian pupils were accustomed to project-based group work, whereas the framing of school work for the Indian pupils was more traditional and rule-bound; b) the Scandinavian pupils were very familiar with interactive technologies and used them throughout their day, whereas the Indian pupils had relatively limited knowledge of them. Taken together, this indicates that if a design project is oriented towards technological solutions and the technological knowledge among participants is limited, the abstract representations offered in the traditional setup of the Future Workshops and Inspiration Card Workshops, which typically work well in Scandinavian settings, may not be sufficient to facilitate active participation. This is compounded if participants are not accustomed to the open and project-oriented format of the methods. On the other hand, the mock-up session as a manifest method was more productive by offering hands-on experiences and prompting participants to explore potential futures through construction.

5 Conclusions and Future Work: Applicability of Participatory Methods across Different Socio-cultural Settings

Our findings from the comparative case studies lead us to argue that PD can indeed be a useful approach in a domain that differs substantially from the Scandinavian setting in which the field emerged. However, it is also clear from our findings that existing methods cannot be expected to yield the same outcomes across domains, to the extent that some methods may not be advisable for use in their current form. While a multitude of factors affect the outcome of a design project – many of which are beyond the scope of a short paper to address or even introduce – we suggest that the concept of abstract and manifest methods can enrich our understanding of the outcomes of user participation. Abstract methods rely on preconceptions and imagination and limited use of materials. Manifest methods involve the use of materials and artifacts as the basis for inspiration and creativity. We speculate that many of existing methods that rely on abstract components may need to be revised if they are to be useful in radically new domains. Encouraged by the success of the mock-up session in the New Delhi case, we suggest that manifest methods may be a good starting point for participatory projects or events, and that researchers and practitioners consider developing new methods for user participation with manifest properties.

Concerning the validity and generalizability of the findings, our work relies on one comparative case study, and we most definitely need more studies and reports from PD practitioners in order to challenge, corroborate and expand upon these findings. We hope to explore these issues in future studies and invite our colleagues in the Interact community to take part.

References

1. Banaji, S.: Beyond Wild Dreams and High-Tech Fetishes: Learning about Media from Children in the Global South. Keynote Address, IDC 2012 (2012)
2. Christiansen, E., Kanstrup, A.M.: Selecting and evoking innovators: combining democracy and creativity. In: Proc. NordiCHI 2006, Oslo, Norway, pp. 321–330 (2006)
3. Dalsgaard, P., Dindler, C., Eriksson, E.: Designing for participation in public knowledge institutions. In: Proc. NordiCHI 2008, Lund, Sweden, pp. 93–102 (2008)
4. Ehn, P., Kyng, M.: Cardboard Computers: Mocking-It-Up or Hands-On the Future. In: Kyng, M., Greenbaum, J. (eds.) Design at Work: Cooperative Design of Computer Systems, pp. 169–196. Lawrence Erlbaum (1991)
5. Gregory, J.: Scandinavian Approaches to Participatory Design. International Journal of Engineering Education 19(1), 62–74 (2003)
6. Halskov, K., Dalsgaard, P.: Inspiration Card Workshops. In: Proc. DIS 2006, pp. 2–11 (2006)
7. Kensing, F.: The Trade Unions Influence on Technological Change. In: Briefs, U., et al. (eds.) Systems Design For, With and By the Users. North Holland (1983)
8. Kensing, F., Halskov, K.: Generating Visions: Future Workshops and Metaphorical Design. In: Kyng, M., Greenbaum, J. (eds.) Design at Work: Cooperative Design of Computer Systems, pp. 155–167. Lawrence Erlbaum (1991)
9. Schuler, D., Namioka, A. (eds.): Participatory design: Principles and practices. Lawrence Erlbaum, Hillsdale (1993)
10. Vidal, R.V.V.: The Future Workshop, Creative and Participative Problem Solving, pp. 1–21 (2006)
11. Zander, P.-O., Georgsen, M., Nyvang, T.: Scandinavian Participatory Design - Beyond Design, Beyond Scandinavia. In: Proc. of The Future of Development Research (2011)

Availability4D: Refining the Link between Availability and Adoption in Marginalised Communities

Fritz Meissner and Edwin Blake

Department of Computer Science / Research Centre in ICT4D, University of Cape Town
{fmeissne,edwin}@cs.uct.ac.za

Abstract. We present a comparative study of mobile and conventional computing technologies applied to providing access to career guidance information to high school students from marginalised communities. Reported high availability of mobile technology amongst these users would be beneficial, but our NGO partner questioned feature phones' applicability for consuming large quantities of information. We created two systems: a text interface exposed through a mobile instant messaging service, and a website targeting conventional computers. Despite positive usability tests for the website and fears of social stigma related to mobile instant messaging, system logging over eight months of parallel deployment showed convincing advantage in engagement for the mobile system. Interviews revealed that computer infrastructure was tied to institutions where access was limited; but greater access to mobile phones (owned or borrowed) made use and advertisement to peers of the mobile system easier. Social stigma was a problem only for a minority.

Keywords: availability, adoption, marginalised communities, feature phones, mobile Internet, M4D, NGOs.

1 Introduction

In late 2010 we were presented with an opportunity to collaborate with a programme called Link at the Cape Town NGO, The Warehouse [1]. The programme aimed to bolster the support available to high school students in marginalised communities of Cape Town as they made decisions that would affect their later success in the job market. They did this by organising career guidance workshops through church youth groups in the targeted communities.

The Link team (The Warehouse staff who ran the Link programme) wanted us to build a website which would support these workshops by providing students with access to information that would otherwise become stale if only presented in a workshop which ran once every few months. For instance, job openings discussed in a mid-year workshop would likely be filled by the time students were able to act on them at the end of the school year.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 762–779, 2013.

Recent research in M4D (Mobile technology for Development) involving fieldwork also performed in Cape Town [2, 3, 4] lead us to believe that it would be beneficial to disseminate this information by some means accessible via feature phones, which had achieved popularity amongst low income youth in Cape Town for accessing entertainment over the mobile Internet, especially mobile instant messaging (IM) services like MXit (a South African service with more than 50 million registered users [5]).

When we first mentioned mobile phones, the Link team shared their awareness of the popularity of the technology amongst youth, and added that its introduction could positively affect negative perceptions of mobile phones (Bosch records perceptions of MXit as time-wasting and harbouring sexual predators [16]):

"It [mobile technology] can penetrate further because you are sending it out to individual locations, and not one central Internet location, so for reach it's better."
– Link coordinator

"...it puts a positive spin on why kids should be using cellphones more effectively. Because at the moment there's such a lot of negative press about cellphones... so, if we can get it to be a more positive thing, that's certainly a good selling point." – Link staff member

However their idea of how it could be applied was limited to reminders which would inform students of when to seek out a computer from which to access new content on the website:

"...this is ... the limitation of mobile phones, is how much information can you access, and ultimately ... [you] will need to find an Internet cafe, but at least you'll know whether to actually bother to go and look for one or not, and that was the attraction of adding the mobile aspect." – Link Coordinator

Later in the conversation the Link team mentioned personal experience of problems viewing content on mobile phones, and some misgivings about the cost of airtime to the students. On the other hand, they were familiar with the capacities of the conventional web, and they already had a plan for reaching their audience: church groups who wanted to support teenagers in their communities could invest in the computer and Internet connection necessary for the website, which would also provide opportunity for interaction and mentorship.

Answers to their concerns about mobile technology did exist: text content need not be accompanied by more data hungry (and therefore costly) pictures or video; expertise learned from using popular mobile social networking platforms like MXit could apply for other purposes [6]; the platform had been used for the M4Lit study in which thousands of teenagers read a 21-chapter short story [3] and for the Dr Maths programme [14], which teaches students mathematics; people who learned to surf on mobile phones prefer the "familiar numeric keypad" to a traditional keyboard [4]. From our perspective, then, the technology had already been demonstrated suitable.

However, it would have been unwise for us to take an uncompromising stance on technology. Botes and van Rensburg highlight a "hard-issue bias" amongst researchers as a major cause of developmental project failure, as the debate can become a distraction from other important issues that must be addressed [7]. Proceeding alone was also unwise: we would not be able to make contact with a suitable group of users on our own, and according to Donner et al, M4D projects are more likely to succeed when the mobile technology element is an addition to a pre-existing developmental project [8].

Further, an honest assessment of existing M4D work would require us to raise some caveats: mobile phone use amongst these users is normally associated with entertainment [2]; for the "serious" purposes of school work and research on health topics, computers were more frequently used than mobile phones [2, 3]; in the M4Lit study the number who chose to finish reading the "m-novel" was only a fraction of the number to whom it was advertised (and similar advertising would normally cost a high fee) [3]; and although people are capable of using relatively complex technology to access the content that matters most to them, their priorities might not match ours [9]. Further, a discussion of sustainability concerns would reveal that both M4Lit and Dr Maths had the backing of large research organisations like the Shuttleworth Foundation [10] and the Meraka Institute [11], organisations with far greater resources to dedicate to ICT concerns than The Warehouse could bring to bear on Link. The Warehouse already maintained one website and the Link requirements did not necessitate any change in technology for the new site.

The point was moot: although the Link team were insistent on developing a website for access from conventional computers, they were happy that we follow that up with a mobile effort, and were willing to let us evaluate the two systems with the same users. Having two systems on platforms of differing availability would allow us to investigate the impact of availability on adoption.

2 Research Methodology

2.1 Action Research

The dual goals of development (providing students with a new channel for accessing information from Link) and research (investigating adoption) matched well with the Action Research framework [12]. Our intention to pursue two different solutions, one after the other fit easily into the cyclical approach of the framework, wherein action precedes evaluation and then more action, based on the outcome of the previous evaluation.

Early results in action research projects shape later methodology, but can also prove interesting in the scope of the project as a whole, and so we report separately on formative (earlier work, relating to design, development and refinement of our systems) and summative (later, comparative) cycles.

2.2 Venues

We operated in four different venues, shown in Fig. 1. Lavender Hill (yellow) and Manenberg (green), both designated "coloured" residential areas under racially discriminatory South African apartheid-era legislation [13] were home to church groups with whom Link had been working since before we joined the programme in 2010. The location of The Warehouse NGO, home of the Link programme, is marked by a red pin.

In 2011 the Link programme began a "homework club" at a church in Mowbray (blue). Mowbray, being formerly designated a "white" area was not disadvantaged by apartheid legislation, but the students who attended were isiXhosa speaking residents of informal settlements (shanty towns) not shown on the map which were formerly designated "black" [13]. These students attended a school in the area, and were attracted to the church by flyers advertising weekly help with homework that were given out at a nearby transport hub which they used daily.

2.3 Participants

The beneficiaries whom we interviewed and with whom we tested were either introductions from Lavender Hill and Manenberg church groups, or students whom we tutored (assisting with school work in mathematics and physical science) at the Mowbray homework club. Our interaction with students in Lavender Hill and Manenberg was restricted to two visits each, for user testing of the Link website. At Mowbray, we were able to engage directly for two to three hours weekly for almost two full school years in 2011 and 2012.

At each venue we worked with a subset of all students, either selected by the church groups or by Link, usually based on whether there were other plans for those participants' time on the day that we visited. We therefore did not have control over our samples. Only in the latter stages of the project at the homework club did we have a direct relationship with students that gave us insight about their technology use habits which could help us to select interviewees according to the data we hoped to gather. Even in that case, we were still constrained by which students would arrive for tutoring on a given week, and by a need to balance time as researcher with availability as tutor.

When describing evidence relating to an individual student, we use initials to protect their identity.

