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

LNCS 6948

Human-Computer Interaction – INTERACT 2011

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

3 Part III

ifip

Springer

Lecture Notes in Computer Science 6948

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

Human-Computer Interaction – INTERACT 2011

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

Springer

Volume Editors

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

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

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

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

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-23764-5 e-ISBN 978-3-642-23765-2
DOI 10.1007/978-3-642-23765-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011935338

CR Subject Classification (1998): H.5.2, H.5.3, H.3-5, I.2.10, D.2, K.3-4, K.8

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

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

Printed on acid-free paper

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

Foreword

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

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

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

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

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

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

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

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

July 2011

Nicholas Graham
Daniel Gonçalves
Joaquim Jorge
Nuno Nunes
Philippe Palanque

IFIP TC13

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

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

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

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

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

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

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

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

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

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

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

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

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

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

IFIP TC13 Members

Australia
Judy Hammond
Australian Computer Society

Austria
Andreas Holzinger
Austrian Computer Society

Belgium
Monique Noirhomme-Fraiture
Federation des Associations Informatiques de Belgique

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

Bulgaria
Kamelia Stefanova
Bulgarian Academy of Sciences

Canada
Heather O'Brian
Canadian Information Processing Society

China
Zhengjie Liu
Chinese Institute of Electronics

Cyprus
Panayiotis Zaphiris
Cyprus Computer Society

Czech Republic
Vaclav Matousek
Czech Society for Cybernetics and Informatics

Denmark
Annelise Mark Pejtersen
Danish Federation for Information Processing

Finland
Kari-Jouko Räihä
Finnish Information Processing Association

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

Germany
Tom Gross
Gesellschaft fur Informatik

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

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

India
Anirudha Joshi
Computer Society of India

Italy
Fabio Paternò
Italian Computer Society

Ireland
Liam J. Bannon
Irish Computer Society

Japan
Masaaki Kurosu
Information Processing Society of Japan

Kenya
Daniel Orwa Ochieng
Computer Society of Kenya

Malaysia
Chui Yin Wong
Malaysian National Computer Confederation

New Zealand
Mark Apperley
New Zeeland Computer Society (NZCS)

Nigeria
Chris Nwannenna
Nigeria Computer Society

Norway
Dag Svanes
Norwegian Computer Society

Poland
Juliusz L. Kulikowski
Poland Academy of Sciences

Portugal
Joaquim A. Jorge
Associação Portuguesa de Informática

Singapore
Henry Been-Lirn Duh
Singapore Computer Society

South Africa
Paula Kotzé
The Computer Society of South Africa

Spain
Julio Abascal
Asociación de Técnicos de Informática (ATI)

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

Switzerland
Ute Klotz
Swiss Association for Research in Information Technology SARIT

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

UK
Andrew Dearden
British Computer Society (BCS)

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

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

Working Group Chairpersons

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

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

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

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

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

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

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

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

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

INTERACT 2011 Technical Committee

Conference Committee

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

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

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

Technical Program Committee

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

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

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

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

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

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

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

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

Short Papers Co-chairs
Daniel Gonçalves, *Portugal*

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

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

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

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

Organizing Committee

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

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

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

Publicity Co-chairs

Paula Alexandra Silva, *Portugal*
Tiago Guerreiro, *Portugal*

Student Volunteers Co-chairs

Tiago Guerreiro, *Portugal*
Xavier Ferre, *Spain*
Effie Law, *UK*

Website Co-chairs

Alfredo Ferreira, *Portugal*

Associate Chairs - Full Papers

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

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

Associate Chairs - Short Papers

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

Reviewers

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

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

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

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

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

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

Sponsors

Gold

Silver

Bronze

Supporters

Organization

Table of Contents – Part III

Long and Short Papers

Novel User Interfaces and Interaction Techniques I

Novel User Interfaces and Interaction Techniques II

Paper 2.0

Recommender Systems

Social Media and Privacy

Social Networks

Sound and Smell

Touch Interfaces

Tabletops I

Tabletops II

Ubiquitous and Context-Aware Computing

UI Modeling I

UI Modelling II

Usability

A Framework to Develop VR Interaction Techniques Based on OpenInterface and AFreeCA

Diego Martínez[1], J-Y. Lionel Lawson[2], José P. Molina[1], Arturo S. García[3], Pascual González[1], Jean Vanderdonckt[2], and Benoit Macq[2]

[1] Laboratory of user Interfaces and Software Engineering,
University of Castilla- La Mancha, Spain
{diegomp1982,jpmolina,pgonzalez}@dsi.uclm.es
[2] Laboratoire de Telecommunications et Teledetections,
Université Catholique du Louvain, Belguique
{jean-yves.lawson,jean.vanderdonckt,benoit.macq}@uclouvain.be
[3] SymbiaIT, Parque Científico y Tecnológico, Albacete, Spain
arturo@symbiaIT.com

Abstract. Implementing appropriate interaction for Virtual Reality (VR) applications is one of the most challenging tasks that a developer has to face. This challenge is due to both technical and theoretical factors. First, from a technical point of view, the developer does not only have to deal with non-standard devices, he has to facilitate their use in a parallel a coordinated way, interweaving the fields of 3D and multimodal interaction. Secondly, from a theoretical point of view, he has to design the interaction almost from scratch, as a standard set of interaction techniques and interactive tasks has not been identified. All these factors are reflected in the absence of appropriate tools to implement VR interaction techniques. In this paper, some existing tools that aim at the development of VR interaction techniques are studied, analysing their strengths and, more specifically, their shortcomings, such as the difficulties to integrate them with any VR platform or their absence of a strong conceptual background. Following that, a framework to implement VR interaction techniques is described that provides the required support for multimodal interaction and, also, uses experience gained from the study of the former tools to avoid previous mistakes. Finally, the usage of the resulting framework is illustrated with the development of the interaction techniques of a sample application.

Keywords: Novel User Interfaces and Interaction Techniques,Tools for Design, Modelling, Evaluation, Multimodal interfaces, Virtual Reality.

1 Introduction

The WIMP (Window-Icon-Mouse-Pointer) metaphor has been broadly accepted and this has provided a standardized set of Interaction Techniques (ITes) and Basic Interactive Tasks (BITas) for these applications. As a result, stable and mature GUIs have been produced, rules to guide interaction have been formalized [1] and

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 1–18, 2011.

integrated development environments (IDEs) are available to design and implement them. On the other hand, the current status of Virtual Reality (VR) interfaces is far less mature. Conceptually, it has been long accepted that a set of universal BITas exists [2] and many possible classifications have been suggested [3], [4], [5], [6]. However, a standardized set of BITas and ITes has not been identified [5] or, rather, agreed.

The importance of defining this set of universal tasks and techniques has been deeply discussed in the 3DUI mailing list [7] (see Figueroa in August, 2004), as they would be the starting point for the development of toolkits to implement 3DUIs. As a result of this situation, the tools available do not provide the required support to deal with the complexity of 3D interaction. In some cases, developers will have to implement their own solutions mostly from scratch, relying on their previous experience and personal criteria. Taking into account the complexity and the multidisciplinary nature of 3D interaction –3DUIs, multimodal interaction, ergonomics, human factors, etc.-, designing interaction can result one of the most challenging tasks that the developer has to face. In some other cases, the support provided forces developers to be bound to a specific VR platform and programming paradigm. This factor restricts developer's freedom, limiting him/her to the BITAs and ITes considered when creating the platform, which may not fit the requirements of the system implemented.

This paper describes a framework to assess VR developers in the implementation of VR ITes. To achieve our goals, a selection of previous proposals, identifying their weaknesses and strengths, is studied in the next section. Based on the weaknesses and strengths identified, a detailed list of goals is provided. A conceptual model to describe ITes has been proposed and, once the conceptual bases have been established, a framework to implement VR ITes has been implemented on top of OpenInterface [10] and AFreeCA [11]. Finally, a case of study describing some ITes developed with this framework is presented.

2 Related Work

Most of the VR platforms that have appeared during the last years have provided some support for user interaction. At a lower level of support, some systems such as, DIVE [18] or MASSIVE [19], provide integrated support for the most popular VR devices. Besides supporting VR devices, MR Toolkit [25] includes the concept of decoupled simulation, that allows device management and VR simulation to be executed in separate computers. This factor is especially useful when resource consuming processes -i.e. visual tracking- are involved.

VRJuggler [26] and VRPN [27] extend this kind of support. Besides providing support for many devices and a decoupled simulation model, they rely on the concept of *abstract input device*. Applications are implemented using a set of abstract devices to gather user input. At run-time, any set of concrete devices that satisfies the requirements of those *abstract input devices* can be used to run the application.

However, in all these approaches, the processing required to adapt the information gathered from the devices to the requirements of the application –smoothing filters, algorithms, etc- is up to the developer. Some platforms, such as InTML [8] or

OpenTracker [28], have been proposed that provide appropriate support for this task. These systems use a graphical notation to describe both the devices and the algorithms that process those data. Using these data-flow graphs includes some benefits, such as improving the comprehensibility of the ITe by depicting it as a diagram or the possibility to reuse the components in other diagrams. These two factors encourage fast prototyping and the interaction of the developers with experts from other domains –i.e. human-computer interaction experts-. However, InTML does not provide an execution platform to generate executable code from the diagrams. OpenTracker, provides an execution platform but is does not provide a graphical editor for the diagrams. Besides, this platform is discontinued nowadays.

There are, however, some platforms, such as NiMMIT [9] and the IFFI [30], inspired in the usage of data-flow models to describe ITes and that do provide an execution platform to create their diagrams and transform them into executable code. Other platforms, such as CHASM [29] or StateStream [31], even extend this concept, mixing data-flow graphs with state-based models in order to provide better support for event-based systems. However, these approaches tend to exceed the limits of the ITes. Many of these platforms consider collision detection, selection or modifications of the appearances of objects -feedback or highlight- in their diagrams. These elements are considered BITas by many authors and, as a result, these diagrams end up covering the whole description of the logic of the application. As a result, the execution platforms of these platforms are not just tools to describe ITes, but they are closed development systems to implement VR applications.

In contrast to these approaches, the intention of the tool proposed in this paper is to provide a framework to implement VR ITes that includes the features discussed in this section, but that can also be integrated in any custom VR development, letting the developer choose the most appropriate platform to describe the rest of the aspects of the system. The specific list of goals that guide the definition of the framework are:

- The framework must identify the borderline between the ITe and the BITas. The theoretical basis of 3DUIs will be reviewed, and the required models proposed, in order to identify the elements that must be addressed.
- The framework must support the relevant features identified during the study of previous proposals, such as the management of devices, a decoupled simulation model and abstract input devices.
- The framework must support the easy and fast prototyping of VR ITes - provide graphical notations when possible, reuse previous components, provide support to implement these components, etc.-.
- The framework must be easy to integrate in any VR development.

3 Conceptual Model

The framework proposed in this paper is intended to provide support to implement ITes, independently of other issues of the VR system, such as the contents or the way the logic of the application is described. However, given the lack of maturity of the 3D field, identifying the functionality that it should and should not cover was one of its main concerns. With this purpose in mind, the theoretical bases of 3DUIs have been studied, proposing a model of VR ITe that can be used to identify the functionality that must be addressed.

The model proposed is inspired in [13] and reuses its definitions and vocabulary to describe the process required for an ITe in the context of VR applications. This model has been chosen as it successfully summarizes relevant previous models [14] and it provides a common vocabulary that removes some existing ambiguities in the field.

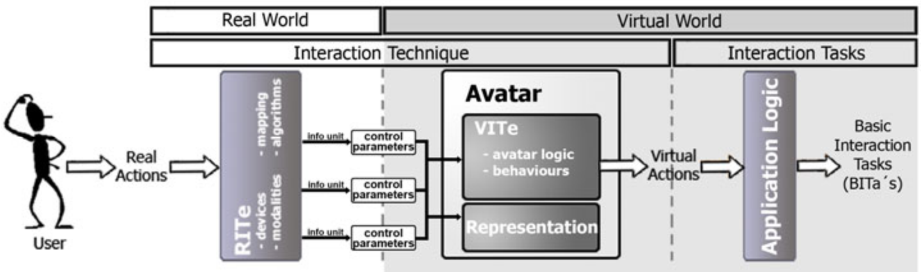


Fig. 1. Elements in the model and their role in the process of an ITe

In [13], the author describes two main functionalities for ITes: 1) to translate user's actions, produced in the real world, into the space of virtual actions, and 2) to generate the information units required by the BITas to be executed.

The first definition includes the idea that ITes cover all the process since the user performs an action in the *real world*, this real action is measured and the *avatar* transforms it into a virtual action that is processed by the *virtual world*. The three main elements in this definition will receive special attention in this model: the *real world*, the *virtual world* and the *avatar*.

The second definition assumes that the ITe has to generate the information units used by an existing set of BITas. This definition implicitly describes two stages: a first stage in which the required information units are generated -see *RITe* bellow- and a second stage that encapsulates and sends these information units to the BITas -see *VITe* bellow-.

The elements identified from the first definition –the *real world*, the *virtual world* and the *avatar*-, will be the key elements to divide the ITe in the two stages demanded from the study of the second definition. As a result, six main concepts -see Fig. 1- build this model of ITe: the three elements in the first definition -the *real world*, the *virtual world* and the *avatar*-, the two stages identified -*RITe* and *VITe*- and the borderline between these stages –*control parameters*-.

3.1 Virtual World

VR applications create a synthetic world that users can explore and interact with (the Virtual Environment (VE)). The virtual 3D space and the domain of the tasks to perform in the system characterize this *virtual world*.

3.2 Real World

The goal of a VR application is to create the *virtual world* described above. However, there are many elements outside the scope of this *virtual world* that participate in the task of letting a user experience it. The combination of all these elements outside the *virtual world* -the real 3D space, the user, the input devices, the algorithms that process the data, the presentation devices, etc.- is defined as the *real world*.