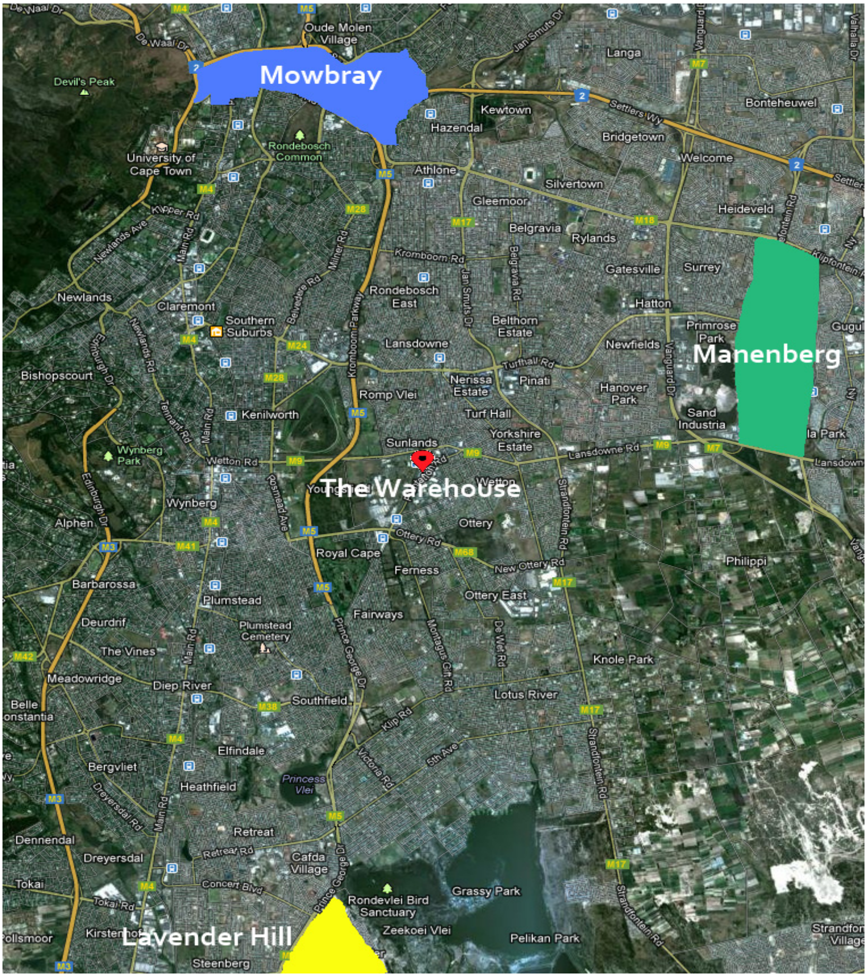


Fig. 1. Map of the southern suburbs of Cape Town, showing the locations in which we worked

3 Artefact Descriptions

We describe here the two systems that we designed and implemented in order to help the reader follow discussions of user reactions in later sections. The web system is known throughout the paper as the “Link website”, while the mobile system is called “LinkChat”. Information is exposed in the same format in each, as “entries” which are a discrete piece of information such as a job advert or a description of a university course. A single content management system, maintained by the Link team, serves search results and content to both systems.

3.1 Link Website

The website we created allows users to perform full text searches either of the whole site, or restrict themselves to information in one of three categories: study (tertiary courses and bursaries), jobs (job adverts and internships), and skills development (short courses and internships). A fourth section, start a business, was inoperable for the duration of our study. Entry detail pages (see Fig. 2) can be chosen based on their title and the first few lines of description on a search results page similar in appearance to Google.

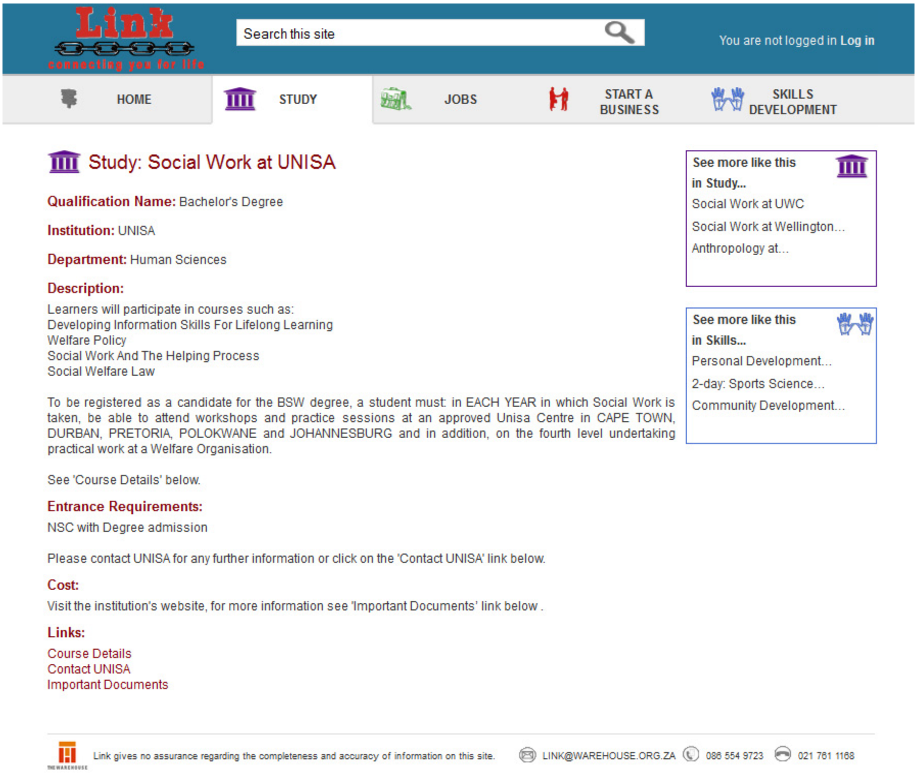


Fig. 2. Website entry detail page

3.2 Mobile System – LinkChat

The mobile system we developed was named LinkChat for the fact that it communicates with users via mobile IM. Although it could be reached from any IM platform which interoperates with Google Talk, all of our users used the MXit platform. Messages are relayed using the XMPP IM protocol [17] to and from our server-side code. Initially developed as proof of concept only, the system serves content in only one of the three categories that the website does: study. Considerably more data has been captured in the study category than in the others.

Fig. 3 shows the process of a search using LinkChat as three consecutive screenshots – the system prompts the user to search, in response to which the user sends a search query. The system then responds with a numbered list of results. After sending the number corresponding with an interesting item, the system sends the entry detail to the user.

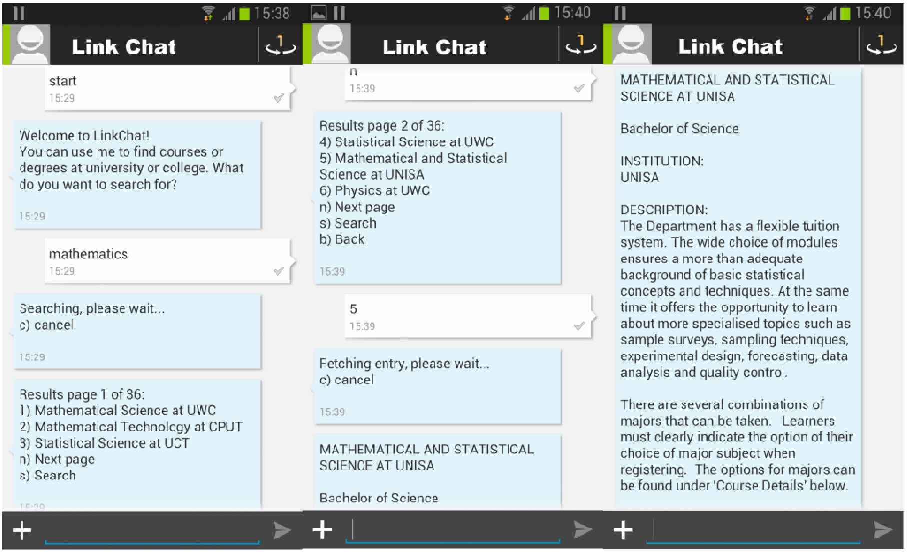


Fig. 3. Screenshots (read in columns top to bottom, left to right) showing the LinkChat search process, choosing an entry from search results and viewing academic course information

Note that for legibility's sake, screenshots in Fig. 3 have been taken with an Android smartphone which is more advanced than the smaller feature phones normally used by beneficiaries. Fig. 4 shows an example of a more typical user device, the Nokia 2700.

4 Formative Work

In the timeline of our engagement with Link, we pause in early 2012 at the conclusion of the first round of user testing of the mobile system, which was also the last work of the formative cycles. Before that point, we had spent time learning about our users and the Link programme, as well as collaborating with Link on the design, development and evaluation of the website. We report the findings that have relevance to the final project results from each of these activities in roughly chronological order.

Fig. 4. Nokia 2700 feature phone typical of devices owned by users

Table 1. Formative Link beneficiary technology use survey

Measure	Location Warehouse (n=9)	Manenberg (n=13)	Lavender Hill (n=2)	Mowbray (n=11)	Total (n=35)
Ever used computer at home	3	9	1	3	16 (45%)
Ever used computer at school	7	9	1	1	18 (51%)
Ever used computer at library	4	0	0	6	10 (29%)
Ever used computer at Internet cafe	0	2	1	1	4 (11%)
Ever used computer (Total)	9	13	2	11	35 (100%)
Computer yesterday or today	1	7	2	2	12 (34%)
Cellphone yesterday or today	8	13	2	11	34 (97%)
Have used Google on computer	6	10	2	9	27 (77%)
Have used MXit on cellphone	7	11	2	10	30 (86%)

4.1 Learning about Beneficiaries – Technology Use Survey

In order to learn about the work of Link and its beneficiaries, we attended workshops at The Warehouse and at partner churches (see Section 2.2). While at these locations we asked participants about their technology use habits (in fact we also asked participants at later evaluations of the website, but for convenience we report all responses together). The results (see Table 1) confirmed our expectations that far more of the students would have very regular access to mobile phones than computers, and that most would be familiar with MXit. Somewhat surprising was that all had used a computer at least once, and most students had used the Internet on a computer in the form of Google – a positive sign for our website.

4.2 Website Evaluation

In preliminary evaluation of the website in August 2011 at Lavender Hill and Manenberg (see Section 2.2), users had significant difficulty using the site, but after changes to the user interface a larger second evaluation (twenty students in twelve groups at all three church venues, employing constructive interaction [15]) showed that if relevant data had been captured by the Link team, our site could help users to find it:

- Only two users expressed doubt about their ability to use the site on their own
- When given general instruction to use the site, in all but one case participants acted by searching, without needing to be told how
- A lack of computing skill and awareness of web search norms slowed task completion, but only in one extreme case did it prevent task completion
- Despite this being their first time using it, students began to develop skill at adapting their input to forms that the site could better work with.

We were aware that the constructive interaction method allowed the students whose computer skills were weaker to hide this from us, but in the context of use envisioned by the Link team, students with poor computer skills should be able to receive assistance at computers from church staff or fellow youth group members. At this stage, it appeared as though the Link team's original plan of information dissemination through the website might be adequate.

4.3 Changing Relationship between Link and Churches

In late 2011 contact between Link and the Manenberg and Lavender Hill churches became less frequent. As a result, we did not visit either group again, and the focus of our engagement was solely on the Mowbray group. Significantly, we (researcher and Link team) interacted with this group directly, rather than having a layer of church leadership between us and beneficiaries. Because students were not directly affiliated with this church, it did not assume responsibility for providing access to infrastructure in the way that the Link team had hoped.