3.3 Avatar

The *avatar* is usually defined as the user embodiment in a VE [6]. Apart from representing the user in the VE, the *avatar* is the entity that allows her to interact with the objects of the VE. Thus, it can be seen as an intermediary existing in the *virtual world* but controlled by the user from the *real world*.

Under this perspective, the *avatar* provides three main functionalities. First, the *avatar* generates the *virtual actions* that allow the user to perform her tasks. Secondly, the *avatar* provides a *control interface*, so that the user can exploit its capabilities in the VE.

Finally, the *avatar* acts as the representation of the user in the VE. This representation must help the user -or other users in the case of a multi-user VE- to understand its capabilities -what it can do and how it can do it-. As it is defined in [16], this representation must be consequent to the metaphor it represents, its affordances and its constraints.

The study of the functionality of the *avatar* is relevant for the goals of the work presented, as it identifies the extents of the ITe. ITes describe how the *avatar* generates its *virtual actions* or how it is controlled. However, ITes should not affect other virtual objects –i.e. 3D widgets-. Interaction with these objects would be the result of executing BITas using the information units contained in the *virtual actions* created by the *avatar*.

3.4 Control Parameters

A *control parameter* is a piece of information that is somehow relevant to control the *avatar*. They can be seen as the cords we pull to move a puppet [15] and they separate two spaces, the real world -where the *RITe* generates updated values for the *control parameters*-, and the *virtual world* -where the *avatar* processes these updated values-.

Two main functionalities are considered. First, *control parameters* can contain information to affect the *avatar's* representation –describe positions of parts of the avatar (head, hands, etc.), its status (standing, crouching, sitting, etc.)-. Secondly, *control parameters* can contain information to make the *avatar* produce a given *virtual action* -see VITe in 3.6-.

The set of *control parameters* of an *avatar* defines its *control interface*. Any user will be able to control the *avatar* by producing the appropriate values of this set of *control parameters* -see *RITe* in 3.5-. As a result, it can be seen as an extension of the concept of *abstract input device*, describing how the *avatar* is controlled. However, this description is not strictly bound to the devices used, but can include domain-specific concepts, such as status (i.e. run, crouch), actions (i.e. SELECT, POINT, CUT), etc.

3.5 Real-World Interaction Technique (RITe)

As it was described above, during the first stage of an ITe, the information units that the ITe will transmit are gathered. This part of the process, since the user performs an action in the *real world* until it is transformed into one or more of the information units required by the *avatar* -see 3.4- is known as the *Real-world Interaction Technique* (*RITe*). It receives this name because it deals with all the elements of the *real world* that affect the ITe -the user, the input devices or any other real objects and the spatial relationships among them (i.e. the relative position of a user to a screen, which real world object the user is holding), etc.-. Real world values are measured -positions, temperatures, key-presses in a device, etc.- and processed to generate meaningful information for the *avatar* -values for a certain *control parameter*-.

Many kind of tasks can take part during this process: algorithms -digital signal processing, filtering or algorithms to transform real world coordinates into virtual world coordinates-, processes typical from multimodal interaction, such as multimodal fusion -combining different input modalities into the required information units [17]-, etc.

3.6 Virtual-World Interaction Technique (VITe)

The *Virtual-world Interaction Technique* (*VITe*) is the last stage of the ITe and covers all the process that occurs since the *control parameters* of the *avatar* are updated until it generates one or more *virtual actions* -to grab an object, or to point and select it-. It receives this name because it happens in the context of the *virtual world* -coordinates refer to the *virtual world*, messages are related to the domain of the tasks to perform in the application, etc.-. The *real world* is completely unknown during this stage.

One issue that has a big impact in the definition of this stage is that, just like it happens with ITes or BITas, there is neither universal format for *virtual actions* nor a standardized way to transmit them. These factors will vary according to the VR platform used, but three main paradigms can be identified [32]:

- **Data-flow graphs:** Some systems, such as VRML [20] or Avocado [32], describe a logical graph that traverses the 3D scene, interconnecting objects and allowing them to exchange events –virtual actions-.
- **Callback subscription:** Some systems, such as DIVE [18] or SecondLife, describe a set of high-level events –virtual actions- and provide a process that detects when any of these events occur. Objects can subscribe call-backs to each of these events, which will be invoked every time the event occurs.
- **Spatial model of interaction:** Some systems, such as AFreeCA[11] or MASSIVE [19], use the space as the layout that rules object communication. Objects define volumes in the 3D space where they can receive information -Focus [19] or receivers [11]- and volumes where the virtual actions can be received -Nimbus [19] or messages [11]-.

The nature (and format) of the *virtual actions* used is different, according to the paradigm chosen. Given that the *VITe* has to generate these *virtual actions*, as long as a unified format for *virtual actions* is not proposed, this stage of the ITe will be dependent on the VR platform chosen and, thus, *VITes* will not be reusable between VR platforms.

4 Prototype Framework

The framework implemented will be explained in this section. As it was discussed in section 3.6., as long as a unified format for virtual actions is not available, it will not be possible to create a fully reusable ITe. It will be necessary to adapt the last stage of the ITe, according to the specific VR platform used.

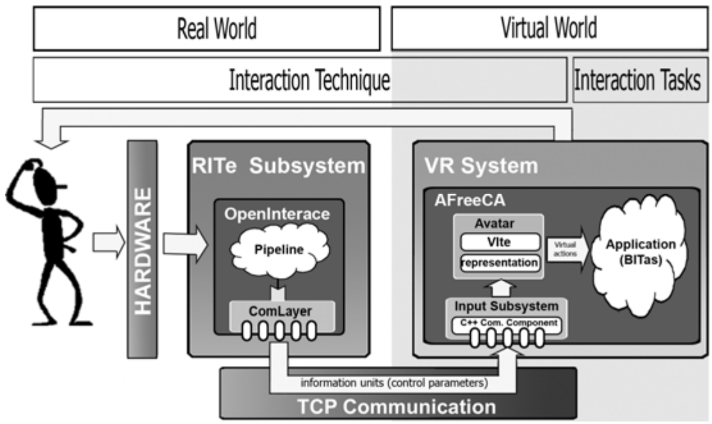


Fig. 2. Structure of the framework and its relation with the conceptual model proposed

In order to limit the impact of this restriction, and according to the model of ITe proposed, the framework has been divided into three main components –see Fig. 2-. The first two components *–RITeSubsystem* and *CommunicationLayer*- are in charge of satisfying most of the requirements and goals of the framework and can be reused in any VR development. The third component *–VRSystem*- deals with the *VITe* stage and, thus, will need to be modified, according to the VR platform chosen by the development team. In order to accomplish this sample implementation of the framework, AFreeCA, a CVE development platform created at the LoUISE group [32], has been chosen.

The implementation details of each of these components will be detailed in the following subsections.

4.1 RITe Subsystem

This component is in charge of transforming the user real actions into the information units -values of the *control parameters*- required by the *avatar* (see section 3.3) and it has been implemented using the OpenInterface Platform [10].

OpenInterface aims at providing an extensible open source design and development workbench for supporting the rapid development of multimodal interactive systems, and it provides support for many of the features desired.

First, is uses a graphical notation that describes programs in terms of components and connections between components. Components can encapsulate anything, from a device to a processing algorithm, and they describe their communication interface with other components using input and output ports. Connections can be used to connect components' ports, describing the flow of data and messages (data-flow graph) among the components of the program.

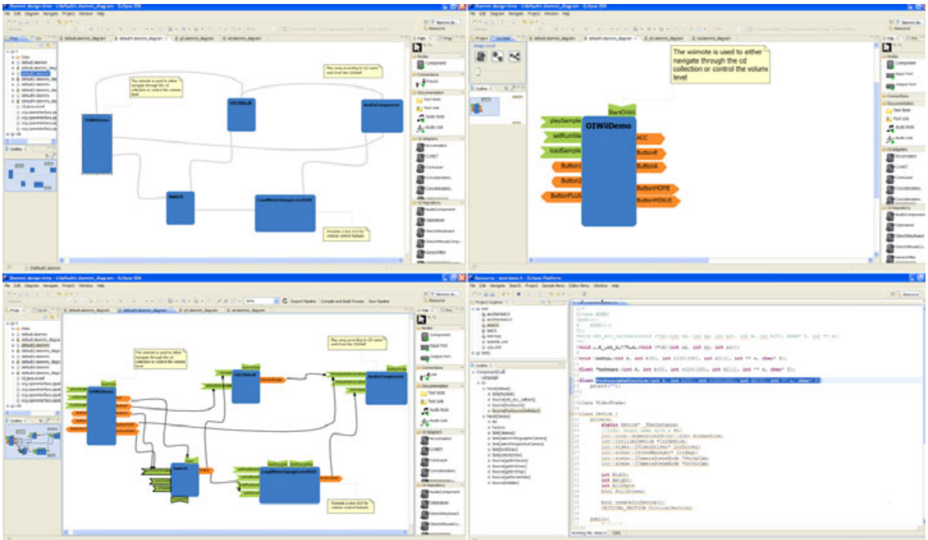


Fig. 3. OpenInterface includes a visual editor called SKEMMI that can be used to design both the diagrams and the components that participate in an OpenInterface application. Diagrams can be visualized at several levels of detail and executable code can be authomatically generated from them.

Secondly, OpenInterface provides an IDE for both components and programs called SKEMMI. This IDE is integrated as a plug-in in Eclipse and it supports all the development life-cycle of an OpenInterface application –see Fig. 3-. Components can be graphically designed, and the skeleton of its source code –basically its input and output methods- generated both in C++ or Java. Developers can then implement the body of the components using Eclipse and start using them for their OpenInterface programs –called pipelines-. Pipelines can also be designed graphically, compiled and run by the OpenInterface kernel with a very low overhead.

Besides providing integrated support for the developers, OpenInterface diagrams can be visualized at diferent levels of abstraction, which facilitates interaction with experts from other fields of knowledge or final users, which can be useful given the multidisciplinary nature of these interfaces.

Thirdly, OpenInterface provides an on-line growing repository of reusable components to implement multimodal systems. These components range from components to control input devices, algorithms to filter data, and even low-level algorithms to support multimodal fusion. This repository can be extended by any developer willing to share its components with the rest of the members of the research community.

All these features make OpenInterface an almost perfect choice to implement the *RITe Subsystem*, as it can be used as an integrated tool to implement *RITes*, including graphical notations, reusability and fast prototyping capabilities. There is, however, one important issue that needed to be addressed. OpenInterface assumes that its execution kernel will lead execution, and all the elements in the program will be OpenInterface components. This philosophy was conflicting according to out goal to allow the framework to be integrated with any external VR system. Instead of forcing VR developers to encapsulate their systems as OpenInterface components, it was decided to include an OpenInterface component that allowed programs to send the values of the *control parameters* calculated –the result of the *RITe*- to an external VR application. This design allowed interoperability but, also, provided support for a decoupled simulation model, that can be important as the complexity and processing requirements of the *RITes* increases.

4.2 Communication Layer

This component is in charge of getting the updates of the *control parameters* generated by the *RITe Subsystem* and deliver them to the *VR System*, where they are encapsulated into *virtual actions* and used to execute BITas.
The data types that these components can send and receive are described in Table 1. This set of data types has been selected so that it can satisfy the requirements of any BITa –it can send the information units that any BITa can possibly require-. In order to identify the data types that should be supported, the most accepted classifications of BITas have been studied [5], [3], [6], identifying the types of data they can require –see Table 1-.

Two pieces of software have been implemented to cover this functionality, an OpenInterface component –*ComLayer*- to send information via TCP/IP to an external program and a generic C++ component to receive this information. Thus, developers simply hava to integrate this C++ component in their VR systems to receive the updates of the *control parameters* sent by the OpenInterface pipeline.

Table 1. Data types supported and BITas that could support them, acording to several authors

Name	C-style type	BITas that require this data type per author
Vector3	`double[3]`	wayfinding, exploration, selecting, positioning, rotating [5]
		navigation [3]
		movement and navigation, object manipulation and interaction [6]
Matrix44	`double[4][4]`	wayfinding, exploration, selecting, positioning, rotating [5]
		navigation [3]
		movement and navigation, object manipulation and interaction [6]
Discrete	`int`	system control [5]
		data input [3]
		object manipulation and interaction, conversation with agents [6]
Continuous	`float`	system control [5]
		data input [3]
		object manipulation and interaction, conversation with agents [6]
DataStream	`char*`	system control [5]
		data input [3]
		object manipulation and interaction, conversation with agents [6]

CommunicationLayer uses the OpenSoundControl formatting to transmit these updates, given its high performance. Even this decision can be technically correct, most VR platforms do not support this protocol. A second implementation of this component is currently being tackled, this time using the VRPN protocol for communication. This will allow straightforward integration of the framework with many VR systems.

4.3 VR System

This component encapsulates all the details about the VE –the BITas, functions and behaviours of the objects-. Regarding the ITes, it encapsulates an application dependent implementation of the *avatar* and its associated *VITes* -see *avatar* in Fig. 2-. AFreeCA is the VR platform chosen for this part of the framework. AFreeCA aims at providing an open source platform to implement immersive and highly interactive Collaborative Virtual Environments (CVEs). The main factor that differentiates this platform from other current VR platforms is its interaction model [11, 12]. This model uses space, time and the collaborative structure of the CVE to rule the communication and interaction among the virtual objects. This allows the platform to provide native support for some interesting features, such as immersive ITes - virtual-hand, go-go, ray-casting, etc.-, collaborative manipulation, subjective perception of the VE - depending on users' locations, their collaborative contexts, etc.-. Additionally, it provides many of the technical features required in a CVE application nowadays - stereo visualization, 3D audio, haptycs, avatars with inverse kinematics, device abstraction, etc.-.

In order to integrate the VR system with the *RITe* Subsystem, an *InputSubsystem* -module used in AFreeCA to gather information from input devices- was implemented using the C++ communication component described in 4.2.

At this point, the possibility to improve the framework, providing a toolkit of reusable implementations of the most common *VITes* -virtual hand metaphors, ray-casting techniques, world in miniature, etc. [16]- was considered. Actually, this was the approach taken in some previous proposals. However, the fact that this toolkit would have been platform dependent, together with the fact that a set of standard ITes does not exist discouraged us from doing so. Instead, reusing the native implementations of these techniques that AFreeCA provides was found a more sensible approach.

5 Example Usage of the Framework

The following subsections illustrate the usage of the framework proposed to a sample VR system. The application reproduced a building game that allowed users to build different figures by joining LEGO-like pieces. The development followed the user centred process described in [21], interleaving four evaluations -two heuristic evaluations, a user centred evaluation and a summative comparative evaluation- with five implementation iterations.

Some initial efforts were necessary during the user task analysis -see 5.1- to identify the BITas, the virtual actions and the control parameters required. However, once this effort was done, the development of the system was facilitated.

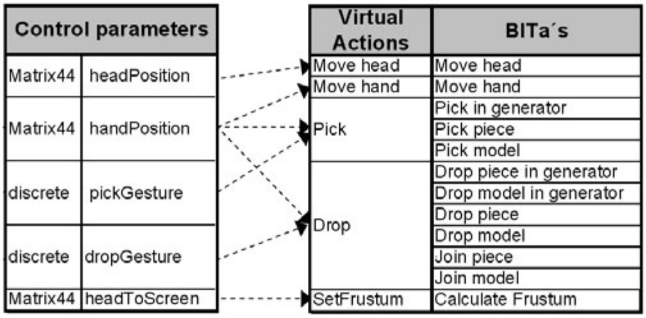


Fig. 4. BITas identified (right), virtual actions that trigger them (centre) and relationships with the control parameters required to generate the actions (left)

First, once the *control parameters* required to control the *avatar* had been fixed, the development of the *VR System* –contents, tasks, avatar, etc- could be independent of the implementation of the *RITes*.

This independence together with the graphical editor and fast prototyping capabilities of the *RITe Subsystem* allowed interaction developers to easily test new techniques or refine the existing ones. This was especially useful after each of the four usability evaluations. After each evaluation, modifications on the ITes, or even new ITes, were proposed and prototyped according to the data and observations gathered from the evaluations.

This resulted in the implementations of up to eleven different prototypes, each of them using different devices and techniques. Different presentation devices–monitors, head-mounted-devices, rear-projection screens-, input devices –mouse, P5 data glove, Wiimote and Fakespace pinch gloves- and techniques –composite and increased manoeuvring, composite positioning [24]- were compared during this process. Some prototypes even evaluated the usage of voice as an input modality, but its performance was very low, abd this approach was discarded. The four most promising ITes were evaluated in a summative comparative evaluation. The detailed description of these ITes, and the relevant conclusions gathered from this study can be found in [24]. However, the current paper will only focus in one of these prototypes, in order to provide enough details about its implementation using the framework presented.

5.1 The Virtual World

The VE implemented reproduced a building game that allowed users to build different figures by joining LEGO-like pieces (see Fig. 5, left). Users could create pieces in the piece generator and join them together to create figures. These figures could be dropped, destroyed by dropping them in a piece generator or joined to other figures. The user task analysis performed during the first stage of the methodology

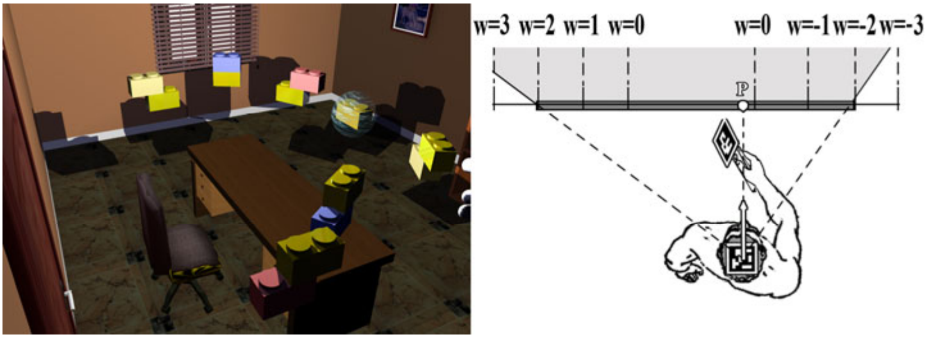


Fig. 5. Screenshot of the Virtual Environment created and sketch of one of the ITes implemented

highlighted that positioning -control of the position of the virtual hand [22]- and manoeuvring -precise control of the viewpoint [22]- would be the most challenging tasks that the user would need to perform in the VE. Besides, this study was very relevant for the design of the ITes in the system. This information was used to identify the BITas that had to be performed in the system -see the third column in Fig. 4-, the virtual actions required to trigger them -second column and the information units these actions would encapsulate –first column-.

5.2 The Real World

Several real world factors were considered that affected interaction:

- Space availability: An empty room of 8x3 meters was available to deploy the prototype. This allowed the usage of rear projection screens and visual tracking.
- Devices available: Several devices were available: A P5 data glove, a wiimote, mice and keyboards as input devices; Head-mounted devices, flat monitors and rear-projection screens as presentation devices; Finally, even though a commercial tracking system was not available, a visual tracking system was implemented using fire-wire cameras.
- Users: The profile of the users was studied and their abilities taken into account. The way this profile affected the ITes is described into more detail in [24].
- Ergonomics: The position and movements that the users would have to perform to use the ITes was considered, to avoid discomfort or tiredness.

The relationships among the previous elements were also studied, which was relevant for the later design of the *RITes*. As an example, during the implementation of the visual tracking system, the field of view of the cameras (devices) and the space availability determined where the cameras had to be located (attached to the roof of the room). This affected the size and location of the ARToolkit markers on the devices used. Their final location was decided taking into account user's comfort and ergonomic criteria.

5.3 Virtual-World Interaction Techniques

In order to choose the *VITes* to use in the system, the selection tool described in [23] was used. With this purpose, the BITas required in the system and the particular features of the VE were studied. The manipulation requirements of the tasks, the limited size of the VE and the required ability of the avatar to navigate while having its hands free encouraged using a virtual hand [16] *VITe* for manipulation. The ability to have a close and detailed view of the model while constructing it encouraged the usage of first person navigation. Instead of having to implement these *VITes*, the native techniques shipped with AFreeCA were used.

5.4 Avatar

Once the *VITes* to use were chosen, the time to design the perceptual representation of the avatar came. Two possibilities were considered: a human-like avatar -they are natively supported by AFreeCA- and an invisible avatar in which the only visible part was the hand. The main drawback of the human-like avatar was that it did not help users to understand the capabilities and constraints of their avatar. This avatar presented many body parts that could not really be controlled by the user -except for the reduced control inverse kinematics permitted-. Another negative point was the fact that its body sometimes occluded other objects of the environment. In contrast to the inconveniences of this human-like avatar, the invisible avatar provided a simple and functional design, representing its exact capabilities. It avoided the user to get wrong ideas or expectations and, also, minimized occlusions. As a result, - and according to the ideas described in 3.3. - the invisible avatar was chosen.

5.5 Control Parameters

As it explained in section 3.4, control parameters describe the information units required for the BITas of the system and the ones required to move the representation chosen for the avatar. Given the invisible avatar chosen, no additional control parameters are required to move the avatar representation itself. As a result, the only control parameters required are the ones obtained in the study of the BITas of the system -see the third column in table 3-.

5.6 Real-World Interaction Techniques

This section describes one of the prototypes implemented. The prototype used a rear projection screen of 1.65x1.20 square meters, a wiimote and two ARToolkit markers, one on the head and another one on the top of the wiimote –see Fig. 5, right-. The rear projection screen was used as a window to the virtual world, displaying the part of the VE that would be visible according to the position of the user's real head, the real position and size of the screen, and the location of this window in the VE -see the grey area in Fig. 5, right-. In order to see objects outside this grey area, a way to rotate the window in the VE was required. To achieve this, when the user looked to the side of the screen -see projection ***P*** *i*n Fig. 5, right-, the virtual window started rotating, showing other parts of the VE. The angular velocity depended on the point where the user was looking at -see ***w*** in Fig. 5, right-. To control the movement of the hand, a stretch go-go technique was used. Also, the A and B buttons in the wiimote were used to pick and drop elements of the VE.

These *RITes* were implemented using the OpenInterface pipeline depicted in Fig. 6. The red pins in every component identify components' services that are invoked by the execution kernel during the start-up. The services are usually used to configure the component or to start the execution of the *RITe* thread.

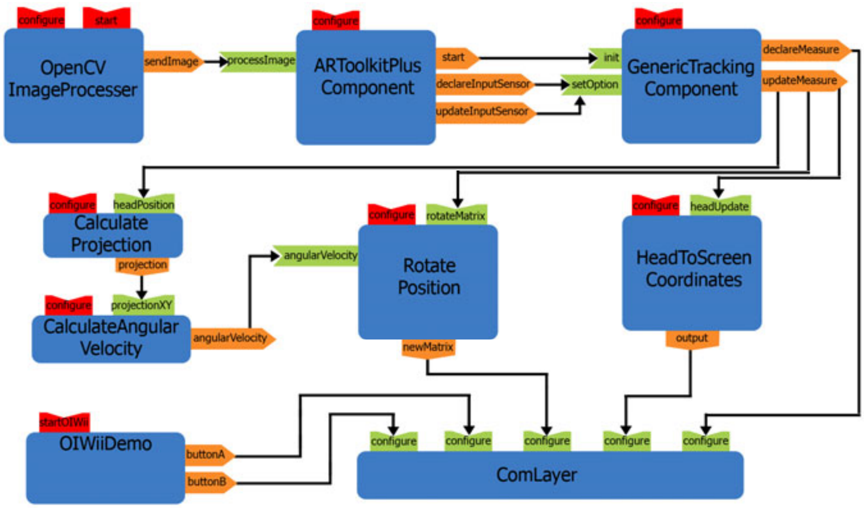


Fig. 6. OpenInterface Pipeline used for prototype number 4

In the first line of the figure, three components collaborate to implement the visual tracking system. The first component (*OpenCVImageProcesser*) is used to get an image from a camera. The second component (*ARToolkitPlusComponent*) encapsulates a modified version of the ARToolkit Plus library, used to detect the 3D position and orientation of the markers in the image captured. Modifications to the library were required to allow the effective detection of several non-related markers in the same image, a feature that is not correctly supported The third component (*GenericTrackingComponent*) is used to transform the coordinates generated by *ARToolkitPlusComponent* from the system of reference of the camera to a useful system of reference for the application, that is, to *virtual world* coordinates. It also encapsulates some noise reduction filters, implemented to improve the quality of the tracking. The component produces a pair containing the name of a *control parameter* -i.e. headPos, handPos,etc.- and a transformation matrix as a result, that are sent to other components through the *updateMeasure* port.

The second line of components implements the *composite manoeuvring* technique [24] that controls the position and orientation of the virtual head. *CalculateProjection* uses head position updates –and ignores hand updates- to calculate the 2D point of the screen where the user is looking at –see point **P** in Fig. 5.right-. This value is used by *CalculateAngularVelocity* to calculate the angular velocity of the virtual head according to the point the user is looking at and the size and location of the screen.

This value is used by *RotatePosition* to affect both the position of the head and the hand. *RotatePosition* contains a transformation matrix that accumulates the rotations performed by the user using the *low-precision manoeuvring technique* [24]. This matrix is updated according to the current angular velocity and the pass of time and it is applied to any transformation that refers to the *virtual world*. This second set of components illustrate an important concept about *RITe: RITes* do not simply deal with the devices and how the information gathered from them is filtered or adapted to the requirements of the application, *RITes* can also encapsulate algorithms or rules to describe complex and rich ITes.

Finally, *HeadToScreenCoordinates* calculates the position of the user's head from the system of reference of the screen. This is used by the *avatar* to calculate the projection volume (OpenGL frustum) displayed in the rear projection screen. *OIWiiDemo* is used to retrieve users key presses of the wiimote. All the information unitsgenerated by this pipeline are finally sent to the *VR System* using *ComLayer*.

6 Conclusions and Future Work

The work presented in this paper describes a framework to assess VR developers in the design and implementation of ITes.

In order to identify the functionality that the framework had to address but, also, to guide the later design and implementation of the ITes, a novel model of ITe has been proposed. This model is the result of the study of previous models of the elements of 3DUIs, but it goes deeper in the analysis of the elements that participate in the ITes -*real* and *virtual world* and the *avatar*- and the stages required for an ITe, dividing it into two sub stages -the *RITe* and the *VITe*-. This model had an important impact during the development of the final framework. First, the stages identified in the model –*RITe* and *VITe*- influenced the components of the framework –a *RITe Subsystem*, a *VR System* and a *Communication Layer* to interoperate them-. Secondly, the study performed during the definition of the model revealed that, given the lack of both standard ITes or a standard format for virtual actions, any *VITe* would be dependent on the VR platform used. As a result, it is currently impossible to provide a toolkit with a closed set of ready-to-use ITes that can be reused by any VR platform. Figueroa stated in [7] that the definition of standard ITes and BITas would be the key element to make the implementation of these toolkits possible. Even though this statement remains true, it would be the definition of a standard format for *virtual actions* what would allow different platforms to reuse complete ITes. Once reusing ITes was possible, the standardization of the most used ITes and BITas would be simply a matter of time.

Besides the ITe model, the second, and major contribution of this work, is the ITe development framework presented. This framework shares the strengths identified in previous proposals, but it avoids some of their deficiencies. These features are summarized in Table 2.