4.4 LinkChat Evaluation

LinkChat formative evaluations were intended to follow the same process as development of the website, but only a small first evaluation and the first part of the second evaluation could be run before spontaneous unsolicited usage lead us to abandon controlled evaluation. The following notable reactions from our users occurred in that first evaluation:

- *SN* was very enthusiastic about MXit and skilful in general mobile operation, but not enthusiastic about LinkChat: she lost interest in the evaluation after a brief attempt at use, choosing to message friends instead while we worked with other participants. She did ask for the LinkChat contact name for later use for the whole group.
- *LA* told us that she did not use MXit, and was reluctant to discuss her reasons. She was willing to try using LinkChat, and was capable of entering text for a search. However, she decided to stop using the system before she viewed an entry. She specifically refused a piece of paper with the LinkChat contact name on it when we wrote it down for others at the table with her.
- *NK* used LinkChat from her own MXit account on our phone. She demonstrated the ability to operate it, but said that she would prefer to use the website. She asked us to write down the website address, expressing confidence in her ability to access and use computers at the library.
- *OM* performed several searches, stopping only when the venue was closed for the day. He had indicated at the start of the session that he hoped to leave almost 15 minutes before he eventually did, from which we deduced his enthusiasm.

These results demonstrated users' ability to operate LinkChat, but their responses demonstrated almost every outcome imagined in our initial dialogue with The Warehouse: MXit and mobile phones preferred for entertainment (*SN*), outright rejection of MXit (*LA*), computers preferred for content consumption (*NK*), and unchecked enthusiasm (*OM*). The outcome of our planned comparative evaluation apparently rested on which of these users' attitudes the majority of our eventual audience reflected.

5 Comparative Work

The move from formative to comparative work occurred swiftly; instead of waiting for us to complete changes to LinkChat and launch it at a specific event like we had the website, students from our first LinkChat evaluation began to use the service whenever it was online. With the website already online, usage data for both systems began to accumulate before we had planned it.

5.1 Methodology

System Logging. Log files for the website were gathered between 22 November (official launch) and 31 October 2011. Log files for LinkChat were gathered from the day of first unsolicited use on 24 February 2012 until 31 October 2012. As far as possible, users identifiable as non-beneficiaries have been excluded from the logs. This was easiest with LinkChat, because all communication on MXit is tied to a user account which made it possible to identify users who were not students when calculating usage. It is likely that some visits from non-beneficiaries remain in our website logs.

Audience. The Link team had advertised the website's launch to 18 students (all but one from Lavender Hill), and in our formative evaluations of LinkChat in early 2012 we discussed or evaluated the site with a further four beneficiaries, for a total of 22. The Link team also advertised to colleagues at The Warehouse, but we do not have an accurate record of to whom or how many.

Students did not appear to need specific instruction to use LinkChat in their own time after learning about the service, making demonstration or evaluation equivalent to advertising. We advertised in this way to eight students, including the five discussed in section 4.4 above. LinkChat was also shown to non-students, including Link staff and colleagues in our research group.

Apart from these numbers, both systems were demonstrated to students who attended the Mowbray homework club on April 17, and the following week flyers advertising both systems were handed out. Unfortunately we do not have attendance records for those weeks, but over a four week period the next term the average number of students who signed the attendance register each week was 31.

Interviews and Demonstrations. Between late March and Mid June 2012 we conducted semi-structured interviews with six students who had were regular attendees at the Mowbray homework club, and demonstrated the two systems to five newcomers who only attended the homework club for the first time after our April advertising. In the interviews we asked students about their search behaviour before and after the Link intervention; in the demonstrations we attempted to understand students' operation skill while guiding them through the use of LinkChat and the website.

Reported Usage Questionnaire. After our April demonstrations, we handed weekly questionnaires to students at the homework club for eight weeks. The questionnaire asked students to inform us what searches they had performed on each system in the previous week, with the aim of supplementing our system logs by providing a way to identify users as beneficiaries or not. We report this primarily to acknowledge that the form may have had some effect as a reminder about the systems; as a source of data it was poor. Students often left without completing it, or filled it out incorrectly. The most useful information recorded was obtained after we included a section for suggestions they could give us about how to improve the systems. These suggestions were mostly requests for content on new topics, and not relevant to our question of availability and adoption.

5.2 Quantitative Results

Comparative numerical results from system logs are shown in Table 2. Despite being deployed for longer, the only measure which is higher for the website than for LinkChat is the number of unique users recorded. This number is subject to quirks of web analytic tools – if a visitor used more than one browser, or cleared locally stored website tracking data after their last visit, they would be recorded as a new and different user. By contrast, users communicating with LinkChat were identified by their MXit

ID, which provided a more reliable number. We note that 102 of the web users (92%) visited on one day only while 34 chat users (56%) visited on five or more different days.

Other numbers are significantly higher for LinkChat, despite it being available for almost three months less than the website, as LinkChat visitors engaged more and more frequently.

Table 2. Numerical results from LinkChat and website system logs

System	Measure					
	Days Live	Unique Users	Searches Performed	Entries Viewed	Searches / Day	Entries / Day
LinkChat	251	57	811	796	3.23	3.17
Website	345	117	116	52	0.34	0.15

The higher usage of LinkChat is consistent over the entire duration of its deployment. Fig. 5 shows a chart of the number of searches performed per month on each system, while Fig. 6 shows the total number of daily visitors in each month. The only month in which the lines representing the two systems touch is February 2012, in which LinkChat was only online for six days. The maximum in LinkChat usage (both graphs) is due to diffusion in March 2012, while the increase of usage on the website in October 2012 (Fig. 6 only) can be explained by a number of visitors who discovered the site through Google, i.e. not people to whom we had advertised the site.

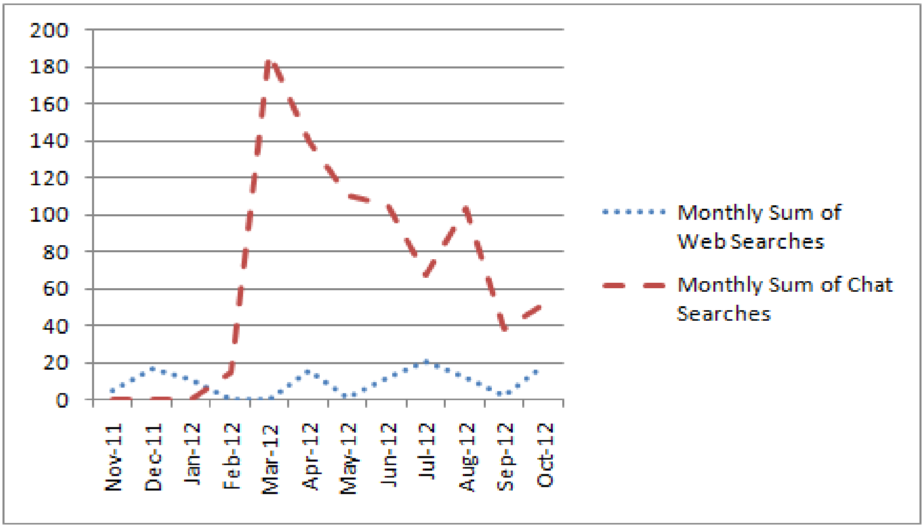


Fig. 5. Total number of searches performed on LinkChat and Link website per month

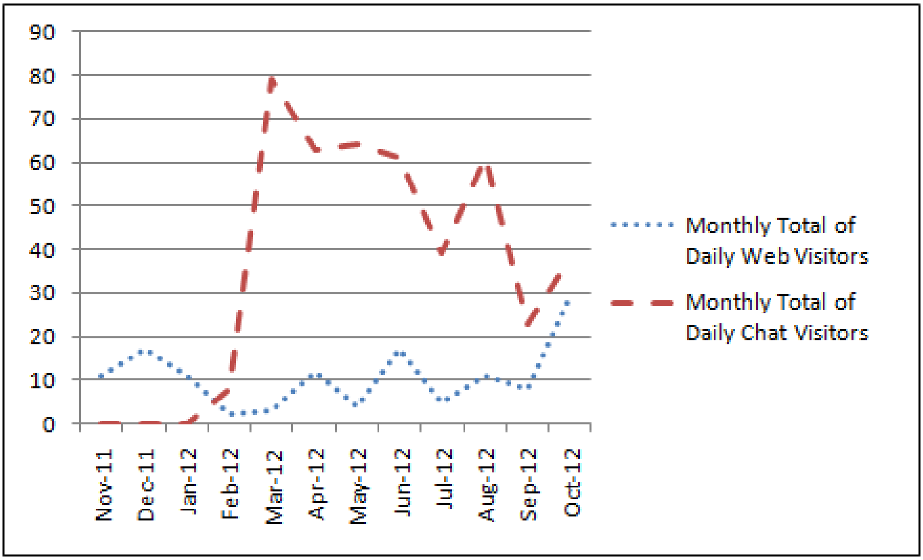


Fig. 6. Total daily visitors per month to Link website and LinkChat

6 Discussion

The clear quantitative advantage for LinkChat was already visible by March 2012. Although other studies had demonstrated the phenomenon of mobile adoption, this result was still interesting because it was in opposition to both the expectations of the NGO and some of our formative results. In order to refine our understanding of the results we saw, we began to gather evidence from interviews, observations of demonstrations, and observations in the formative cycles to understand the greater popularity of LinkChat. In so doing we hoped to discover design implications for future work in marginalised communities.

6.1 Unhelpful Existing Computer Infrastructure

Evidence we gathered about computer access was related to home, school or public library. What we learned underlined the lack of access which students had to conventional computing resources.

Home. *AM* explained to us that it was not normal for his peers to have access to computers at home, estimating that perhaps twenty percent of his classmates at school might have a computer in the home. Even then, he suggested that there would be family members with other priorities for the computers, such as a brother who wanted to play games (his hypothetical example). Others who spoke about computers at "home" confirmed that access was only possible for them because of their relationship with the owner: *MG* used his older sister's work laptop "once or twice a month" when he visited her, *NM* mentioned her father having a computer for work and *AS* had access to a computer once a week when he visited his mother's former employer.

School. *NK* estimated that less than a quarter of her school was in the science "stream" that was given priority use of the school computers. When access did happen for the science students, it was shared with others at the same computer. *MG* was in the science stream, but when we asked how often he could use a computer he referred to visits to his older sister rather than to time at school. This may mean that his access at school was less frequent than visits to his sister, or that he did not see the school computers as relevant for the purposes we spoke about (they may have been reserved for school work only).

Library. A public library was near to the students' school and the tutoring venue. Library cards are free and we heard of the students taking shelter there on rainy days after school. The library had working computers, but the queue for use was long, and the allowed period of use was less time than *NK* and *YM* wanted. "Not good... I didn't finish my things", said *YM*, who felt that the worst part of the experience was the slow Internet connection. *NK*'s priority at the library was using Google to search for information about university courses. Despite her expressed preference for the Link website when she saw LinkChat (see section 4.4 above), the site had "slipped her mind" when she was actually at a computer. Having used LinkChat, she felt that while in front of a computer she preferred the breadth of Google results to the specific content furnished by Link. *YM* searched for similar content, but also mentioned social networking as a higher priority: "I have to Facebook first... have to check".

Lack of Skill and Confidence with Computers. In *AM*'s opinion, the lack of computer access lowered his and his peers' ability to operate computers: "I'm not used [to computers]", making access when it did occur less fruitful. *YM* and *NK* also found information search on computers slow work: "It's a process", said *YM*.

6.2 Confidence, Expertise and Convenience with Mobile Phones

Students had greater access to mobile phones than to computers. Some students had their phone with them at the homework club, which other students did not seem to find unusual. *NK* and *YM* did not have phones with them, but spoke of using a friend or family member's device. Mobile phones were therefore not restricted by venue.

Cost too, was not problematic, at least on MXit; some students had SIM cards from an operator which gave them free data for MXit. *OM* estimated that an evening on MXit chatting to friends would cost him R0.30 (0.03 USD), while *MG* told us that he preferred LinkChat to the Link website because it cost less.