The framework -implemented on top of OpenInterface Workbench and AFreeCA- facilitates the implementation of the ITes. On the one hand, the usage of OpenInterface facilitates the design and fast prototyping for the *RITe* stage, covering most of the features the ITe development platforms must contain –graphical notations,

Table 2. Summary of the features achieved by previous existing platforms and features included in the framework proposed

	VR Devices	Decoupled simulation	Abstract input device	Graphical notation	Reuse	Visual editor	Execution platform	Integration
DIVE	yes	no	no	no	no	no	no	no
MASSIVE	yes	no	no	no	no	no	no	no
MRToolkit	yes	yes	no	no	no	no	no	no
VRJuggler	yes	yes	yes	no	no	no	no	no
VRPN	yes	yes	yes	no	no	no	no	no
InTML	yes	no	no	yes	yes	yes	no	no
OpenTracker	yes	no	no	yes	yes	no	yes	no
NiMMIT	yes	no	no	yes	yes	yes	yes	no
IFFI	yes	no	no	yes	yes	yes	yes	no
CHASM	yes	no	no	yes	yes	yes	yes	no
StateStream	yes	no	no	yes	yes	yes	yes	no
OpenInterface +AFreeCA	yes	yes	yes	yes	yes	yes	yes	yes

IDE, modularity, component reusability and fast prototyping capabilities-. These features, together with the on line repository of ready-to-use components can save many developing efforts. On the other hand, the usage of AFreeCA facilitated the design of the VITe stage, given that it provides a set of the most usual *VITes*. Finally, both elements can interoperate using a decoupled simulation model.

Thirdly, the prototypes developed revealed that the contributions presented in the paper can have a good impact on the development process. First, the model supports the design of the ITes, helping in the identification of relevant factors. Secondly, the framework supports their development. Besides, the architecture of the framework allows the *RITe Subsystem* and the *VR System* to be implemented and tested with a high independence of each other, once that the *control parameters* that they will exchange have been determined.

Several lines of research can continue the work presented here. First, the usage of the proposed framework in the development of other VR systems is considered as relevant. This way, its ability to adapt to the requirements of different applications could be checked. Also, as a side effect, more OpenInterface components would be generated during these developments, components that could be reused in the future, saving development time and efforts.

Secondly, the definition of a universal format for virtual actions would be definitively interesting. As it has already been discussed, this would be a key element to produce reusable toolkits of VR ITes. A detailed study of the different formats of virtual action that the most common VR platforms use and the way they are processed by the VE would be the first step. Then, it would maybe be possible to suggest a format of virtual action that could be reused in all these platforms.

Acknowledgements. The current work has been partially supported by the Junta de Comunidades de Castilla-La Mancha (PEII09-0054-9581) and the Ministerio de Ciencia e Innovación (TIN2008-06596-C02-01).

References

1. Pittarello, F., Celentano, A.: Interaction locus: a multimodal approach for the structuring of virtual spaces. In: Proc. of HCITaly Symposium, Florence, Italy (September 2001)
2. Herndon, K.P., van Dam, A., Gleicher, M.: The Challenges of 3D Interaction: A CHI 1994 Workshop. ACM SIGCHI Bulletin, vol. 26(4), pp. 36–46 (October 1994)
3. Boyd, D., Sastry, L.: Development of the INQUISITIVE Interaction Toolkit - Concept and Realisation. In: User Centred Design and Implementation of Virtual Environments (UCDIVE) Workshop, York, United Kingdom (September 1999)
4. Barrilleaux, J.: 3D User Interfaces with Java3D. Manning Publications (2001)
5. Bowman, D.A., Kruijff, E., LaViola Jr., J.J., Poupyrev, I.: An Introduction to 3-D User Interface Design. Presence, vol. 10(1), pp. 96–108 (February 2001)
6. Sutcliffe, A.: Multimedia and Virtual Reality: Designing Multisensory User Interfaces. Lawrence Erlbaum Associates, Mahwah (2003)
7. 3DUI community web page, `http://www.3dui.org/`
8. Figueroa, P., Green, M., Hoover, H.J.: InTml: a description language for VR applications. In: Proc. of the Seventh International Conference on 3D Web Technology, Web3D 2002, Tempe, Arizona, USA, February 24 - 28, pp. 53–58. ACM, New York (2002)
9. Boeck, J., Vanacken, D., Raymaekers, C., Coninx, K.: High-Level Modelling of Multimodal Interaction Techniques Using NiMMiT. Journal of Virtual Reality and Broadcasting 4(2) (September 2007)
10. Lawson, J.-Y., Al-Akkad, A., Vanderdonckt, J., Macq, B.: An Open Source Workbench for Prototyping Multimodal Interactions Based on Off-The-Shelf Heterogeneous Components. In: Proc. of the First ACM SIGCHI EICS, July 14-17. ACM Press, USA (2009)
11. Martínez, D., García, A.S., Martínez, J., Molina, J.P., González, P.: A model of interaction for CVEs based on the model of human communication. Journal of Universal Computer Science 14(19), 3071–3084 (2008)
12. Martinez, D., Molina, J.P., Garcia, A.S., Martinez, J., Gonzalez, P.: AFreeCA: Extending the Spatial Model of Interaction. In: International Conference on Cyberworlds, pp. 17–24 (2010)
13. Molina, J.P., García, A.S., Martínez, D., Manjavacas, F.J., Blasco, V., González, P.: An Interaction Model for the TRES-D Framework. In: Proc. of 13th IEEE Mediterranean Electrotechnical Conference (MELECON 2006), Benalmdena, Málaga. special session New interaction paradigms in Virtual Environments (May 16-19, 2006), Electronic Proceedings
14. Foley, J.D., van Dam, A., Feiner, S.K., Hughes, J.F.: Computer Graphics: Principles and Practice, 2nd edn. Addison-Wesley, Reading (1996)
15. Molina, J.P.: Un enfoque estructurado para el desarrollo de interfaces de usuario 3D, doctoral thesis, University of Castilla- La Mancha (February 2008)
16. PoupyrevI, Weghorst, S., Billinghurst, M., Ichikawa, T.: Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. Computer Graphics Forum 17(3), 41–52 (1998)

17. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young, R.: Four Easy Pieces for Assessing the Usability of Multimodal in Interaction the CARE Properties. In: Human Computer Interaction, Interact 1995, pp. 115–120 (1995)
18. Frécon, E., Stenius, M.: DIVE: A Scaleable network architecture for distributed virtual environments. Distributed Systems Engineering Journal 5(3), 91–100 (1998)
19. Greenhalgh, C., Purbrick, J., Snowdon, D.: Inside MASSIVE- 3: flexible support for data consistency and world structuring. In: Proc. of the Third International Conference on CVEs, CVE 2000, San Francisco, California, USA, pp. 119–127. ACM Press, New York (2000)
20. Web3D consortium, `http://www.web3d.org`
21. Gabbard, J.L., Hix, D., Swan, J.E.: User-Centred Design and Evaluation of Virtual Environments. IEEE Computer Graphics and Applications 19(6), 51–59 (1999)
22. Bowman, D.A., Kruijff, E., LaViola, J.J.: Poupyrev E I. 3D User Interfaces: Theory and Practice. Addison-Wesley, Reading (2004)
23. García, A.S., Molina, J.P., González, P.: Exemplar VE design guidance tool for selection and manipulation interaction techniques. In: Proc. of 11th International Conference on HCI, HCI International 2005, Las Vegas, Nevada, USA (July 2005)
24. Marítnez, D., Kieffer, S., Martínez, J., Molina, J.P., Macq, B., González, P.: Usability evaluation of virtual reality interaction techniques for positioning and manoeuvring in reduced, manipulation-oriented environments. The Visual Computer 26(6-8), 619–628
25. Mark, C.S., Green, M., Liang, O., Sun, Y.: Decoupled simulation in virtual reality with the MR Toolkit. ACM Trans. Inf. Syst. 11(3), 287–317 (1993)
26. Bierbaum, A., Just, C., Hartling, P., Meinert, K., Baker, A., Cruz-Neira, C.: VR Juggler: A virtual platform for virtual reality application development. In: SIGGRAPH Asia 2008: ACM SIGGRAPH ASIA 2008 Courses, pp. 1–8. ACM, New York (2008)
27. Russell, M., Hudson, T.C., Seeger, A., Weber, H., Juliano, J., Helser, A.T.: VRPN: A device-independent, network-transparent vr peripheral system. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST 2001), pp. 55–61. ACM, New York (2001)
28. Reitmayr, G., Schmalstieg, D.: OpenTracker-an open software architecture for reconfigurable tracking based on xml. In: VR 2001: Proceedings of the Virtual Reality 2001 Conference (VR 2001), vol. 285, IEEE Computer Society, Washington, DC, USA (2001)
29. Wingrave, C., Bowman, D.: ”CHASM”: Bridging Description and Implementation of 3D Interfaces. In: Proc. of IEEE VR Workshop on New Directions in 3DUIs, pp. 85–88 (2005)
30. Ray, A., Bowman, D.A.: Towards a system for reusable 3D interaction techniques. In: Proc. of VRST 2007, pp. 187–190 (2007)
31. Haan, G., Post, F.H.: StateStream: a developer-centric approach towards unifying interaction models and architecture. In: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2009), pp. 13–22. ACM, New York (2009)
32. Tramberend, H.: Avocado: A Distributed Virtual Reality Framework. In: Proceedings of the IEEE Virtual Reality (VR 1999). IEEE Computer Society, Washington, DC, USA (1999)

Exploring Interaction Strategies in the Context of Sleep

Dzmitry Aliakseyeu[1], Jia Du[1], Elly Zwartkruis-Pelgrim[1], and Sriram Subramanian[2]

[1] Philips Research Europe, 5656 Eindhoven, The Netherlands
[2] Computer Science, University of Bristol, U.K.
{dzmitry.aliskseeyeu,jia.du,
elly.zwartkruis-pelgrim}@philips.com,
sriram@cs.bris.ac.uk

Abstract. This paper highlights opportunities and considerations when developing interaction techniques in the relatively unexplored area of sleep. We do this by first describing the sociological aspects of sleep such as the need for scheduling and coordinating sleep within a collective, followed by a description of the physiological aspects such as the circadian rhythm and sleep stages. We then examine how some external factors like the location of sleep and family settings can affect sleep to highlight potential design opportunities. We finish the paper by describing five distinct themes such as scheduling for a collective, supporting sleep transitions and feigning and inhibiting sleep around which design opportunities are explored. The main contribution of this paper is a discussion on the phenomenon of sleep and its position in the social life to provide a rich understanding of sleep and a set of opportunities for interaction design research around sleep.

Keywords: Sleep, interaction design, sleep quality, bedroom context.

1 Introduction

Sleep is a fundamental and pervasive behavior among members of the animal kingdom. An average human needs about 7 to 8 hours of sleep a night [28]. Sleep affects our physical and mental health and is essential for normal functioning of all processes within our body. For example, sleep has a major restorative function for the human body and plays an essential role in memory processes and learning.

Although there isn't a clear understanding of the fundamental purpose of sleep from a neuroscience or biology perspective, the functional necessity of sleep is demonstrated by showing that failing to do so for even a few days leads to a broad repertoire of pathological outcomes [43]. Lack of sleep can lead to a decrease in metabolic activity of the brain, a weakened immune system, decreased alertness and ability to focus. Sleep reduction is also linked to weight gain [40], increased risks of breast cancer [54] and impaired glucose intolerance [42]. It is therefore essential that humans get the right amount of sleep.

However, a busy lifestyle and work pressures have led to an increase in waking hours at the expense of sleeping hours. In the 2007 US National Sleep Foundation Survey, about 65% of the people stated having some kind of sleep problem and 44%

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 19–36, 2011.

reported sleep problems almost every night resulting in symptoms close to insomnia [36]. The pharmacological market for sleep disorders was valued at $7bn in 2008 and growing.

Sleep has gained increased awareness not only because of its widely recognized functional significance in health but also its influence in people's social life and performance. We may look at sleep as an activity for resting the brain and body, but we can also look at sleep as a special stage, e.g. bedtime ritual, where we interact with the world and people around us in a unique style.

Experiences around sleep are an interesting area that has been dominated by the bedding industry for years. Only recently has one seen the emergence of solutions in the bedroom that focuses on creating new experiences around people's sleep. For example, products for directly improving sleep or sleep transitions include devices such as Ecotones Sound Therapy Machine© and Philips' Wake-up Light©. They improve sleep by reducing the environmental disturbances (like sound) and help people to fall asleep by playing relaxing sounds, creating a dusk lighting ambiance or waking people up by gradually increasing light.

While addressing specific sleep related problems through sensing and interaction technology is important, we argue that sleep should be considered as a holistic part of our lifestyle. Understanding the social, psychological, and physiological aspects of sleep can lead to design opportunities that not only lead to improved quality of sleep but also an improved quality of life.

In this paper we examine the various design opportunities that exist in the context of sleep. We start by reviewing HCI literature that is related to sleep followed by a review the physiological, psychological and sociological aspects of sleep. We finally present various design opportunities that arise from clever integration of the factors with socio-physiological aspects.

The paper has three main contributions:

- Through a review of the literature we describe the phenomenon of sleep and its position in the social life to provide a rich understanding of it.
- Expand the design space of interactive technologies to include sleep by describing factors that need to be considered when designing for sleep.
- Identify five sleep related themes that have high potential and need for interaction design.

2 Sleep and HCI

Research into interactive technologies in the home is growing in popularity. There are numerous research articles on all aspects of the interconnected living room and kitchen yet there is very little research into HCI for the bedroom. This is despite the fact that on average a person spends most of their time in their bedroom when at home; for most of this time they will be asleep. This limited research interest in the past is perhaps a consequence of considering sleep as a state of inactivity. In this section we review sleep monitoring products and fairly limited amount of HCI literature that looks at interactive solutions to support sleep.