For *CM*, greater access to mobile phones had created comfort with feature phone interfaces. She preferred Linkchat on the smaller screen of a feature phone to the website on a 15" laptop because a full sized computer screen took more time to scan for the information she wanted. *AM* contrasted his experience with mobile phones with his reservations about computers, saying "I know the phone". When we used a feature phone as conversation aid during an interview, *MG* told us that we were operating the phone too slowly.

A theme of convenience regarding LinkChat emerged as a result. *AM* preferred LinkChat, because "It doesn't waste time," and *YM* felt it was useful "... when in a hurry". *NK* preferred to use Google to the Link website (despite her earlier expressed preference for the website over LinkChat – see Section 4.4), but that expressed preference did not seem to be an issue in practice – she was the most frequent user of LinkChat, performing 118 searches on 37 different days.

6.3 Unsolicited Use of LinkChat

Some of the students who saw LinkChat in our evaluations continued to use it before we had intended them to do so. *SN*, whom we reported on in Section 4.4 had asked us to write the name of the LinkChat contact down but had not messaged it any further then, communicated with LinkChat a few days after. *OM*'s visit the next day took us by surprise, because we had not written the contact down for him. We later realised that MXit stores users' contacts on its own servers rather than the device on which the chat client is installed, and so students who used their own accounts on our demonstration phone had the contact ready when they next signed in on their own phones.

6.4 LinkChat Diffusion amongst Peers

More LinkChat users had been recorded in system logs than we had advertised the system to directly, and we met students who were attending the homework club for the first time, but who already knew of LinkChat. We learned of three ways in which peers had passed on knowledge of the system.

Face to Face. *NK* and *YM* told us that they had shown LinkChat to school friends on the friends' phones.

Online, through MXit. Although we had seen MXit purely as a platform for disseminating text, its social features also proved important. *OM* told friends in another city about LinkChat while messaging with them on MXit. *CM* gave a friend at a different school her MXit password so that the friend could use her own phone to log in and message the LinkChat contact from *CM*'s account.

Classroom Demonstration. An unusual, but noteworthy incident demonstrated the difference between the two systems in terms of opportunities for diffusion. Two students whom we had tutored took it upon themselves to advertise our systems. On March 11 2012 we received the following message on Facebook from *MG*:

"Me and (AS) have came up with a briliant idea on how we can spread the word about your wabsite and we gona d it at school starting from tomorrow yeah. And wa our names there if theres space on the 'thank you list' bt if ther isnt no sweat w doin this 4 ya"

We were pleased with *MG*'s initiative but uncertain about how we could acknowledge his contribution and how the Link team would respond to the idea of singling out students. We suggested that we talk at the tutoring programme the next day about the idea, but the conversation did not take place because we did not see *MG* there. Both had been part of our second formative evaluation of the Link website, and because of the wording of the message we assumed that they proposed to publicise only the Link website (we had not personally introduced them to LinkChat).

On the 15th, *MG* sent a free "please call me" message to us, and when we responded he informed us that he wanted to demonstrate LinkChat at school but that it was offline. We were making code changes at the time, but started the server so they could proceed. By the end of the day twelve new MXit IDs had used LinkChat to perform 89 searches – the most of any day before or since.

MG later informed us that he and *AS* had told their Life Orientation teacher at school of the two systems. The teacher had asked them to demonstrate LinkChat to their class, but told them that it would not be possible to demonstrate the website until the school computer room was available for students to visit. When the demonstration of LinkChat took place, the website was not mentioned.

6.5 Reaction to LinkChat from Non-MXit Users

The question raised in Section 4.4 of which observed reactions to LinkChat would be most typical of students was avoided by the spread of LinkChat beyond the Mowbray group. Regardless of how many of the students at the homework club were using LinkChat, it had proven better at attracting users than the website. Nonetheless, the question of why some users were negative and unresponsive about MXit remained.

One student in this position was willing to discuss: *NM* struggled with reading, and this made her uninterested in computers or mobile phones, especially text-centric IM.

In order to understand other reasons, we asked *MG* – himself an enthusiastic mobile phone and MXit user, but similarly to Bosch's interviewees, aware of negative perceptions [16] – what reasons he thought people might have for not wanting to discuss the topic. His opinion was that these students were a minority. A few might be embarrassed about only having "tilili", a slang word with which his peers described a phone with no features beyond voice calls and SMS. MXit also had "a bad side and a good side", and some students might be constrained by society's expectations of them because of their parents' position: "...because I'm a pastor's son, I'm gonna ruin his reputation [if I use MXit]". The same might go for teachers' children, although *MG* also noted that some of his teachers at school used MXit.

7 Implication for Design – Complications of Context

At the start of our engagement, we understood our dialogue with the Link team in terms of three variables: cost, availability and fit for purpose. We agreed that availability would be better for mobile phones, but disagreed about cost and fit for purpose. A greater awareness of technical information at the start of the project allowed us to correctly predict the broad outcome, but subsequent qualitative enquiry revealed more nuance. The qualitative data demonstrates a greater range of contextual factors:

- Setting, affecting the availability of technology through institutional rules: mobile phones were not tied to a single setting and so even though it is unlikely that no institutional rules apply at all, students did not raise difficulties of the sort faced when using computers, such as library time limits.
- Software, affecting the range of operations that technology can perform: the MXit software provided an interface with which students were familiar, and social features through which knowledge of LinkChat could spread.
- The user, who has capacity to operate a technology in different ways: students were familiar with mobile phones, and as a result skilled, confident and willing to read for long enough periods to consume information.

- Personal resources, which can be sufficient or insufficient for a given purpose, or offer workarounds making a particular use redundant: personal resources included devices on which to consume data, time, and money for airtime if necessary.
- Surrounding persons, with the ability to supplement resources, but also bringing social norms which may restrict access: other students could lend a phone when one personally owned was not available, or knowledge of how to operate LinkChat to aid a novice. On the other hand, some parents might consider their children's use of MXit a negative effect on their own reputation.

Most of these factors worked in favour of the mobile system we built, but will affect adoption in different ways in other contexts. They may, like the negative social implications of mobile IM we encountered, drive people away from a technology even while other factors are positive.

This list of contextual factors is not necessarily exhaustive; other data may reveal more. We note also the inter-related nature of factors; for instance personal resources could include a SIM card from the operator Cell C, making the cost zero when the software in use was MXit because of an agreement between the two.

8 Conclusion

Our work has shown how using mobile phones to access the internet suffered from misconceptions from the NGO supplier of online content (e.g., that mobile phones are not suitable for consuming large amounts of content) as well as negative associations amongst some members of the target user group (e.g., chat systems have negative moral implications). In the context of an NGO-led intervention in marginalised communities of Cape Town, we developed two systems, one a website for use on conventional computers, and the other a text interface for consumption on the mobile instant messaging platform MXit that works on feature phones.

We show that amongst marginalised youth in urban South Africa, mobile instant messaging as platform for content provision has a substantial advantage over the conventional web. It lends itself to word of mouth adoption, conveys information adequately in spite of the limited display capacity and access is cheap enough not to be an obstacle to adoption. We show that the platform is popular and well-suited to developmental purposes. A qualitative investigation into the reasons for its popularity revealed that a range of contextual factors caused this advantage: the technologies were affected by setting, software in operation, users' capacity for operation, their personal resources and surrounding persons' resources and attitudes.

References

1. The Warehouse, Serving the Church in its Response to Poverty, Injustice and Division, `http://warehouse.org.za/` (accessed: January 17, 2013)
2. Kreutzer, T.: Internet and Online Media Usage on Mobile Phones among Low-Income Urban Youth in Cape Town. International Journal of Education and Development using ICT 5(5) (2009)

3. Vosloo, S., Walton, M., Deumert, A.: M4Lit: a teen m-novel project in South Africa. In: Proceedings of mLearn 2009, Orlando, FA (2009)
4. Donner, J., Gitau, S.: New Paths: Exploring Mobile-Only and Mobile-Primary Internet Use in South Africa. Presented at the Workshop on Mobile Media in South Africa. University of Cape Town (2009)
5. Mxit - Home, `http://site.mxit.com/` (accessed: January 21, 2013)
6. Donner, J.: Blurring livelihoods and lives: The social uses of mobile phones and socioeconomic development. Innovations: Technology, Governance, Globalization 4(1), 91–101 (2009)
7. Botes, L., Van Rensburg, D.: Community participation in development: nine plagues and twelve commandments. Community Development Journal 35(1), 41–58 (2000)
8. Donner, J., Verclas, K., Toyama, K.: Reflections on MobileActive08 and the M4D Landscape. In: Proceedings of the First International Conference on M4D, pp. 73–83 (2008)
9. Smyth, T.N., Kumar, S., Medhi, I., Toyama, K.: Where there's a will there's a way: mobile media sharing in urban india. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 753–762 (2010)
10. The Shuttleworth Foundation is an innovative non-profit organisation, `http://www.shuttleworthfoundation.org/` (accessed: January 22, 2013)
11. The Meraka Institute: CSIR research and development in information and communications technology (ICT) `http://www.csir.co.za/meraka/` (accessed: January 22, 2013)
12. Hayes, G.R.: Taking action in your research. Interactions 19(4), 50–53 (2012)
13. Besteman, C.: Transforming Cape Town. University of California Press (2008)
14. Butgereit, L.: Math on MXit: using MXit as a medium for mathematics education. In: Meraka INNOVATE Conference for Educators, CSIR, Pretoria, pp. 18–20 (2007)
15. Jones, M., Marsden, G.: Mobile interaction design. John Wily & Sons Ltd., West Sussex (2006)
16. Bosch, T.: Wots ur ASLR? Adolescent girls' use of MXit in Cape Town. Commonwealth Journal of Youth Studies 6(2), 52–69 (2008)
17. Internet Engineering Task Force (IETF), "Extensible Messaging and Presence Protocol (XMPP): Core, `http://xmpp.org/rfcs/rfc6120.html` (accessed: January 22, 2013)

Communication Choices to Engage Participation of Rural Indonesian Craftspeople in Development Projects

Ellya Zulaikha and Margot Brereton

Electrical Engineering and Computer Science, Faculty of Science and Engineering,
Queensland University of Technology, 2 George St,
4000 Brisbane, Australia
ellya.zulaikha@student.qut.edu.au
m.brereton@qut.edu.au

Abstract. In participatory design projects, maintaining effective communication between facilitator and participant is essential. This paper describes the consideration given to the choice of communication modes to engage participation of rural Indonesian craftspeople over the course of a significant 3 year project that aims to grow their self-determination, design and business skill. We demonstrate the variety and subtlety of oral and written forms of communication used by the facilitator during the project. The culture, the communication skill and the influence of tacit knowledge affect the effectiveness of some modes of communication over the others, as well as the available infrastructure. Considerations are specific to the case of rural Indonesian craftspeople, but general lessons can be drawn.

Keywords: Communication Mode, Rural Craftspeople, Participatory Design, Participatory Development.

1 Introduction

In theory the use of Information and Communication Technology (ICT) should assist rural business to access and compete in global markets. However, scholars widely acknowledge that ICT application in rural business is problematic [6,9]. There is a lag in uptake of ICTs in rural areas and review of a wide array of disparate initiatives found that interventions are best undertaken through multi-channel approaches and internediaries [3,5]. However, imposing the use of ICT may not always meet the needs of rural people nor might ICTs suit the local context. Instead of beginning with the assumption that ICT in some form is needed to enhance the livelihoods of rural people, this project begins by examining how rural craftspeople communicate and how they engage to participate in projects to develop their livelihoods. Such understandings would in any case need to underpin efforts to adopt ICTs. This paper describes a case study of a participatory approach to engage participation of rural craftspeople in a development project, focusing upon the communication modes used by the intermediary.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 780–787, 2013.

Rural craftspeople producing traditional handmade products currently struggle to stay in business. The common problems of rural craftspeople are lack of innovation, management skill and capital. One rural craft industry which currently faces these difficulties is the Jombang glass bead craft industry in Indonesia. There used to be hundreds of craftspeople running glass bead businesses, but in the last decade the number has decreased to just over 20, and it is still decreasing.

Advisory programs have been provided to address these problems such as design and management training, and facilitation to sell craft products. Such advisory programmes are common for rural craft industries in Indonesia. While these programs offer value and give new insights to craftspeople, they do not fully engage the imagination, capacity and innovation of craftspeople. As top-down policy initiatives, these programs overlook the unique potential of the participants as well as the local context. Additionally, the complexity of the rural craft community with its inter-related aspects of communal life was ignored.

This action research project has undertaken participatory design with rural craftspeople in order to explore how they can develop their craft practice and livelihoods. Participatory design involves them fully in decision making processes, is grounded in their local context, respects their skills, allows them to exercise self-determination, and respects the unique potential of each participant.

Communication is one key to success of a participatory project. Therefore, it is necessary to examine how communication occurs and effective modes of communication. This knowledge can inform application of ICTs for rural craftspeople.

2 The Participatory Approach

Participatory methods have been used in many fields. The fundamental tenet of participatory methods is empowering participants to have a better say in determining their future. Good communication enables better quality participation. However, engaging participants can be challenging. As stated by Bebbington et.al, rural people will not share their problems nor their ideas in situations that are not conducive [2]. Knowledge of cultural sensitivity and the local political situation is essential to creating conducive situation. Interpersonal and political skills of facilitators are also needed to create conducive situations [11].

2.1 Communication as an Essential Aspect of Participatory Project

The most crucial phase of participatory projects is “entry to the field” and building an initial relationship with participants [4,11]. Nevertheless the success of this phase depends upon the communication skill of the facilitator.

In the literature that examines the effectiveness of different communication channels, Maltz argued that the chosen communication medium will affect the quality of information, in terms of its richness, spontaneity and speed [10]. Hollingshead argued that communication medium is less influential on the quality of group decision making [7], which is influenced more by the equality of social status among members in a group. However, Hollingshead did not consider the role of facilitators. In participatory projects involving internediaries, the quality of communication and resulting decisions also depends on the facilitator's skill to mediate the discussion in order to avoid domination etc.

2.2 The Challenge of Initiating the Participatory Project

Different societies hold different understandings of participation based on their local value systems [13]. Facilitators should understand the values in the community [13] and how to deal with them before initiating a participatory project. Western techniques for engaging participation are not necessarily suitable for eastern cultures [13]. Negative consequences of relationships caused by misunderstandings can occur. There has been little research exploring communication for engaging participation from the context of eastern culture.

2.3 Overview of the Three Year Participatory Project

The communication strategies described below have been developed as part of a three year project. The first phase in year 1 involved ethnographic study, interviewing craftspeople about their livelihoods, artistic and business perspective and attitudes towards sharing. The second phase in year 2 involved a series of collaborative design workshops between industrial design students from the local urban university and the rural craftspeople in which they worked together to develop new designs that both reflected their own values and local culture but that also tapped into the design expertise of formally trained students. This series of workshops led them to collaboratively create new designs that were sold in government craft outlets, and served to increase their design confidence. The third phase in year 3 involved the craftspeople developing their own project. They decided to increase local awareness of their craft by taking a road-show to local schools and engaging school children in making glass products with them. They were very proud to demonstrate and teach their skills and to raise awareness of the local craft. This paper discusses communications in year 3.

Fig. 1. Napkin Rings designed together by craftspeople and designers

3 Method

The researcher acted as facilitator, along with 4 young design students from local institutions and 2 professional designers in the participatory project. All have a similar cultural background with the rural community, enabling them to understand sensitive issues within the culture. Additionally, the first author has 10 years experience of working with rural craftspeople, which is advantageous for approaching participants.

However, a similar cultural background is not enough to build trust and a close relationship for further collaborative projects. Strong interpersonal skills gained from experience and maturity are more influential in fostering the process. For example, when the conversation with craftspeople unintentionally moves to a sensitive issue within the community, the facilitator must give an appropriate response to maintain a neutral position as well as to keep a positive outlook.

Facilitators were also aware of social hierarchy as a hallmark of Javanese Society [1]. Therefore, the facilitator identified and approached the community leaders, whose recommendation usually influenced the community decision.

3.1 The Community Meeting to Decide a Project

Once the researcher gained the trust of the community, she organized a community meeting on a weekly basis for two months in 2012. Each meeting was attended by six to twenty craftspeople. Most of participants had graduated from high school.

The first meeting began with icebreaking in order to create an informal and joyful situation, using a game that involved pairing Indonesian celebrities, and a discussion of punishments for participants who made unconstructive comments, yielding laughter and a relaxed atmosphere. A modified 'string game'[1] was employed to make sure that everyone had a turn to speak and share their opinion. These initial activities were crucial to get craftspeople comfortable in participating. Then, through the intensive discussion in the community meetings, the craftspeople decided to organize events to promote the industry to local buyers through a glass-bead-making workshop. Local high schools surrounding the industry were selected as targets for pilot projects.

The glass bead making workshops for high school students aimed to provide insight into glass beads as a part of Indonesian history and culture, as well as allowing the students to experience making glass beads with assistance from craftspeople. Each workshop lasted 3 hours.

By the end of the events, craftspeople gained a positive response from schools. In total, there were 160 participants in the workshops. Two schools asked for further workshops and enthusiastically allocated the project as an alternate extra-curricular activity for students. Meanwhile, some students expressed their desire to implement their design and meet craftspeople in their place. Overall, the participants showed that they respected and were interested in the local products.

---

[1] Each participant chooses one of provided strings, then he must talk as they slowly wind the piece of string around their index finger. The person must keep talking until reaching the end of the string. (source: `http://www.icebreakers.ws/small-group/string-game.html`).

There are social and economic benefits for craftspeople through these workshops. The social benefits were indicated by the new network built with local high schools while the economic benefit was gained by the product sales and payment as a tutor for the workshop. The glass-bead-making workshop has a strong potency to enhance the craft industry because; firstly it spread the information about the industry to local buyers, secondly it enables future networking between the craft industry and schools. It also developed confidence in the craftspeople to engage externally in this way.

Fig. 2. The glass-bead-making workshop: high school students learnt to make a glass bead, assisted by craftspeople

4 The Choice of Communication Mode

The researcher used both oral and written communication modes during the project. Oral communication was delivered by face-to-face contacts and phone calls while the written communication was mainly by texting (Short Message Service / SMS). Researchers also used printed material (a letter) and email, but did so rarely. Communication was not only about the project, but also to express appreciation and greetings related to special community occasions such as the coming of Ramadan.

Table 1. shows that the oral mode of communication was more commonly used than written mode. Collective face-to-face interaction was the oral mode most frequently used during the project, followed by individual face-to-face, then by phone. Phone calls were used to contact community leaders; first to introduce the researcher prior to meeting, second to make an appointment, and also when certain issues needed to be discussed with certain people. Meanwhile, SMS was the only frequent mode used for written communication. Most SMSs were sent collectively, i.e from the facilitator to multiple recipients. However, it was one way rather than reciprocal communication. Mail and email were the least used modes. The researcher had tried to

spread information through this mode, but there was insufficient response, so she used oral mode and SMS instead.

5 The Use of Language in Each Communication Mode

The researcher combined Indonesian and Javanese language for communication during the project. The bilingual use of language is common in current Javanese society. The Javanese language provides a sense of closeness and facilitates building relationships with participants. The local language is commonly used in oral communication, but is rarely used in written communication. On the other hand, Indonesian as the language of formal instruction in school is used for reading and writing purposes, as well as in formal situations. Thus, the researcher used the Javanese for oral mode, but Indonesian language for written mode. In collective meetings, both Javanese and Indonesian were used.

Table 1. The Use of Communication Modes in The Participatory Project

	Communication Mode					
	Oral			Written		
Purpose	Face to face Individually	Face to Face Collectively	Phone Call	Texting (SMS)	Mail/Flyer	Email
Approach						
Introduce the researcher	V	V	V*	-	-	-
Initial discussion	V	V	-	-	-	-
Organize The Project						
Make an appointment	*Only if needed*	V	V*	V**	-	V*
Make an invitation	*Only if needed*	V	-	V**	V	-
Share problems/ideas	V	V	*Only if needed*	V**	-	-
Make a resume	-	V	-	V**	-	-
News Update	V	V	*Only if needed*	V**	-	-
Arrange work distribution	V	V	*Only if needed*	*Only if needed*	-	-
Miscellaneous						
Appreciation	V	V	-	V**	-	-
Greetings	V	V	-	V**	-	-

V* : only to community leaders
V** : sent collectively
Only if needed : means an additional communication effort

Nevertheless, the use of local language must be done carefully. Javanese speech consists of some levels that reflect politeness and honorific expression [6]. Inappropriate use is impolite and reflects cultural insensitivity. Accordingly, the facilitator used a high level of Javanese speech (*Kromo*) when speaking to an older or respected person; and medium level (*Ngoko*) to a person of similar age and social status.

The other important issues of communicating with craftspeople orally are the intonation and the type of sentence. The facilitator kept speaking in an informal intonation to maintain friendliness, expressing ideas as questions to solicit feedback. This gave craftspeople a chance to interact, engage their responses and avoid passivity. Otherwise the facilitator would miss key information about their ideas or feelings.

The use of Indonesian language in written mode was to ensure that the message was understandable. Nevertheless, it created a formal impression which means distance. Therefore, the facilitator must maintain friendliness feeling and politeness when sending SMS, by using a structure of: "greeting (thanks to God) + main content + thank you". Craftspeople did not necessarily reply to messages. Replies usually comprised only one or two words, such as "Yes mam". Nevertheless, no response did not mean they were not interested in the issue.

6 Discussion

The most effective mode of communication during the project was face-to-face interaction and SMS. Face-to-face meeting conveyed tacit knowledge by the style of speech, dress, or even means of transportation used; in addition to the verbal message. Tacit knowledge builds trust and rapport which are essential to effective networking practice [12]. In addition, collective face-to-face interaction enabled immediate response from community leaders and other members. Rural communities have a strong cohesive culture in which their decision will be greatly influenced by a respected person or persons. Collective face-to-face meeting accelerated effective engagement.

The low usage of email or mail was apparently caused by the limited written communication skill of craftspeople. Email was also less popular as the infrastructure is insufficient to support a reliable communication, creating more obstacles to use than SMS or phone.

7 Conclusion

Communication choices to engage participation in this context considered several aspects: the communal culture, the limited written skill and the influence of tacit knowledge. The influence of community leaders, as is characteristic in rural communal culture, affected the communication effectiveness as well as engaging participation. The limited written skill of participants meant that oral modes were used more than written mode, especially when immediate responses were needed. The use of language and mode of communication must be carefully selected, as it taps into tacit knowledge which is essential in building trust and rapport while communicating with rural people.

Acknowledgements. I would like to first appreciate the work of Jombang glass bead craftspeople during the project. This work is supported by a scholarship from the DIKTI on behalf of the Ministry of National Education, Republic of Indonesia.