2.1 Sleep Monitoring Tools

There are a number of technologies available for monitoring the sleep. Polysomnography (PSG) [34], is often used in clinical settings and allows researchers to view physiological status of a person during sleep e.g. eye movements, chin muscle activity, brain activities and heart rate. It is regarded as the gold standard to measure sleep and wake activities [23]. However, current PSG measurement requires attaching many physiological sensors on the body, which cannot be done unobtrusively [12] and therefore less suited for home use. Actigraphy [7] is commonly used as an unobtrusive alternative to PSG. Actigraphy is a watch like device that is worn on one's arm, and measures movements with an accelerometer and interprets sleep and wake patterns accordingly [19, 46]. It has been shown to be valid in assessing sleep and wake activity but is less accurate for monitoring detailed sleep information like sleep stages and sleep onset [1]. Actigraphy is commonly used together with subjective sleep assessment methods such as a sleep diary [34], especially in longitudinal studies in clinical settings. Many commercial products also incorporated the concept of actigraphy. One such example is the iPhone application EasyWakeup (easywakeup.net) which, when placed underneath the pillow, can detect the movements of a person and it will wake up a person during a relatively lighter sleep phase within a user defined wake up time window.

Zeo [58] is a new sleep monitoring device that is claimed to be able to monitor single PSG by three sensors integrated in a headband. Zeo can provide more accurate information about sleep quality, quantity and sleep staging than an actiwatch. It is also less obtrusive than full PSG measurement.

2.2 Interaction Design around Sleep

Most of the research in HCI and design involving sleep has revolved around the alarm clock. Landry et al. [27] extended a standard alarm clock to include contextual information (traffic, weather, early meeting, etc.) to inform the user of any unusual situations that would require an adjustment in the usual morning routine. Schmidt [48] has developed a networked alarm clock that is connected with the social network of a user. The user can define when to be awoken by including conditions such as 'wake me if persons A & B are already out of the bed', i.e. the wake up time (alarm) can be defined by the status of other members of the network. The reverse alarm clock designed by Ozenc et al. [39] encourages children to stay in bed and avoids disturbing their parents during the night or too early in the morning. Choe et al. [8] carried out interviews and surveys to synthesize concepts for supporting healthy sleep behavior such as Sleep monitoring clock and role playing game to promote healthy sleep.

Kim et al. [24] developed a network enabled alarm clock so people can share their sleeping status within a social group. They found that sharing sleeping patterns allowed participants to feel more connected and intimate [24]. Sleep therefore can become an important part of a co-presence system that contributes to achieving a higher level social connectedness. Other examples of systems that use the bed for connecting remote people to improve intimacy of couples sleeping in different locations are The Bed by Dodge [11], Telematic Dreaming by Sermon [50] and Somnia [47]. Each of these projects connects sleepers in two different locations through sensory immersions such as warming the pillow and transmitting audio and video.

Various creative solutions have been proposed for this unique location and its accompanied user state. However, using existing HCI standards is not sufficient to meet future user needs and expectations. A good understanding of the specific context and user state are needed first in order to come to HCI standards for this specific area. In the following chapters we will present different aspects that can be considered in this respect.

3 Sleep

Humans spend on average about a third of their life sleeping. Although historically sleep has been considered to be a state of inactivity, scientists now believe that sleep is a dynamic process with brain activity.

Sleep is the most important form of periodic remission which requires scheduling, coordination and careful consideration for ones vulnerability [57]. These activities of sleep are of social significance and present rich opportunities for exploring design solutions. Most of the social aspects of sleep are often dictated by the human sleep physiology and brain activity. For example, bed time rituals of transitioning from being awake to asleep and vice-versa have emerged from our intuitive understanding of the physiology of sleep. Important differences in sleeping practices in different countries (monophasic sleep vs. the siesta culture) can be better understood by looking at the human biological clock.

To identify what role interactive technologies can play in supporting sleep and activities around sleep we first need to understand what role sleep plays in our life and what sociological, cognitive and physiological processes define it. In this section we give a very brief overview of relevant work on the physiology, sociology and psychology of sleep.

3.1 Human Sleep from a Physiological Perspective

The sleep and wake cycles are driven by the circadian clock and homeostatic sleep drive.

Circadian Clock. The internal circadian clock regulates rhythmic variations in physiological and behavioral processes, such as core body temperature, food intake and sleep [41]. The rhythm of this circadian clock is depicted in figure 1. The circadian clock potentiates wakefulness (and alertness) at one phase of the diurnal cycle, while facilitating sleep and its attendant processes at the opposite phase [44]. Engaging in behaviors that are not synchronized with the circadian clock, such as travelling through different time zones or shift work, will result in a misalignment between the external environment and the circadian clock.

This internal clock naturally maintains a periodic rhythm of slightly more than 24 hours. External factors, most importantly light, provide daily resets of this internal mechanism to a 24 hour cycle. Studies have shown that bright light administered during the early biological night will delay the timing of the circadian system, whereas bright light administered during the late biological night will advance the timing of the circadian system [22]. Exposure to light suppresses the release of melatonin, a hormone that promotes sleep.

Sleep Onset. Sleep onset refers to the transition from wakefulness to sleep. The period is defined as the time spent after a person goes to bed and the light is turned off to entering stage one sleep. This process is usually observable via brain signal changes [14,16]. Transition from sleep onset to light sleep has been described as the process of gradual reduction of arousal through thalamic deactivation, a disappearance of saccades, a reduction in endogenous blinking and an appearance of slow eye movements. During sleep onset, humans still maintain their reaction to external stimuli, even till stage 1 and 2 sleep, though responses and reaction time decrease slowly as we enter deep sleep [37].

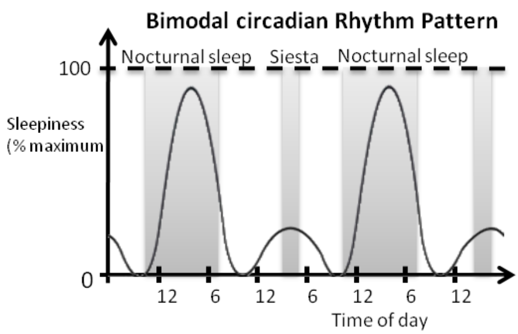


Fig. 1. The biological clock drives a person's sleep and wake pattern throughout a day, being most sleepy in the early morning and second in the afternoon. Based on [31].

Waking Up. The process of awakening is accompanied with reactivation of consciousness and alertness. Opposite to sleep onset, increased core body temperature and heart rate corresponds to the reduced subjective sleepiness are observed [2]. When awakened abruptly from sleep, for example by a sudden noise, one may feel drowsy and find it difficult to remain awake. This is known as sleep inertia, and may last for minutes up to several hours [25] (sleep inertia and accompanied feeling will also depend on factors such as sleep stage people are awakened from, the amount of time they have slept, their circadian phase and so on). So it is essential that a person does not wake up abruptly from sleep but that the wake up process is gradual.

Sleep Drive. Just as a person needs to drink when losing fluid during exercising, a person needs to sleep when she has not slept enough. The drive for sleep and the need for drinking are regulated by the human's internal mechanism known as homeostatic drive [31]. In terms of sleep, homeostatic drive not only takes into account the hours a person has been active during the day but also how many hours were lacking from the person's previous nights. Therefore he/she may feel sleepier when homeostatic sleep drive functions as a result of an insufficient amount of sleep.

Sleep Stages. Human sleep is not a binary switch which is ON or OFF but goes through stages and cycles, alternating between light and deep sleep stages. Scientists have characterized those stages as Rapid Eye Movement (REM) sleep and Non-REM sleep. Normal nocturnal sleep consists of four or five sleep cycles [20] each including several stages of Non-REM sleep and REM sleep. A complete sleep cycle includes progression from stage one to stage four sleep and back to stage one and then entering REM sleep, after which a new cycle starts again, each lasting around 90 minutes [28]. Figure 2 shows the various stages of sleep and how a person transitions through these stages over the course of a 7 hour sleep.

After sleep is initiated, one enters Stage 1 sleep. In this stage, most cognitive and physiological activities slow down, though body movements still occurs once in a while as sleep travels into a deeper stage. A person wakes up easily from Stage 1 sleep but might recall it as only being drowsy. Stage 2 is seen as an intermediate stage between light and deep sleep, lasting about 10 to 25 minutes. Stage 3 and Stage 4 are considered as deep sleep stages. The person becomes less responsive to external stimuli and the auditory threshold of adults increases compared to Stage 2 and REM sleep [10]. Thus it is relatively harder to awaken a person who is in deep sleep. When awakened from this stage one may feel disoriented for a few minutes. On average, an adult spends 10-20 percent of her sleep time in this slow wave sleep.

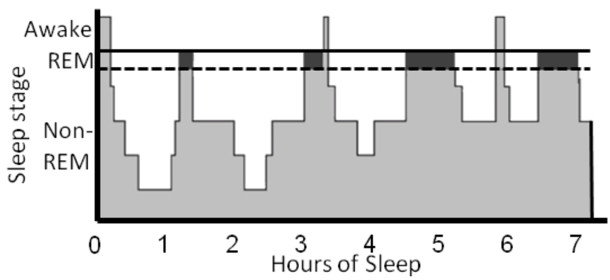


Fig. 2. Sleep architecture: REM sleep and Non-REM sleep and sleep stages. Based on [28].

As a sleep cycle completes and transits to lighter sleep, one enters into REM sleep, also called 'Dream Sleep' [28]. During REM sleep, one's respiration rate and brain activities increase, along with the occurrence of dreams. When awakened in this stage, one is likely to recall his/her dream. But a person is unable to move due to the inability to use his/her voluntary muscles, e.g. postural or skeletal muscles.

3.2 Sociology of Sleep

Since as far back as the 16th century there exists literature on the etiquette of sleep and the sharing of a bed. It was not uncommon [56] for medieval households to share a huge bed among all household members including servants and knights. Often only wealthy nobles with large houses could afford separate bedrooms. Sleep in those times was a public affair that required ordering and coordination.

Although sleeping habits have changed and evolved in the 21st century it is still the case that members of a collective coordinate "time schedules" for sleep. For example, the arrival of a new child compromises the sleeping schedule of a couple to the extent that synchronization of the baby's sleep with the parents' becomes an imperative. The sleep synchronization of the youngster not only allows parents to reduce disturbance during their sleep but it also hastens the child's integration into the family's pattern of activity. This is not only true of a family but of many other groups. Apartment dwellers in urban environments have noise level restrictions at night to reduce sleep disturbance of others.

The scheduling within a collective generally serves two main purposes: first, when sleeping, it reduces external stimuli that can disturb the sleep of an individual, and second, when awake, it allows members of the collectivity to interact with each other.

Another important sociological aspect of sleep is safety. The sleeping society is extremely vulnerable since it not only lacks the sense of location but also control over its social and physical environment. The collective coordination of sleep potentially minimizes this vulnerability to an extent.

Schwartz [49] argues that the transition from sleep to the wake state is a necessity for protecting a person against sudden shifts or impositions of a wake status, since immediately after waking up a person needs time to gain full control of his or her body and mind. The time needed for the transition will vary depending on a person's mental and physical state, age and other factors. Similarly, bed time rituals play an important role in the transitions from being awake to sleep [49, 57]. Familiar cues, such as the end of a TV program, often set into motion the bedtime ritual. These rituals (such as undressing, washing or donning the sleeping costume) are not only a response to sleepiness but they are also triggers to induce sleep [49].

Another aspect of sleep is how the presence of the sleeper affects those that are awake. The effect depends on the context and the relation between those sleeping and other wakeful persons e.g. an insomniac being wide awake at night might feel helpless realizing other people are asleep.

There are also different sleeping cultures, which are: monophasic (sleep consolidated into one block), bi-phasic (two blocks of sleep during the day and night), and polyphasic (one night's sleep and several day's napping). These sleeping cultures are often associated with a particular country or region e.g. bi-phasic is common for southern countries like Spain (i.e. siesta), while polyphasic is most common in Japan (i.e. napping called inemuri – "to be present and sleep"). It is important to understand these differences when designing for sleep. For example based on the physiology of sleep napping has a positive effect on the well-being of a person, however social norms might not accept napping at work.

Other cross-cultural differences regarding sleep also include bedtime rituals, co-sleeping (of infants with parents), the sleeping environment and sleep related problems [38]. For example co-sleeping with young children in Japan (59%) was found to be much higher than in US (15%) [38]. As a result, a product intended for supporting healthy bedtime routines for US families may not work well for Japanese families and vice versa.

Many sociological aspects of sleep are dictated by sleep physiology and how the body regulates its sleep and wake cycles. Understanding the physiology of sleep is also essential for designing supporting technologies.

3.3 Psychological Aspect of Sleep

Many sleep problems are caused by psychological factors, such as worries, a busy mind, stress and anxiety [4,15]. According to the survey from the National Sleep

Foundation [36], 35% of respondents reported that worries and not being able to empty one's mind are the most disturbing psychological factors affecting sleep. Psychological responses, such as stress and anxiety correspond to physiological changes like increased respiration rate, heart rate, decreased skin temperature and increased body temperature [6]. These physiological changes enhance alertness and arousal and are contrary to the decreased physiological state of a person falling asleep. Studies that compare good and poor sleepers provided further evidence that poor sleepers have increased anxiety and alertness at bedtime, resulting in decreased skin temperature [26], increased number of body movements, and increased core body temperature [32].

Therefore, many sleep enhancing products are designed based on meditation principles [51], such as muscle and mind relaxation and breathing techniques to induce relaxation. These products have the potential to reduce sleep onset when used at bedtime, such as the paced breathing devices of emWave© (www.emwaveforsleep.com) and nightWave© (www.nightwave.com).

3.4 Sleep Problems and Coping Strategies

According to the American Academy of Sleep Medicine, there are various types of sleep disorders, ranging from respiratory-related sleep disorders to developmental or neuropsychiatric sleep disorders. Some of these disorders have a medical cause, while others are related to behavioral or cognitive factors. The latter implies that the treatment should focus on the behavioral and / or cognitive basis of these conditions. The term commonly used for these type of treatments is Cognitive Behavioral Therapy (CBT-I). It is a treatment for insomnia, since misperceptions regarding sleep and stressful events often play an important role in the development of insomnia. For example, people might have problems falling asleep and therefore become more and more anxious as the night passes. As a result they might engage in behaviors that are detrimental for sleep, such as working behind the pc in the bedroom. The aim of CBT-I is to break such patterns by providing insight into people's actual sleeping behavior and educating them to associate the bedroom with sleeping. Review studies have shown the effectiveness of such treatments, reflected by a significant reduction in sleeping problems, which exceed those of pharmacological treatments [52] and last longer [45].