References

1. Beard, V.A.: Individual determinants of participation in community development in Indonesia. J. Environment and Planning C: Government & Policy 23(1), 21–39 (2005)
2. Bebbington, A., Dharmawan, L., Fahmi, E., Guggenheim, S.: Village politics, culture and community-driven development: insights from Indonesia. J. Progress in Development Studies 4(3), 187–205 (2004)
3. Cecchini, S., Scott, C.: Can information and communications technology applications contribute to poverty reduction? Lessons from rural India. J. Information Technology for Development 10(2), 73–84 (2003)
4. Dearden, A., Rizvi, H.: Participatory IT Design and Participatory Development: A comparative review. In: Participatory Design Conference 2008, pp. 81–91. Trustees of Indiana University, Indiana (2008)
5. Galloway, L., Mochrie, R.: The use of ICT in rural firms: a policy-orientated literature review. J. Policy, Regulation and Strategy for Telecommunication, Information and Media 7(3), 33–46 (2005)
6. Galloway, L.: Can broadband access rescue the rural economy? J. Small Business and Enterprise Development 14(4), 641–653 (2007)
7. Hollingshead, A.B.: Information suppression and status persistence in group decision making the effects of communication media. J. Human Communication Research 23(2), 193–219 (2006)
8. Irvine, J.T.: Status and style in language. Annual Review of Anthropology, 557–581 (1985)
9. Kshetri, N.: Barriers to e-commerce and competitive business models in developing countries:A case study. J. Electronic Commerce Research and Applications 6(4), 443–452 (2008)
10. Maltz, E.: Is all communication created equal?: An investigation into the effects of communication mode on perceived information quality. J. Product Innovation Management 17(2), 110–127 (2000)
11. Park, P.: What is Participatory research? A Theoretical and Methodological Perspective. In: Park, P., Brydon-Miller, M., Hall, B., Jackson (eds.) Voices of Change: Participatory Research in the United States and Canada, pp. 1–19. JF Bergin & Garvey, Westport (1993)
12. Swan, J., Newell, S., Scarbrough, H., Hislop, D.: Knowledge management and innovation: networks and networking. Journal of Knowledge Management 3(4), 262–275 (1999)
13. Winschiers-Theophilus, H., Chivuno-Kuria, S., Kapuire, G.K., Bidwell, N.J., Blake, E.: Being participated: a community approach. In: 11th Biennial Participatory Design Conference, pp. 1–10. ACM, New York (2010)

Content Prototyping – An Approach for Engaging Non-technical Users in Participatory Design

Maletsabisa Molapo and Gary Marsden

ICT4D Research Centre, University of Cape Town, Rondebosch 7707,
Cape Town, South Africa
{mmolapo,gaz}@cs.uct.ac.za

Abstract. Many in the developing world have little to no experience with computers - they have never used software as part of their daily lives and jobs, so there is always a challenge for how this class of users can be engaged in Participatory Design in a manner that the value of their participation is not limited by their computing experience. This paper looks at previous work that addressed this challenge, and introduces an approach called content prototyping, which is an adaptation of existing practices to fit the needs of non-technical users. We also discuss the lessons learned from using this approach, and give recommendations for related projects in the developing world.

Keywords: HCI4D, Prototyping, Low-Literacy.

1 Introduction

One of the goals in designing new technology for use in developing countries is to design such that the technology integrates into existing cultural structures and community ecosystems, and this can be achieved by seeking guidance from people belonging to the particular cultural groups [8], and involving them in the design process through participatory design [15]. However, for people to be in a position to make such a contribution, they need to fully understand what the new technology is capable of, and be able to visualize how it may integrate into their daily lives.

These questions arise, therefore: how can we co-design new technologies with users who have little to no technology experience? What methods can be used to conduct participatory design in such a manner that users' limited technology exposure does not become a hindrance to their ability to contribute to the design process?

This paper explores answers to these questions through lessons drawn from previous work in the field of HCI for Development (HCI4D), and introduces an approach termed 'Content Prototyping', wherein we recommend that designers seek to develop prototypes that fit their users' current realm of understanding and experience, instead of typical software-based prototypes which inexperienced users may have difficulty conceptualizing. The core of our proposed method of increasing user participation is asking the question: *what representation of the design concept can inexperienced users relate to the best?* In our case, the best representation of the design concept was the output (content) that software would produce, so we prototyped the output for our users, not the software, and designed back from output to output-producing software.

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 788–795, 2013.

2 User Centered Design

User Centred Design (UCD) is a broad methodology based on focusing on the user from the beginning to the end of the design process, ensuring that the needs, wants, and limitations of users are given extensive attention throughout the design process [1]. One form of UCD that has gained acceptance over the years is Participatory Design (PD) [1,12], which permits joint design between the designer and the user.

The success of PD is based upon the assumption that users have experience with digital technology [9], and can appreciate what the technology can do for them. This is hardly the case for most developing world users [3,5]. Because of their limited exposure to technology, such users would not be able to contribute to the design process as they would have limited understanding of how the technology can integrate into their daily lives and jobs, much as they would not have enough computing experience against which to judge what is good or bad technology [10].

In classical PD, prototyping is used to elicit user input on design ideas, where users are presented with prototypes of differing fidelity, and their feedback is used to inform design and motivate refinement of design ideas [6]. Normally, users would be started off with low-fidelity prototypes such as paper prototypes (typically paper-based simulation of user interface elements [8]). However, previous research in the developing world has revealed that users with low computer proficiency levels have difficulty interacting with low-fidelity prototypes because: it's difficult for them to conceptualize prototypes and abstract design concepts, e.g., associating paper sketches with software [9], so they mostly misinterpret and misunderstand design abstractions [11]. This means that PD techniques must be refined to be appropriate to the (computer) literacy and experience of prospective users, so as to encourage their interest in the process and increase the value of their participation.

3 Related Work

Different approaches that have been used to encourage participation of non-experienced users in design are discussed below, which are the works based on whose guidance we developed the idea of content prototyping.

3.1 Simple Technology Artifacts with Instant Utility

According to Ramachandran et al. [14], one way of getting users with little exposure to technology involved in the design process is by introducing simple technology artifacts whose capability is immediately obvious, and presenting these to the users at an early stage in the design process. This approach helps stimulate dialog between the users and the designers within the users' context, and gives a platform for users to easily contribute their local knowledge and expertise to the design process in a manner that they wouldn't if a typical low or high fidelity prototype were used [10] . So the introduction of simple technology artifacts with immediately obvious capability in early stages of design works better than the introduction of low fidelity prototypes at the same stage.

3.2 Scenarios of Use

When users are presented with usage scenarios of the future system within the context of their current work or daily life, it becomes possible for them to envisage the use of the technology in their existing structures, and hence they are able to participate in the design process [13].

3.3 Progressive Design: Increasing Participation through Experience

Maunder et al. [10] and Kam et al. [7] recommend progressively improving the user's technology experience to get them ready to participate in the design process. The designers would engage with the users in their natural work environment, developing the users to a point where they are comfortable with basic technology, while also building supportive structures within their environment. The authors indicate that this approach (termed Progressive Design [4]) "would ensure the progression and development of the users' knowledge base and skill set, thereby enabling the user to better understand the technology, the benefits it offers and how to utilize it effectively….the result is an empowered, confident, motivated user that is able to actively participate in every phase of the design process," [10].

4 Context and Stakeholders

In developing countries, the shortage of health facilities and qualified health professionals is supplemented by employing Community Health Workers (CHWs). CHWs (who are textually illiterate) are trained by public health professionals who are based in rural health centers. Our goal was to assist this training process by designing a content creation model wherein the trainers would create non-textual digital content for the CHWs. We worked together with health centers in Lesotho and Sierra Leone. To understand the CHWs' training context we conducted interviews, user observation, and contextual inquiry. These were followed by persona definition (of trainers and CHWs), task analysis and the design of the local content creation model. In the content creation model, there would be a computer application developed, which would be used by trainers to create non-textual content for CHWs (using images and recorded voice), and the content would be shared to CHWs via Bluetooth when they visited the health center for their monthly training sessions. Our study of the user space revealed that most trainers have low computer proficiency skills, mostly acquainted with basic office applications and web browsers, and all the CHWs had never used a computer before, but all of them owned mobile phones.

5 Methodology

The understanding of our users' skill set led us to rethink the classical prototyping approach we had initially planned to use, which would involve designing a technology (software) that implements the content creation model, starting with low fidelity prototypes, and then going back to the trainers and CHWs with the low fidelity prototypes for them to give us feedback on the design. However, at this stage, we were unsure whether the trainers understood what the introduction of a new technology would mean for them, and how they could integrate it into their daily work. We needed to communicate the possibility of integrating a technology into the training process in a manner that they would understand and relate to [10]. Additionally, we had already established that CHWs are major role players in the flow of health information from the health trainers, via themselves, and on to the communities they serve. Therefore, we also decided that it would be important to involve them in the design process, to give them a say on the content that they would not only consume, but also distribute. Input from the CHWs would be especially valuable from a local cultural perspective. A low fidelity prototype of a computer application (even a fully developed software prototype) would not make sense to a village woman (a CHW) who had never used a computer before, and was never going to interact with the software, only the content produced.

5.1 The 'Content Prototype' Approach

We decided to postpone designing an application and introduced what we term a "content prototype" to mimic the concept of "a simple technology artifact with instant utility" [14], to develop the trainers' and CHWs' mentality to the possibility of using technology in training [4,7,10], as well as to present them with usage scenarios for digital content in their existing training process [13].

To achieve this, we would present sample content to the users, the kind that would be produced in the content creation model we had designed, and use this content as a platform to start the conversation around the idea of digital training content and the process of creating it. We envisaged that both health trainers and CHWs would relate better to digital version of the content they already knew, than a paper prototype of an application whose use they may not clearly understand.

With sample content presented first, we believed that introducing software later on would make sense to them (the trainers especially) as "a tool that creates the useful content we saw earlier". Moreover, based on the work of Ramachandran et al. [14], the expectation from this early stage prototyping using the "simple technology artifact with instant utility," the content prototype in our case, is that we would be able to attract the users' interests in the technology (in this case being the digital content produced for consumption on mobile phones), expose local attitudes towards the technology, elicit design ideas for subsequent stages in the design process, stimulate dialog between the users and the designers within the users' context, and to give a platform for users to easily contribute their local knowledge and expertise to the design process.

5.2 How Does Content Prototyping Compare to other PD Approaches?

Content Prototyping is based on recommendations from other designers who have used PD in developing world projects, but centers on the question:

what representation of the design concept can users relate to the best? In this case digital samples of existing content would be the best representation of the idea of digitizing available content into multimedia formats.

5.3 Creating the Content Prototype

We revisited the content used in training (image books, flash cards, posters) and translated some of it into sample digital content (mock-up multimedia content), resembling the kind that the trainers would produce according to the content creation model we were proposing. We extracted some of the images on the posters and image books and used them to create sample content in the form of "mobile videos". Most posters and image books are made of images accompanied by a line of text that describes the concept represented, as in Figure 1(left). Per concept, we placed an image on a separate PowerPoint slide, then recorded the descriptive line in voice-over in the local language; then saved the overall presentation as a PowerPoint show. This meant that when the trainer opened the PowerPoint show, they would see, in full screen per slide, an image showing with voice-over playing. On the slides, we framed the images with a mobile phone in a person's hand to demonstrate that the videos (series of images with voice over) would play on mobile phones.

Fig. 1. Left: A page from an image book. Right: Three PowerPoint slides, showing a mock-up video made from the image book. Descriptive voice was recorded over each slide.

5.4 Introducing the Content Prototype

When the content prototypes had been created, we introduced them to the trainers and CHWs. We first held a meeting with the health trainers, where we made an introduction and then started playing the samples that were created. The day after meeting the trainers, we held a focus group meeting with 20 CHWs. We did not make the introduction of the content in this meeting, but the chief nurse at the health center did, explaining to them what the content meant (which showed that she had understood it clearly from the meeting we held the previous day). She explained the concept in the simplest terms, and got the CHWs excited even before seeing the content. After the briefing and the playback of the content, we got into a discussion facilitated by one of the junior nurses at the health centre.

5.5 Feedback from the Content Prototype

The results of our ‘early-stage prototyping’ by the use of our content prototypes are compliant with those reported by Ramachandran et al. [14]. The sample content helped to ground our interactions with the users (both trainers and CHWs), and started a conversation about the possible use of mobile digital content, how it would be used, CHWs’ familiarity with mobile technology, etc. Seeing the mock-up multimedia version of their already existing content gave the health trainers an idea of what digital content could do for them. The mock-up content enabled them to ask more questions and express their concerns. Beyond this, we, the researchers, gained more clarity and insight from their comments for the next stages in the design process.

Feedback from the Trainers: The first opportunity spotted by the chief trainer from Lesotho was that through mobile digital content, CHWs would be able to retain information more. She recalled that on several occasions, they would give instructions to the CHWs on what to do for patients in the villages, and the CHWs would get the procedures wrong due to forgetfulness. Beyond training, she also saw the potential of the mobile digital content helping them give elaborate instructions to CHWs. While on the subject of getting procedures right, she suggested that it would be useful if the content produced would include moving pictures, i.e. videos clips. She indicated that sometimes they would wish to demonstrate a procedure to the CHWs, e.g., how to inject a patient. Apart from seeing the potential borne in the use of multimedia content, she also expressed an interest in being able to create or modify the digital content. She emphasized that for their CHWs, it would be best if the voice recordings were in the local language spoken by the CHWs. We informed her that we would provide software that allows them (the trainers) to create such digital content on their own, at which her primary concern was how easy the software would be to use.

Feedback from Community Health Workers: When asked for their opinions on the introduction of digital content, the CHWs’ main comment was that the content would be useful only when the voice is recorded in the local language (Sesotho in Lesotho). They indicated that if the content is in Sesotho, they could use it to counsel their patients. CHWs also saw the opportunity to have medical information with them at all times, seeing that the content “in their pocket”(meaning their phones), could make it easy to refer to the content in cases of emergency.

Evidence of A Two Layered User Base: The trainers saw the potential to disseminate information and instructions to the CHWs more effectively, while the CHWs saw the potential to do their jobs in the community more effectively, and the platform to share content in their communities. This revealed to us that our two sets of users have, to a certain extent, different goals and perceptions, and that our design should embrace these differences. The content prototype enabled this revelation.

6 Discussion

The centre of content prototyping as a method is identifying an understandable artifact, which users can relate to, and use it to guide participatory design exercises. In our case this was sample digital content. The trainers of CHWs do understand the content communicated to their trainees more than they do software, so we chose to use samples of digital content to elicit their needs, interests and concerns. Seeing the content prototypes, the trainers were able to visualize how digital content could assist their existing processes, and even expressed interest in creating such content themselves, also expressing needs that we had not initially designed into the content creation model (e.g., the need to include moving video clips in the content.

This manner of content prototyping also helped engage the CHWs in the early stages of the design process; an opportunity they would not have had if our first prototype were a software prototype, or a low- or high-fidelity prototype of a computer application. The CHWs were able to contribute to early discussions and played a role in influencing the decisions made in the design. Later on in the project, the CHWs' feeling of involvement in the project also encouraged their adoption, appropriation, and ownership of the digital content, as also observed by other researchers, e.g., [2] .

7 Conclusion

Maunder et al. [10] discussed the challenges of using techniques like paper prototyping with people who have limited technology experience, and along with Ramachandran et al. [14], recommend the use of simple technology artifacts with instant utility, introduced early in the design process to expose users to the technology and to elicit requirements and contextual issues from the users' interaction with the technology artifact. Other researchers recommend depicting technology usage scenarios to develop ideas around the use of the technology in everyday life, while other recommendations involve progressively preparing the user for participation in the design process by exposing them to technology bit by bit.

We adopted all these recommendations in our design, but instead of introducing a technology, we introduced "content prototypes," which were a representation of the output that a computer application would produce. This was identified as a representation of the design idea that our users would relate to the best. We learned from this that our two layers of users (content creators - the trainers, and CHWs - content consumers/distributors) were able to participate in the design process as they could relate to the content prototype.

We make a further recommendation therefore, alongside those made by other researchers whose work guided this approach, that where a technology being designed will produce a certain product, it is beneficial to deploy content (or output) prototypes and design the way back from output to output-producing software.

References

1. Abras, C., Maloney-Krichmar, D., Preece, J.: User-Centered Design. In: Berkshire Encyclopedia of Human-Computer Interaction, pp. 753–767 (2004)
2. Anokwa, Y., Smyth, T., Ramachandran, D., et al.: Stories from the Field: Reflec-tions on HCI4D Experience. Information Technologies & International Development 5(4), 101–115 (2009)
3. Blake, E., Tucker, W.: Socially Aware Software Engineering for the Developing World. In: IST-Africa 2006 (2006)
4. Chin, G., Rosson, M.B.: Progressive design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1998, pp. 611–618. ACM Press (1998)
5. Gitau, S., Diga, K., Bidwell, N., Marsden, G.: Beyond being a Proxy User: A look at NGOs ‘ Potential Role in ICT4D Deployment. In: Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development, ICTD 2010 (2012)
6. Jones, M., Marsden, G.: Mobile interaction design. John Wiley and Sons (2006)
7. Kam, M., Ramachandran, D., Raghavan, A., Chiu, J., Sahni, U., Canny, J.: Practical considerations for participatory design with rural school children in un-derdeveloped regions. In: Proceeding of the 2006 Conference on Interaction Design and Children, IDC 2006, p. 25. ACM Press (2006)
8. Maguire, M.: Methods to support human-centred design. International Journal of Human-Computer Studies 55(4), 587–634 (2001)
9. Marsden, G., Maunder, A., Parker, M.: People are people, but technology is not technology. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 366(1881), 3795–3804 (2008)
10. Maunder, A., Marsden, G., Gruijters, D., Blake, E.: Designing interactive sys-tems for the developing world - reflections on user-centred design. In: 2007 International Conference on Information and Communication Technologies and Development, pp. 1–8. IEEE (2007)
11. Medhi, I., Sagar, A., Toyama, K.: Text-Free User Interfaces for Illiterate and Semiliterate Users. Information Technologies and International Development 4(1), 37–50 (2007)
12. Nielsen, J., Bødker, M.: Collaborating with users. In: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design: Open 24/7, OZCHI 2009, p. 325. ACM Press (2009)
13. Nielsen, J.: Paper versus computer implementations as mock up scenarios for heuristic evaluation. In: Human-Computer Interaction (INTERACT 1990), p. 10 (1991)
14. Ramachandran, D., Kam, M., Chiu, J., Canny, J., Frankel, J.F.: Social dynamics of early stage co-design in developing regions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, p. 1087. ACM Press (2007)
15. Winschiers, H.: The Challenges of Participatory Design in a Intercultural Context: Designing for Usability in Namibia. In: Proceedings of the Participatory Design Conference, PDC 2006, pp. 73–76 (2006)

Designing a Platform for Participatory Urbanism: Transforming Dialogue into Action in Underserved Communities

Leonardo Giusti, Amelia Schladow, Amar Boghani, Steve Pomeroy, Nicholas Wallen, and Federico Casalegno

MIT Mobile Experience Lab, 20s Ames Street, Cambridge 02139, MA
{lgiusti,schladow,amarkb,spomeroy,nwallen,casalegno}@mit.edu

Abstract. Participatory urbanism platforms must balance stakeholder needs to both empower citizens and exact change from the local authority. While many platforms can trigger discussion, changes will only be achieved through successful collaborative efforts. This paper outlines the challenges and opportunities of designing for participatory urbanism, drawing on a case study completed with UNICEF and underserved communities in Rio de Janeiro, Brazil. Our design approach helped to generate physical changes in the community infrastructure, and the beginnings of behavioral changes for community residents.

Keywords: Participatory urbanism, civic media, location-based platforms.

1 Introduction

In this paper we investigate how participatory platforms, mediated by mobile and web-based technologies, provoke new methods and possibilities for individual citizens to become actively involved in their neighborhood, city, and urban landscape. In particular, we address how these platforms can support participatory urbanism in underserved and low-income communities, discussing challenges and opportunities.

The discussion presents a case study: UNICEF - Youth Mapping in Rio de Janeiro [1]. The project aimed to empower youth in the favelas of Rio de Janeiro with digital mapping technology for reporting environmental and structural hazards. A specialized version of the Open Locast [2, 3, 4] web and mobile platform (UNICEF GIS) was designed to fit the specific constraints of the local context. With the help of UNICEF and the involvement of local authorities, we organized a series of workshops to train 111 youths to use the technology. More than 300 reports were collected and the local government has planned and implemented several interventions to solve the problems reported by these local residents.

The presentation of the case study and its results helps to articulate a discussion about the design of participatory platforms and the challenges related to their adoption, sustainability, and impact in developing communities. Specifically, we discuss

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 796–803, 2013.

our contribution to one of the challenges identified by Ho in "Human-Computer Interaction for Development: The Past, Present, and Future" [5]: supporting an ecosystem around affordable computing. In this paper, we advocate a design approach that integrates the development of resilient and affordable technological systems with a social and cultural strategy, which includes the early involvement of the local authorities and community in activities. In particular, we discuss the importance of thoughtfully balancing the different voices involved in the discussion to both empower local residents, and activate the local municipality in taking a course of action.

This approach helped to successfully implement a media platform for participatory urbanism. In addition to concrete results documented by improvements to the local environment, user interviews showed how the project sparked a cultural shift, encouraging self-reflection and a sense of commitment toward the quality of the urban environment.

2 Related Works

A number of recent projects [6, 7, 8] show the ways in which mobile media is valuable for enhancing information and knowledge exchange in developing communities. According to Ho et al. [5], a key challenge of HCI for developing countries is creating an ecosystem around affordable computing. He stressed the importance of replicable, low-cost approaches and hardware that can be appropriated and adopted by community-based organizations with minimal requirements for external support.

However, designing for underserved communities requires significantly more than just technological considerations. As outlined by Dearden et al [6], the form of new technologies represents physical capital for the designer to work with, but attention must be paid to human capital, social capital, financial capital, and the transforming structures and processes (i.e. organizational situations in the locality where the technology is intended to be used). Furthermore, they mention the importance of planning the way external agents will interact with local people and communities, since the goal of intervention is to work with locals to envision, create, and adopt sustainable systems that ultimately empower the locality.

This overarching framework provides a starting point for socio-technological considerations in the design of civic platforms for developing communities. Civic media platforms can enable community engagement and introduce multiple points of view into urban discussion. However, a number of challenges emerge when they are applied to participatory urbanism in addition to the challenges implicit in participatory practices in general.

Historically, participatory practices in urbanism can be traced to user-centric visions for architecture in the 1960s. Of particular interest is the work of Yona Friedman, an architect who advocated for architecture to remain a framework for further development by the inhabitant [9]. The notion of the open framework is also visible in the work of Cedric Price et al. [10], an architect who argued for engaging inhabitants as co-designers through temporary architecture and the integration of technology. However, in engaging ideas of user-participation in urbanism, we should also consider potential drawbacks. In Nightmare of Participation [11], Markus Miessen introduced his series of essays by challenging the current state of such practices. “Both historically and in terms of political agency, participation is often read through romantic notions of negotiation, inclusion, and democratic decision making. However, it is precisely this often questioned mode of inclusion that does not produce significant results,” as the voices of the needy are frequently drowned out by the concept of the majority. The risk, as Miessen points out is that the idea of participation can be used as a strategy for consensus and political control: “too often it becomes an expedient method of placation rather than a real process of transformation” [11].

The debate about participatory urbanism in architecture highlights fundamental aspects that are often underestimated in the design of media platforms for civic engagement. The effectiveness of a particular platform correlates with its ability to inform and support the transformation of dialogues into actions: as a consequence, any conversation is inherently somewhat political. In order to make effective decisions and direct interventions a mechanism for selecting, filtering, and prioritizing content needs to be instated; furthermore, a local municipality or governing body has to be involved as a stakeholder from the beginning of the process. A civic media platform is a complex socio-technical system where the power relationship between the involved actors needs to be carefully articulated and specific strategies of engagement, negotiation, and conflict resolution should be designed.