3.5 Social and Demographic Factors Influencing Sleep

The physiology as well as sociology of sleep define how the different external (environmental) and internal factors influence sleep and sleep quality. Factors, such as location, the people around us and our age can have a direct effect on the way we prepare ourselves for sleep, how we cope with sleepiness, and how we wake up. In this section we describe factors that we believe have the biggest influence on sleep and therefore must be considered during designing sleep related interaction technology.

Location. Our ability to fall asleep and the quality of sleep heavily depends on where we sleep. An individual can sleep well at home while suffering from insomnia in a hotel room.

Factors such as unfamiliar noises [17, 35], additional stress (as in a hospital) [13], or an uncomfortable bed can all influence our ability to sleep. Alongside the unfamiliarity of a location, comes our anticipation of safety [49, 56]. For example being able to sleep for half an hour in a train may refresh us but the anxiety about safety while sleeping will make it very difficult to fall asleep. People living in war-zones or in refugee camps often suffer from sleep related disorders due to anxiety. Getting a good night's sleep has been found to be difficult for the 62 million people living in relatively unsafe refugee camps.

Another aspect that affects our sleep is the social expectations and norms of the location. The way we can make ourselves comfortable (say undressing or curling up) at home in our bed may not be socially acceptable in an airline lounge even if they have a flat bed. In general, understanding what effect the location has on sleep can trigger the creation of new supportive interaction technologies.

The People around Us. Another important contextual factor is the people who are around us when we are about to sleep. For example, families without children have very different sleeping routines, sleep duration and problems when compared to families with children or to people living alone [3, 18].

Partners sharing one bed face disturbances such as, snoring, individual differences in room temperature preferences and different bedtime habits (wanting to read before sleeping or watching TV). In families with children, pre-sleep as well as waking-up routines are important activities for ensuring good sleep for both the children and the parents. There maybe other routines in place for when the parent wakes up due to young child having difficulty falling asleep. Interestingly partners often say they have a better sleep when they are together instead of sleeping alone, while objective measurements show the opposite.

Recently, a number of different concepts for re-connecting couples and family members have been proposed [9, 11, 29, 30]. These concepts could indirectly help to improve sleep quality by creating ‘peace of mind’. However they can also be extended to include sleep as an essential part of re-connecting by creating a more intimate link between partners.

Age. People of different ages sleep differently, have different routines and experience different problems with sleep [5, 21]. For example, elderly people suffer from insomnia more often and overall have a lower quality of sleep. Their sleep tends to be more fragmented and their waking up time is earlier than young adults [28]. Elderly people tend to undertake fewer physical activities, have less light exposure and often nap during the day; these factors hamper their night's sleep. Consequently, new solutions should help them to manage a healthier and more physically active routine, not just before sleep but throughout the day. In contrast, busy adults, are often more stressed which, despite generally being in good health, might lead to problems with falling asleep. Solutions for stress management, relaxation guidance and enabling power naps could be suitable for this target group. Sleep enhancing solutions should therefore be designed differently for the two age groups.

4 Opportunities for Design

Thus far, we have discussed how sleep is a dynamic process and affected by various external and internal factors. When designing products or services that support the ritual or activities that influence sleep, we should take into account its basic physiological, psychological and sociological characteristics, as well as contextual factors such as location and culture. Below we propose design opportunities that relate to the aforementioned four factors.

The opportunities are structured around five themes: scheduling for the collective; sleep management; supporting transitions; context of and around sleep; and feigning and inhibiting sleep.

4.1 Scheduling for a Collective

As we noted in the section on Sociology of Sleep, scheduling for a collective is an important sociological aspect of sleep. Traditionally the ordering and time coordination of sleep has been a loosely scheduled activity. But with busy lifestyles and more distributed collective, scheduling sleep these days becomes a more complex activity that can benefit from careful technology mediated design. Below we list some examples of problems with coordination and potential opportunities for HCI.

Although the authors of the reverse alarm clock [39] and the networked alarm clock [48] do not discuss their designs from the perspective of scheduling for a collective, both these designs are examples that promote coordination between members of a collective. In the first instance the alarm clock is encouraging the child to coordinate his awakening to that of his parents.

However, with the latest developments in sleep sensing technologies (such as Zeo [58] described in the previous section) it is possible to redesign the coordination strategies through a much richer user experience. For example, the reverse alarm clock's design can be extended to include children in the age group of 4 – 7 yrs by incorporating appropriately presented information about their parents' sleep state. This could encourage children to only awaken their parents when it's absolutely essential or if they are in a sleep state that will allow them to respond immediately upon being awakened (which can be detected by monitoring the sleep stage of a person). Such awareness of a sleeping member is valuable for all members of a household (not just children). For example if the door of an elderly person living in cared facility could provide awareness cues of the sleep state of the person (information like – when they fell asleep and when are they likely to wake-up) it might help care-givers to decide whether to return later or wake them up. Another usage of sleep awareness was proposed in the Aurama system [9]. The Aurama is a context aware system that provides caregivers with information about elders living independently [9]. The load sensors in the bed were used to distinguish between, moving and staying still (i.e. wake and sleep states); this information together with overall sleep patterns was available to the caregiver. However detailed information about a sleep state, such us lying in bed awake, falling asleep or sleeping, was not available. Such awareness possibilities are only possible now with the relatively robust single electrode EEG sets like Zeo [58].

This form of awareness information can also be shared via network enabled devices so a distributed family can coordinate and schedule their activities. There are interesting design challenges in supporting such activities for a distributed collective that is spread across multiple time-zones. Family calendars and devices could include asleep or in bed mode to let remote participants know that they are not available.

When shared within a social network, it may help strengthen the social relationships between friends and therefore may add a new dimension to the current social relationships in groups.

Such collective scheduling of sleep could also be used by people to plan activities around time when others might be asleep (in a way "abuse" sleep time). For example, one may choose to set-off on a long journey at night to avoid traffic or use the quiet time to work or play on their media console. In these cases a person might not only choose to not disclose their sleep related awareness information but it is also possible to use the relevant networked information to plan their activity to reduce their vulnerability (can I get a head-start on the local traffic?) or capitalize on others vulnerability (to carry out anti-social behavior). This will need to be taken into account when designing a system for supporting sleep awareness for families as well as for social networks.

4.2 Sleep Management

Sleep management refers to how we organize our sleep. As we have described in the sleep section many sleep problems come from psychological disturbances, inappropriate sleep routines or busy lifestyles. As a result people start to sleep less [28]. In this situation being able to get a short nap during the day is not anymore a luxury but a necessity. Design solutions aimed at sleep management can therefore not only address night time sleep but also daytime naps. Below we present two examples of designing systems for improving sleep management.

Event Driven Alarms. Most HCI work related to sleep has revolved around the redesign of the alarm clock. The main focus of the redesign lies in adjusting the wake up time based on predetermined condition(s) and in providing contextual information that would help to adjust the usual morning routine if necessary. Despite adding context and events the proposed alarm clock modifications are still time based. While these are useful in a bedroom setting these alarms are of relatively little use in locations like the office or train. It might be interesting to explore alarms that can awaken a person based on a specific event without including any time constraints.

Although the network alarm clock [48] is an event driven alarm clock where the event is what other members of the social network of the person are doing, the authors do not interpret their alarm clock as an event-driven clock and potentially miss some interesting design possibilities that emerge from this form of interpretation which leads to more possible forms of sleep management. For example, a person napping in a train would find it useful if the clock went-off when their station is approaching. Rather than having a time-based alarm, it would go off at the anticipated station arrival time. The clock could monitor the train information and the user's sleep stage to trigger the alarm by optimizing arrival time with sleep stage to wake the person at the appropriate moment. Similarly a project coordinator might choose to nap until he receives a critical email from a colleague. An event driven alarm could help the person have a more "intense" nap knowing fully well that they will be awakened by the alarm only when an email from the right person arrives. This can lead to well rested employees and improved sleep scheduling through napping.

Improving Sleep Hygiene. Opportunities exist to combine remote day time activity monitoring and sleep monitoring technologies to draw a complete picture of day time and night time behavior. As discussed previously, sleep is influenced by many psycho-physiological factors, such as light exposure, physical activity, drinking and eating behavior throughout a day. Thus, more tailored advice can be provided to individuals as a persuasive tool to improve their sleep behavior.

4.3 Supporting Transitions

Supporting transition phases of sleep (sleep onset and waking up) can help a person fall asleep and wake up feeling fully rested. One needs to carefully consider the physiological and psychological aspects of falling asleep and waking up.

The main challenge during winding down is to reduce the disturbances from the environment by supporting consistent bedtime routines by adapting familiar routines for different location and people around us. For example, bedtime routines from a bedroom can be adapted and re-appropriated for an airline lounge or office nap. Similarly, routines can be adapted for minimal disruption when a bed partner is away on business trip.

Recently several exciting products that support waking up such as EasyWakeup© (easywakeup.net), and Philips Wake-up Light©, have become available. Making the waking up more gradual (e.g. such as with the Philips Wake-up Light) can insure a pleasant wake up experience. However these products rarely consider supporting different requirements of couples who have different wake-up schedules. The additional challenge here is to achieve a gradual wake up without disturbing the sleeping partner.

Another example is expressive alarm clock described in [55], where the alarm sound of the clock depends on the way the user sets the alarm time (emotional state and level of urgency). As described in the sleep section the waking moment as well as how a person was awakened affects how the person feels. By using sleep monitoring device to detect the sleep stage the waking up experience can be further enhanced by alarm sound selection and by shifting the actual wake up moment (e.g. based on the level of urgency).

Sleep Onset in the Presence of Multi-media. Nowadays it is increasingly common for the bedroom to include a television and for bedtime rituals to include watching TV in bed. Watching TV in bed is generally considered a bad influence on sleep. The constantly changing multimedia images and evolving storylines in the program can cause anxiety and increase arousal, whereas to induce sleep onset one needs to be relaxed and calmed. In addition, the bedroom should be associated only with sleep [53]), thus watching TV should be kept outside the bedroom. While moving the TV out of the bedroom is a simple solution it is unfortunately not the desired solution for most people and that is one reason why every hotel room today includes a TV in the bedroom.

The challenge here is to redesign the TV such that it facilitates the transition from awake to sleep for the user. For example, the TV could have a "bedtime" mode where the TV recommends a program that not only includes appropriate content but also includes visual elements that avoids blue, or slowly reduces the contrast, brightness, volume and image quality to create a blurry ‘sleepy’ effect, that can help to induce sleep onset.

When a user transits from being awake to asleep his/her capabilities as well as needs for interaction changes. For example, when a person is almost asleep he only needs an “off” button whilst when awake a “change channel” button is more appropriate. The decreasing of cognitive and physical state of a person during a sleep initiating process means a bedroom product should adapt and alter its interaction mode from explicit (requiring active user control) to implicit (fully automatic). When a user is fully awake, he/she should be able to take full control over the device, whereas, when he/she is fully asleep, the device should operate automatically. The challenge lies in balancing the two types of interactions for the transition zone from explicit to implicit and on the degree of control the system has. Users do not always value systems that think for them, even when they are not in a state to do so themselves.

At bedtime, bedrooms are quiet and the light level is generally lower, with people avoiding too much physical activity (except for sex, [53]). Therefore, an interface that requires precise user input, e.g. small buttons, or has a light based user interface, e.g. LCD screen, are not suitable in this context. In such situations UIs need to be redesigned to allow quick and effortless interaction; effortless meaning minimal rise in a user’s arousal.

4.4 Context of and around Sleep

People sleep in different contexts which can also affect their quality of sleep, for example, one might require a dark and quiet room to sleep in at night, but one may be less conscious about sleeping on a noisy and crowded train. People might find it disturbing when sharing a bed with a partner but at the same time feel lonely when alone in bed. These different contextual requirements require careful consideration when designing for and around sleep. Below we describe one such context in further detail.

Bed Sharing Context. People who sleep by sharing a bed with a partner need to face and deal with issues such as different bedtime schedules, rituals and preferences regarding the sleeping environment. Therefore knowing the preferences of each individual in the bed and adjusting the context or setting of devices in the bedroom can help reduce conflict. A bedroom TV could include a “partner asleep” mode where the volume and brightness settings are adjusted or the TV adapts its orientation (either by physically turning or by using the emerging multi-view 3D displays) to minimize light for the sleeping partner. Waking up alarms can also be designed to use different actuation modes, such as, vibration when one partner wants to sleep for longer, or audio alarms when both partners want to get up at the same time. These forms of contextual design are becoming important as couples have increasingly different work schedules and demands.

4.5 Feigning and Inhibiting Sleep

Most of our discussion up to this point has been on improving the quality of sleep and the experiences around sleep. However there are several situations where it may be useful to feign or inhibit sleep. For example, yawning (real or not) is an important social marker; it indicates to the collective that a person is tired. It may be rude in some settings but in others, it offers the perfect opportunity to excuse oneself from a gathering; a yawn at a dinner party can be followed by a remark about how tired one feels and the request to be excused. It is also not uncommon to feign sleep to listen in on others' conversations or to avoid conversations with a bed mate.

There are opportunities to explore how systems can be designed to support such social activities. If sleep monitoring technology and awareness tools become prevalent and popular, it is possible to envisage situations where the user re-appropriates the awareness provided by these systems to feign sleep or tiredness. There are bound to be ethical debates on how far one can go with such re-appropriation.

An important aspect of sleep management is inhibition of sleep. Stimulants such as coffee are often used to inhibit sleep but their effectiveness has always been questioned depending on the amount of actually consumed caffeine.

Minimizing the need for sleep has been the holy grail for war planners since time immemorial [33]. American pilots are often given Amphetamines (otherwise known as speed) to keep them alert for 30-hour long missions. There are several military research programs (from US and other governments) in place to explore 90-hour long shifts.