3 Youth Mapping Project

3.1 Context

Rio de Janeiro is vulnerable to floods and landslides, natural disasters that are expected to increase with climate change. The city’s favelas are largely situated along mountainsides, and are already prone to both disasters and socio-environmental risks. In response, UNICEF, with the support of the Municipality of Rio, the Municipal Secretariat of Health and Civil Defense, and Centro de Promoção da Saúde (CEDAPS), decided to start a project in which local adolescents are mobilized to monitor, identify and prioritize social and environmental risks in their community.

In particular, the project had two main design objectives: (1) strengthen participatory governance, and empower local residents to directly improve their local communities; (2) help the local government to be better informed about social and structural conditions, and plan future interventions. The local context posed several technological and social challenges: (1) limited technological infrastructure, in particular slow or absent 3G connection; (2) low penetration of smartphones; (3) social resistance to the adoption of services and products developed outside the local community; (4) lack of faith in real change.

3.2 Design Approach

The UNICEF GIS platform is a youth-led civic media platform composed of a mobile and a web application, designed to engage residents in photo-documenting social and environmental risks in their communities (see section 3.3).

However, the design of the platform was not limited to the technological features; it also encompassed a social and cultural strategy. To foster the adoption of the system and the active participation of the local population, we designed a mediation framework that is both technological (to drive specific actions) and political. A series of workshops (see section 3.4) were organized to instruct youths on the use of the technology, to mobilize them toward environmental and social risks, and to involve local authorities and decision makers in the process. UNICEF played a central role in this process; its local connections were critical to identifying the human, social, and financial capital specific to the territory and to organize the participation process.

3.3 Technology

We worked with UNICEF to build upon the Open Locast [2, 3, 4] platform and customize it for digital mapping in Rio de Janeiro (see fig. 1). UNICEF's knowledge of the community and their experience working with youth helped to contextualize the implementation of Open Locast. The Open Locast platform for this project consisted of a mobile application to capture photos and text, and a website to illustrate and discuss the geo-located collection of media. Visualization for both the mobile and web components were simplified and made more approachable for youth. Finally, maps were organized around themes (e.g. erosion, sanitation, power line problems, social spaces, etc.) to improve organization and comprehensibility for the participants. Particular attention was paid to the creation of a resilient technological infrastructure to cope with connectivity limitations. The Open Locast platform saves all generated content so that it can be automatically uploaded when the connection improves. Furthermore, media content could also be created with other devices (e.g. cameras), and then uploaded using the web application. This provided a series of different opportunities in a context where the penetration of devices is low and presents serious limitations.

Open Locast. Open Locast is a location-based media platform created by the Mobile Experience Lab. It is an integrated platform that combines mobile and web tools to help users create individual and collective narratives, share content, and build local

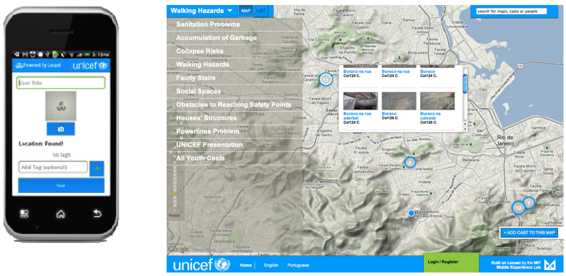


Fig. 1. A screenshot of the mobile application (left) and the website (right)

conversations. Open Locast serves as a tool for horizontal reporting, that is, entirely user-generated content, populated as a centralized space for photographs and videos generated by users, citizens, or specific workshop participants. As it leverages a user-generated tagging system to classify content and add semantic relations between casts, it allows for both user-generated and user-organized content – creating and building an archive solely based on the content present in real-time with location [4]. The Open Locast platform is made up of three components: the Web Interface, the Core (backend and API), and the Mobile application (Android). The Open Locast Core is in Django, which includes GeoDjango to incorporate a geographic web framework enabling powerful storage, querying, and geographic information manipulation. The core features an extensive network-centric API, RESTful and JSON, which allows data stored in Open Locast to be easily accessed and manipulated, allowing a wide- range of visualizations, statistical analyses, and user interfaces. Open Locast Web, which runs through Django, serves as the program's web presence and interface. All map-based interactions and visualizations are handled by OpenLayers – a JavaScript library for displaying map data. It currently functions on Android platforms, and communicates with the Core using its API. The data is stored in an SQLite database based on the Android Content Provider framework.

3.4 Workshops and Activities

The Youth Mapping project involved a group of 111 adolescent (from 11 to 18), community thought-leaders, and local authorities in mapping environmental risks in target communities. The project implementation was carried out in a two-step approach: an initial session to train key stakeholders (see fig.2 on the left) and secondary workshops with local youths for content generation (see fig. 2 on the right).

Fig. 2. Workshop participant training session (left), youths at workshop in Rio de Janeiro (right)

In preparation for a series of workshops, a "trainer of trainers" session was scheduled with key stakeholders: UNICEF, the Public Laboratory for Open Technology and Science, and a selection of Rio community stakeholders, including local NGOs, technical professionals from the Municipal Departments of Health and the Environment, local community members, and adolescent peer mentors. The training session provided space to share knowledge and introduce the technology to the community.

After this initial session, workshops were collaboratively planned. Five 5-day workshops were organized in 5 underserved communities of Rio de Janeiro (Morro dos Prazeres, Morro dos Macacos, Morro do Borel, Morro do Urubu and Rocinha). Each workshop hosted approximately 25 adolescents as well as the aforementioned community leaders, NGOs, and other local authorities. Local experts instructed youth on how to recognize specific potential hazards. For example, mudslides are a major concern in these communities and thus, the youth were instructed to take pictures of certain erosion patterns and dangerous overhangs. The youth were provided phones and guided in using the customized Open Locast platform, and then worked with community organizers and UNICEF to create real-time portraits of their community.

3.5 Assessment

Altogether, the youth generated about 300 casts, which are currently mapped and used as a reference [12]. The following picture (see fig. 3) documents the results of the project, showing specific intervention informed by the mapping activities and conducted by the local authorities.

To assess the workshops, youth and instructors were interviewed according to prepared questionnaires. We interviewed 6 youths (3 males and 3 females) and 2 instructors. Technically, the platform functioned well. At times, a slow 3G connection prevented casts from uploading properly, but the information architecture of the Open Locast platform was designed to cope with this issue.

Fig. 3. Improvements made to communities, as informed by the Youth Mapping project

In addition to creating this resource, the workshops seemed to impact participants' perspective and behaviors; many of the youth were inured to the state of their locale, but capturing specific risks and unpleasant elements helped them learn to expect more from their community. Emboldened, the youth began to see how they could have a voice in local policy. Involvement in the mapping activity helped youths to look at their community in a different way, as exemplified by the following interview responses: "I could see things I had not seen before. Thanks to this activity I realized there is a lot of garbage all around our place", "I got to know places that I didn't know before, and I saw problems that I hadn't noticed before". The youths began to identify a social geography "We realized that the reality of those of us who live in the bottom part of the Morro is much better than that of those who live in the upper parts of it".

More importantly, youths stated an intention to change their habits: "before I was always throwing garbage anywhere, now I've realized that's something bad for the community and I'll try to change it". Yet, as one of the teacher pointed out: "a few students told me they now felt bad about throwing trash on the ground and would try to do it less. At the same time, I still observed them throwing a water bottle on the ground after our conversation." However, the intention to change habits shows an opportunity to build upon and the potential for behavioral shifts.

4 Lessons Learned

Following the Rio de Janeiro experience, the Youth Mapping project was successfully deployed in Haiti in July 2012 [13]. UNICEF adopted the same technological platform (Open Locast) and a similar cultural strategy. In partnership with two local organizations, GHESKIO and the National Office against Violence (ONAVC), UNICEF took on the challenge of identifying places where adolescents and young people are at increased risk of contracting HIV. We are now considering extending the project to 10 cities in Brazil. The interest in continuing the project testifies its effectiveness and impact. In particular, the project was a success in building:

Participation. The Youth Mapping project helped community members to document needed improvements, allowing residents to directly contribute to the amelioration of their neighborhoods;

Governance. This tool for empowerment can provide local governments with constant feedback on problems in underserved areas. The initiative is responsible for real change; as a result of photos and comments from the workshops, repairs have already been made;

Education. Through a hands-on investigation, the youths were encouraged to think critically about how maps and new digital tools can inform community discourse on economic, environmental, and social sustainability.

We strongly believe that the success of this project relies on the design approach we adopted. As previously discussed, we decided to integrate the design of the technological platform with a social and cultural strategy. The role played by UNICEF as a mediator was crucial for the success of the project. UNICEF's long-term presence in the territory, ability to build trust in the communities, and power in activating local resources were fundamental to the adoption of the technology, and to promote a behavioral and cultural change.

The underlying challenge of designing platforms for participatory urbanism is the transformation of dialogue into action, which requires harmony between stakeholder perspectives. Local authorities need to be included in the design and implementation of platforms from the earliest stages. They help to engage the local community and cause real change. And yet, since local authorities are in a position of strong power, their voice in decision-making processes can be louder than the public and they have greater potential to set the final agenda. It is important to mediate the role of local authorities and strategically balance their involvement with that of local residents. This is particularly true in developing communities, where power dynamics and access to resources can be especially skewed. Our approach in assuaging this challenge was an intermediary group of thought-leaders, community activists, and young mentors who mobilize locals and empower them from the ground up. In this way, the public holds local authorities accountable and sets its own goals for the community.

Our contribution to the challenge identified by Ho – supporting an ecosystem around affordable computing – shows how considerations should go beyond technology to encompass social and cultural capital. Specifically, in the context of designing a participatory urbanism platform for underserved communities, we have illustrated the need to include local authorities without compromising the voices of local residents. Our approach of using local leaders as mediators, helped to manage diverse stakeholder needs. As a result of a successful collaboration, we observed the beginnings of a behavioral shift and the potential for sustainable impact in the community.

References

1. http://www.unicef-gis.org/
2. Arslan, P., Brunnberg, L., Casalegno, F., Schladow, Z.: Locast H2flow: creative learning tool for participatory urbanism. In: Proceedings of the Second Conference on Creativity and Innovation in Design, pp. 267–270. ACM (October 2011)
3. Giusti, L., Pollini, A., Brunnberg, L., Casalegno, F.: En Plein Air: A Mobile Learning Approach for Sustainability Education in the Wild. International Journal of Mobile Human Computer Interaction (IJMHCI) 4(2), 44–58 (2012)
4. Chernyakova, I., Villeré, M., Casalegno, F., Giusti, L., Schladow, Z.: Civic Media Platforms and Participatory Urbanism: A Critical Reflection. International Journal of Architectural Computing 10(2), 253–274 (2012)
5. Ho, M.R., Smyth, T.N., Kam, M., Dearden, A.: Human-computer interaction for development: The past, present, and future. Information Technologies & International Development 5(4), 1 (2009)
6. Dearden, A., Rizvi, S.M.H.: A deeply embedded sociotechnical strategy for designing ICT for development. International Journal of Sociotechnology and Knowledge Development (IJSKD) 1(4), 52–70 (2009)
7. Parikh, T.S., Lazowska, E.D.: Designing an architecture for delivering mobile information services to the rural developing world. In: Proceedings of the 15th International Conference on World Wide Web, pp. 791–800. ACM (May 2006)
8. Gitau, S., Marsden, G., Donner, J.: After access: challenges facing mobile-only internet users in the developing world. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 2603–2606. ACM (April 2010)
9. Friedman, Y.: L'architecture mobile: vers une cité conçue par ses habitants, vol. 5. Casterman (1970)
10. Banham, R., Barker, P., Hall, P., Price, C.: Non-plan: an experiment in freedom. New Society 20, 435–443 (1969)
11. Miessen, M.: The nightmare of participation. Sternberg Press (2010)
12. http://rio.unicef-gis.org/
13. http://pap.unicef-gis.org/

Author Index