Such research ventures are not only of interest to the military but also to shift-workers and increasingly for professionals such as stock analysts, surgeons etc. Most legal sleep inhibition approaches in office environments today rely on stimulants, such as caffeine present in coffee and RedBull©. New ways of sleep inhibition through minimal use of stimulants is of great interest to firms that deal in time-critical work.

Design solutions that combine persuasion techniques with a better understanding of the physiology and sociology of sleep can lead to systems that inhibit sleep. Such solutions can explore passive approaches such as adjusting the lighting condition, bringing together people who need to inhibit sleep and so on. While these approaches can work they may be too slow for a rapidly evolving situation. In such cases, design solutions can also explore electro-magnetic stimulation to temporarily suppress sleep receptors [33]. Such approaches need to be carefully debated and considered from a legal and ethical standpoint.

4.6 Summary

The following list summarizes opportunities identified in each subsection:

- Supporting sleep awareness within a family (co-located and distributed) or a community (students living in a dorm or cared living for older people) and social network aided by sleep sensing technologies
- Supporting event driven wake-up, where event include external factors (e.g. traffic, weather, destination) and internal factor (e.g. sleep stage, desired wake up time, urgency, partner sleep stage, specific wake-up triggers defined by the user)
- Using persuasive tools to provide tailored advice for improving sleep hygiene.

- Supporting sleep onset in the presence of multi-media distracters such as TVs. For example, the color and tone-mapping in these devices can be carefully adjusted through detailed understanding of retinal response and color-perception theory to positively influence the cycle.
- Context sensitive intervention that support sleep for an individual while accommodating potentially conflicting needs of the partner. For example, preserving an individual's sleep ritual of reading a book in the bed while simultaneously being sensitive to the needs of her partner to sleep immediately.
- Persuasive tools, techniques and systems to facilitate sleep inhibition in different situations and contexts.

5 Conclusions

The main contributions of the paper lie in the exploration of a newly emerging topic in the area of sleep that is becoming more important due to changes in society and lifestyle and presents challenges for which the HCI community can rise to. Through a review of the literature we describe the phenomenon of sleep and its position in the social life to provide a rich understanding of Sleep. We present a set of factors that have a direct effect on the way we prepare ourselves for sleep, cope with sleepiness, or wake-up and identify five sleep related themes that have high potential and need for interaction design.

We believe this paper opens up a new topic for HCI research and will serve as a beneficial reference for future researchers.

References

1. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcoft, W., Pollak, C.P.: The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper. Sleep 26, 342–392 (2003)
2. Balkin, T.J., Braun, A.R., Wesensten, N.J., Jeffries, K., Varga, M., Baldwin, P., Belenky, G., Herscovitch, P.: The process of awakening: a PET study of regional brain activity patterns mediating the reestablishment of alertness and consciousness. Brain, 2308–2319 (2001)
3. Biddle, J.E., Hamermesh, D.S.: Sleep & the allocation of time. The Journal of political economy 98(5-1), 922–943 (1990)
4. Bertelson, A.D., Monroe, L.J.: Personality patterns of adolescent poor and good sleepers. Journal of Abnormal Child Psychology 7, 191–197 (1979)
5. Bovill, M., Livingstone, S.: Bedroom culture and the privatization of media use. In: Bovill, M., Livingstone, S. (eds.) Children & their changing media environment: a European comparative study. Lawrence Erlbaum, Mahwah (2001)
6. Cannon, W.B.: Bodily changes in pain, hunger, fear and rage: An account of recent research into the function of emotional excitement, 2nd edn. Appleton-Century-Crofts, New York (1929)
7. Chesson, A.L., Ferber, R.A., Fry, J.M., Grigg-Damberger, M., Hartse, K.M., Hurwitz, T.D., Johnson, S., Kader, G.A., Littner, M., Rosen, G., Sangal, R.B., Schmidt-Nowaram, W., Sher, A.: The indications for polysomnography and related procedures. Sleep 20, 423–487 (1997)

8. Choe, E.K., Kientz, J.A., Halko, S., Fonville, A., Sakaguchi, D., Watson, N.F.: Opportunities for computing to support healthy sleep behavior. In: Ext. Abstracts CHI 2010, pp. 3661–3666. ACM, New York (2010)
9. Dadlani, P., Sinitsyn, A., Fontijn, W., Markopoulos, P.: Aurama: caregiver awareness for living independently with an augmented picture frame display. AI & Society 25(2), 233–245 (2010)
10. Davis, H., Davis, P.A., Loomis, A.L., Harvey, E.N., Hobart, G.: Electrical reactions of the human brain to auditory stimulation during sleep. Journal of Neurophysiology 2, 500–514 (1939)
11. Dodge, C.: The bed: a medium for intimate communication. In: Ext. Abstr. CHI 1997. ACM Press, New York (1997)
12. Edinger, J.D., Fins, A.I., Sullivan, R.J., Marsh, G.R., Dailey, D.S., Hope, T.V., Young, M., Shaw, E., Carlson, D., Vasilas, D.: Sleep in the laboratory and sleep at home: comparisons of older insomniacs and normal sleepers. Sleep 20, 1119–1126 (1997)
13. Freedman, N., Kotzer, N., Schwab, R.: Patient perception of sleep quality and etiology of sleep disruption in the intensive care unit. Am. J. Respir. Crit. Care. Med. 159, 1155–1162 (1999)
14. Freedman, R.R.: EEG power spectra in sleep-onset insomnia. Electroencephalography and Clinical Neurophysiology 63(5), 408–413 (1986)
15. Fortunato, V., Harsh, J.: Stress & Sleep Quality: The Moderating Role of Negative Affectivity. Personality and Individual Differences 41, 821–836 (2006)
16. Foulkes, D., Vogel, G.: Mental activity at sleep onset. Journal of Abnormal Psychology 70(4), 231–243 (1965)
17. Griefahn, B., Spreng, M.: Disturbed sleep patterns and limitation of noise. Noise & Health 6(22), 27–33 (2004)
18. Hale, L.: Who has time to sleep? Journal of Public Health 27(2), 205–211 (2005)
19. Hauri, P.J., Wisbey, J.: Wrist actigraphy in insomnia. Sleep 15, 293–301 (1992)
20. Hirshkowitz, M., Moore, A.C., Hamilton, C.R., Rando, K.C., Karacan, I.: Sleep apnea in patients with quadriplegia. Journal of Clinical Neurophysiology, 56–62 (1992)
21. Hume, K., Van, F., Watson, A.: A field study of age & gender differences in habitual adult sleep. Journal of Sleep Research 7(2), 85–94 (1998)
22. Jewett, M.E., Kronauer, R.E., Czeisler, C.A.: Phase/amplitude resetting of the human circadian pacemaker via bright light: A further analysis. Journal of Biological Rhythms 9, 295–314 (1994)
23. Kalra, G.K., Banks, S., Dinges, D.F.: Relationship between polysomnography and simultaneous actigraphy in healthy subjects. Poster Presented at Sleep, 22nd Annual Meeting of the Associated Professional Sleep Societies, LLC (2008)
24. Kim, S., Kientz, J.A., Patel, S.N., Abowd, G.D.: Are you sleeping?: sharing portrayed sleeping status within a social network. In: Proc. CHI 2008, pp. 619–628 (2008)
25. Krauchi, K., Cajochen, C., Wirz-Justice, A.: Waking up properly: is there a role of thermoregulation in sleep inertia? Journal of sleep research, 121–127 (2004)
26. Lack, L.C., Gradisar, M., Van Someren, E.J.W., Wright, H.R., Lushington, K.: The relationship between insomnia and body temperatures. Sleep Medicine Reviews 12(4), 307–317 (2008)
27. Landry, B.M., Pierce, J.S., Isbell, C.L.: Supporting routine decision-making with a next-generation alarm clock. Per.Ubiq. Comput. 8(3-4), 154–160 (2004)
28. Lawrence, L.J., Mardon, S.: The Harvard Medical School guide to a good night's sleep. McGraw-Hill, New York (2007)

29. Lottridge, D., Masson, N., Mackay, W.: Sharing empty moments: design for remote couples. In: Ext. Abstracts CHI 2009, pp. 2329–2338. ACM, New York (2009)
30. Mhorain, A., Agamanolis, S.: Aura: an intimate remote awareness system based on sleep patterns. In: Proc. of the CHI Workshop on Awareness Systems (2005)
31. Miltler, M.M., Miller, J.C.: Methods of testing for sleepiness. Behavioral Medicine 21(4), 171–183 (1996)
32. Monroe, L.J.: Psychological and physiological differences between good and poor sleepers. Journal of Abnormal Psychology 72, 255–264 (1967)
33. Moreno, J.: Juicing the brain. Scientific American Mind 17(6), 66–73 (2006)
34. Morin, C.M.: Insomnia: Psychological Assessment and Management. Guilford, New York (2003)
35. Muzet, A.: Environmental noise, sleep and health. Sleep Medicine Reviews 11, 135–142 (2007)
36. National Sleep Foundation. 2008&2010 Sleep in America Poll, `http://www.sleepfoundation.org`
37. Ogilvie, R., Wilkinson, R.: The detection of sleep onset: behavioral and physiological convergence. Psychophysiology 21, 510–520 (1984)
38. Owens, J.A.: Sleep in children: Cross-cultural perspectives. Sleep and Biol. R. 2, 165–173 (2004)
39. Ozenc, K.F., Brommer, J.P., Jeong, B.-k., Shih, N., Au, K., Zimmerman, J.: Reverse alarm clock: a research through design example of designing for the self. In: Proc. DPPI 2007, pp. 392–406. ACM, New York (2007)
40. Patel, S.R., Hu, F.B.: Short sleep duration and weight gain: a systematic review. Obesity (Silver Spring) 16(3), 643–653 (2008)
41. Piggins, H.D.: Human clock genes. Annals of Medicine 34(5), 394–400 (2002)
42. Punjabi, N.M., Shahar, E., Redline, S., Gottlieb, D.J., Givelber, R., Resnick, H.E.: Sleep-Disordered Breathing, Glucose Intolerance, and Insulin Resistance. American Journal of Epidemiology 160(6), 521–530 (2004)
43. Renger, J.J.: Overview of Experimental and Conventional Pharmacological Approaches in the Treatment of Sleep and Wake Disorders. Current Topics in Medicinal Chemistry 8, 937–953 (2008)
44. Richardson, G., Tate, B.: Hormonal and pharmacological manipulation of the circadian clock: recent developments and future strategies. Sleep 23, 77–85 (2000)
45. Riemann, D., Perlis, M.L.: The treatments of chronic insomnia: a review of benzodiazepine receptor agonists and psychological and behavioral therapies. Sleep Medicine Reviews 13(3), 205–214 (2009)
46. Sadeh, A., Hauri, P.J., Kripke, D.F., Lavie, P.: The role of actigraphy in the evaluation of sleep disorders. Sleep 18, 288–302 (1995)
47. Scherini, T., Melo, P., van Craenendonck, T., Zou, W., Kaptein, M.: Enhancing the sleeping quality of partners living apart. In: Proc. DIS 2010, pp. 171–174. ACM, New York (2010)
48. Schmidt, A.: Network alarm clock. Per.Ubiq. Comput. 10(2-3), 191–192 (2006)
49. Schwartz, B.: Notes on the Sociology of Sleep. The Sociological Quarterly 11(4), 485–499 (1970)
50. Sermon, P.: Telematic Dreaming. Leonardo 33(2) (2000)
51. Shamini, J., Shauna, L.S., Summer, S., Scott, C.R., Paul, J.M., Iris, B., Gary, E.R.S.: A randomized controlled trial of mindfulness meditation versus relaxation training: Effects on distress, positive states of mind, rumination, and distraction. Analysis of behavioral medicine 33(1), 11–21 (2007)

52. Smith, M.T., Perlis, M.L., Park, A., Smith, M.S., Pennington, J.M., Giles, D.E., Buysse, D.J.: Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. American Journal of Psychiatry, 5–11 (2002)
53. Stepanski, E.J., Wyatt, J.K.: Use of sleep hygiene in the treatment of insomnia. Sleep Medicine Reviews 7(3), 215–225 (2003)
54. Verkasalo, P.K., Lillberg, K., Stevens, R.G., Hublin, C., Partinen, M., Koskenvuo, M., Kaprio, J.: Sleep Duration and Breast Cancer: A Prospective Cohort Study. Cancer Research 65(20), 9595–9600 (2005)
55. Wensveen, S.A.G., Overbeeke, C.J., Djajadiningrat, J.P.: Push me, shove me and I show you how you feel. Recognizing mood from emotionally rich interaction. In: Proc. of DIS 2002, London, UK, June 25-28, pp. 335–340 (2002)
56. Williams, S.J.: The Social Etiquette of Sleep: Some Sociological Reflections and Observations. Sociology 41(2), 313–328 (2007)
57. Williams, S.J.: The Sociological Significance of Sleep: Progress, Problems and Prospects. Sociology Compass 2(2), 639–653 (2008)
58. Zeo Personal Sleep Coach, http://www.myzeo.com/

FeetUp: A Playful Accessory to Practice Social Skills through Free-Play Experiences

Andrea Rosales, Ernesto Arroyo, and Josep Blat

Universitat Pompeu Fabra, Carrer de Roc Boronat 122-140
08018, Barcelona, España
{andrea.rosales,Ernesto.arroyo,josep.blat}@upf.edu

Abstract. In this paper we describe the design process of an interactive accessory to play anywhere and anytime while encouraging free-play and practice social skills. We explain the design process, the resulting conceptual design of FeetUp and the preliminary user's evaluation. FeetUp is a playful accessory that takes advantage of children's interest to jump, or perform body stunts. These activities generally include lifting both feet, and FeetUp gives audiovisual feedback whenever this happens to encourage free-play related with jump activities. Preliminary user's experience shows how FeetUp, encourages freeplay.

Keywords: free-play, children, playful, augmented technologies, social interaction, wearable.

1 Introduction

Children are in continuous training and their way of being will be strongly influenced by all their current experiences [12] [8]. By playing children mimic other's experiences and learn about it [10], this way they build their own values and develop new skills.

Many authors have discussed the correlation in children's life between their increased interest on screens (either television, videogames, computer, mobile phone, or others) and their lower physical activity and/or social interaction [1], [11], [14]. More interest on screens prevents them from experiencing valuable face-to-face interaction important for practicing social skills, which could be developed through free-play. Free-play can be open-ended, spontaneous, physical, collaborative, engaging, fun, pleasurable, and usually involves a pretend element [17].

Several projects have explored the use of objects to encourage free-play [1], [9], [4], [16]). Some of these objects are designed to stimulate physical play, and regard social activity as a, secondary goal. Other objects have been designed to stimulate social interaction by providing open-ended scenarios [9], [7]. However, in most cases the object to be augmented lacks well-defined selection criteria, or its interaction affordances does not add value to free-play. E.g., the interaction design proposed by Ridén for an augmented rubber ball was not useful [16]. The authors noticed that "the

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 37–44, 2011.

interaction enabled by its flashing lights was not used at all" and they concluded "the force of affordance of balls is very strong for children".

Our approach focuses on understanding children's interest, habits and routines related to evocative objects and free-play activities. We use this knowledge as initial requirements to design and create evocative objects that encourage social interaction through free-play. In this paper we present FeepUp, an evocative object designed to encourage children's to play using their bodies to jump, roll, perform handstands or any other stunts collaboratively with their friends. We describe the design process, the resulting conceptual design of FeetUp, including the object design, its interaction rules, and the physical setup. Finally we present preliminary results indicating that FeetUp encourages free-play and discuss opportunities for future work.

2 Design Process

The design process started by two ethnographic studies aimed at understanding, the relationship between free-play and children's routines, in order to identify the implications for design [5]. This sets the framework for designing playful objects to promote free-play by involving children in the design process [6] through brainstorming, iterative prototyping and evaluation.

2.1 Children and Free-Play

In a preliminary ethnographic study, we used traditional methodologies, such as observing children in their natural play settings as observer-participants note taking, and then categorizing the collected information [2].

We followed children's between 6-9 years old while playing in Barcelona during 24 sessions in parks, school patios, and their homes. We documented children's relation with free-play, what they do, and what objects they use during free-play sessions. We registered our observations in a notebook, and then we categorized the observations, to extract relations between objects, activities, games and social interaction related with free-play.

From this process emerged that the most common object in children's activities or games is their own body. They naturally played games related with physical challenges with their body. We identified several free play challenges they typically find in their body, such as playing against gravity, aim, coordination or visibility, attention, strength, etc. Each challenge is related to many activities or games. For example, while playing with gravity, they enjoy handstands, handholding, hanging, hand walking, somersaults, tumbles, or carrying each other.

2.2 Children's Routines

A second ethnographic study identified opportunities on children's routines for free-play. We documented the routines of 8, 2 to 4 members families, all of them with at least one kid aged between 6 and 9, as participant-observers [2]. We immersed in their spare time activities and conducted unstructured interviews [2] with parents.

The study indicated that children's routines lack spare time or free time. Family routines are full of duties, and children's time has to be planned according to them. In addition to attending school, practicing sports, and doing extracurricular activities, children are also participants in many of their family's tasks, like going to the supermarket or visiting their aunt's house or following their brother's activities. However, free-play can also happen in parallel to other activities, like playing while shopping, during a family visit, or on the road.

Objects designed to stimulate free-play should be small enough to be taken along while going out, and practical enough so that children can always carry them. Objects could also be embedded into smart clothes and offer additional value as suggested by Steffen [15]. As children wear clothes all the time, they can act as playful objects to allow children to play whenever they have an opportunity, whereas at on the road, on a waiting room, or at the store.

2.3 Brainstorming with Children

The Current design trends suggest that children should be part of the design team in order to voice their opinions and inform the design with their own interest, emotions and feelings [6]. For the design of FeetUp, children collaborated with researchers, providing ideas for interaction and providing useful information about adequate ways to stimulate free-play. At the beginning of the process we presented children with sketches of 5 imaginary powerful accessories. (See Figure 1).

Fig. 1. Children were presented with 5 inspiring sketches of imaginary powerful accessories: a fanny pack, a belt, a brooch, gloves and shoes

We explained to children that these objects were powerful ("magical", and "limitless"). In order to evoke children's imagination, we asked them to describe and explain the sketches: what the pictured kids where doing, and how the pictured devices would work.

This process provided many interaction design ideas for many playful accessories, and also provided valuable insights to identify the most adequate accessories to stimulate body challenges in diverse playful situations. For the final selection, we considered only playful accessories that could be developed with existing hardware. We selected three playful accessories that elicit numerous opportunities to practice social skills and allow them to imagine several situations and evoke possibilities for play making use of body challenges. The selected projects include FeetUp, Statue, and Gloves The two other projects were developed at later time.

2.4 Wizard-of-Oz and Imaginary Experience

We started the iterative design process with early prototypes, to validate the concept, and verify if the prototype could be used for what it was designed for. Several sessions allowed children to test every new improvement, and re-arrange the pieces to fit their bodies and interests. The first iterations demonstrated the ideas with a combination of wizard-of-oz and partially functional prototypes to test and validate the initial concepts. At the beginning, part of the functionality was implemented and the rest was operated manually activated sensors. This method was used to illustrate how some of the accessories would work, facilitating the user to imagine experiencing and playing with each accessory. This gave us important insights to evolve the design according to how users imagine this will be working and what for.

2.5 Iterative Prototyping

We improved the original design according to users' feedback and introduced several new functionalities that emerged from interacting with the objects themselves. For every test we invited a new user to a playground to play with the prototypical accessories and drew conclusions about how to improve the design, considering how users spontaneously used the device? What for? And how users expected the device to work?

The results of every new test guided the following iteration in the design. This process finished when the prototype was fully implemented and the concept was validated with individual users. Trough validation, users confirmed that they managed to practice diverse activities with the accessory and imagined games where they could use it. Important improvements emerged from user's ideas about how to arrange sensor to properly detect jumping, how to arrange the actuators for visibility, or the kind of feedback they expect to have in the games they where imaginary playing.

3 Conceptual Design

FeetUp is a playful accessory embedded in a pair of children's shoes that provides audiovisual feedback whenever the user jumps or is off the ground. (See Figure 2). FeetUp is similar to popular shoes that light while walking, but FeetUp demands additional effort as kids have to jump to get the feedback. Since games and playful activities have to have some degree of challenge [3] [13], increasing the difficulty to get the feedback opens a gap to make more interesting plays around it.

Fig. 2. FeetUp

FeetUp stimulates children to play against gravity, one of their most frequent activities related to free-play. Giving feedback while jumping lets each child share his jumping achievements with his colleagues, and when several children have the same accessory, they have the same interest in common and they can have shared goals. While sharing goals they must explain ideas, argue, negotiate, and make agreements, thus they are practicing quite important social skills.

There is no digital network between users, but the common interest is the excuse for them to create a network of shared goals, represented as simple activities or more complex games. They have to propose how to use this shared feedback to play together with this new shoes. The system incorporates a simple rules system so that children have the opportunity to create their own system of rules, challenges and goals based on the basic, but consistent information they receive.

3.1 Object Design

For the object design we considered the use of an existing object children already wear and have with them all the time. Thus we choose a pair of shoes that can be used anywhere, and offer limitless opportunities to play. The audiovisual feedback is directly embedded in the shoes themselves to maximize its visibility.

3.2 Interaction Rules

The interaction system reacts to only one factor; children jumping. The shoes rules give feedback when children are jumping; they blink and play sounds when both feet are not touching any surface.

3.3 Physical Setup

The setup includes 4 touch sensors, 2 microcontrollers, and 2 emitter and receiver radios, 2 LED arrays, and 2 piezo speakers. The setup includes two pressure sensors placed on the insole inside each shoe (one in the base of the heel and another on the sole), thus allowing us to detect when each foot had been completely lifted from the floor. A microcontroller in each foot reads sensor data, validates that both sensors have been de-activated, and sends a radio frequency signal read by the other device. When both microcontrollers confirm that all pressure sensors have been de-activated, they trigger the LEDs in the surface of the shoe and make a sound alert with a piezo speaker.

4 Evaluation

A pair of children tested the final prototypes playing with the shoes in a playground during 30 minutes, in a preliminary evaluation of the project. Researchers did structured observations, to identify how the FeetUp accessory yielded the characteristics commonly present in free-play. Observers took notes during the session, and at the end of it, answered some open questions.

Observers described 7 concepts related with free-play behavior [17], including spontaneous, voluntary, collaborative, open-ended and engaging play, being physically active and with a pretended element. Some observations are described in the following list:

- They enjoyed discovering the accessory and understanding how it worked; they discovered that it blinked when jumping. They *spontaneously* started to play challenges related on how to cheat the shoes and discovered how to cheat it by making it blink without jumping; moving their feet while sitting, or tiptoeing.
- They exchanged their discoveries, and motivated each other to start new challenges. They paid attention to their partner activities and made comments about them, while *collaboratively* helping each other.
- They were free to do whatever they wanted in the park. However, they *voluntarily* moved-on to other jump related activities, and got involved in making the lights blink by dancing, hanging, sliding, or jumping.
- They played games using simple goals, games they were used to. Later they started to adapt the activities to better see the lights of the shoes, to keep the lights blinking longer, or to make the activity more and more difficult in an *open-ended* encouraging game.
- They *engaged* in the experience, preferring to play with the accessory rather than playing with games in the park where the testing took place, and the activity occupied all the time they usually spend in the park.
- They were active and had continuous *physical* activity during the test.
- They described the accessory as “magical shoes” and they imagined that the shoes could give the super-powers and allow them to control others shoes, or other people. However, the accessory was no used as a *pretend element*.

Fun and pleasure values, also fundamental in free-play [17] were not considered, as they will be observed using an extended framework in future work.

Children used verbal and non-verbal communication while discovering the accessory, while imagine what they could do with the powerful shoes, and finally while playing. With verbal communication they described their hypothesis about the shoes, described their achievements, express satisfaction or frustration, and give support to his friend. Non-verbal communication challenged them to imitate each other, or establish new challenges. This indicates that they where practicing social skills.

Some inadequate ergonomics of the design generated frustration and disappointing experiences in children. This was due to the size and intrusiveness of the devices' hardware, as is also mentioned by Steffen [15].

5 Conclusions and Future Work

We have presented the design process of a playful accessory to practice social skills through free-play. FeetUp is a playful accessory to play against gravity; it encourages children to perform jumping challenges.

The ethnographic studies gave us important insights to create a conceptual framework which argued that children could benefit with playful clothes, playful bags or playful accessories and that these objects can support their innate need for playing.

The participatory design process, and the activities proposed to involve children in the design process, allow us to involve children's insights, during the entire process and make informed design decisions to fit their interest and preferences.

Future work includes, testing the prototypes with larger groups of children, and evaluating the accessory using a specific framework to identify how FeetUp encourage fun and pleasure. Future work will also consider 1) Developing conceptual designs and prototypes for other physical activities often performed by children in their games. 2) Comparing the results of the evaluation of each design. 3) Define guidelines to design objects to stimulate practicing social skill.

Acknowledgments. This work has been partially funded by the Learn 3 project, (TIN2008-05163/TSI). The authors would like to thank the cultural and technical partners of the C3I2 workgroup for their support and ideas.

References

1. Bekker, T., Sturm, T., Eggen, B.: Designing playful interactions for social interaction and physical play. Personal and Ubiquitous Computing 14, 385–396 (2010)
2. Blomberg, J., Burrel, M., Guest, G.: An ethnographic approach to design, The Human-computer interaction handbook. Lawrence Erlbaum Associates, New Jersey (2003)
3. Costello, B., Edmonds, E.: A study in play, pleasure and interaction design. In: Conference on Designing Pleasurable Products and Interfaces, pp. 76–91. ACM, Helsinki (2007)
4. Creighton, E.: Jogo An Explorative Design for Free Play. In: 9th International Conference on Interaction Design and Children, pp. 178–181. ACM Press, Barcelona (2010)
5. Dourish, P.: Implications for design. In: Conference on Human Factors in Computing Systems, pp. 541–550. ACM, Montréal (2006)
6. Druin, A.: The Role of Children in the Design of New Technology. Behaviour and Information Technology 21(1), 1–25 (2002)
7. Ferris, K., Bannon, L.: ...a load of ould boxology! In: 4th Conference on Designing Interactive Systems, pp. 41–49. ACM, London (2002)

8. Ginsburg, K.: The Importance of Play in Promoting Healthy Child Development and Maintaining Strong Parent-Child Bonds. American Academy of Pediatrics, Pediatrics (2006)
9. Iguchi, K., Inakage, M.: Morel. In: Conference on Advances in Computer Entertainment Technology, Hollywood, California (2006)
10. Nachmanovitch, S.: Free Play:Improvisation in Life and Art. Penguim Putnam, New York (1990)
11. Leal Penados, A., Gielen, M., Stappers, P.J., Jongert, T.: Get up and move: an interactive cuddly toy that stimulates physical activity. Personal and Ubiquitous Computing 14(5), 397–406 (2009)
12. Piaget, J.: Science of education and the psychology of the child. Orion Press, New York (1970)
13. Prensky, M.: Digital Game-Based Learning. In: Fun, Play and Games: What Make Games Engaging. McGraw-Hill, New York (2001)
14. Seitinger, S.: An ecological approach to children's playground props. In: 5th International Conference on Interaction Design and Children, pp. 117–120. ACM, Tampere (2006)
15. Steffen, D., Adler, F., Weber Marin, A.: Smart Semantics Product Semantics of Smart Clothes. In: International Asociation of Societies of Design Research, pp. 79–88. ACM, Seoul (2009)
16. Ridén, A., Wiberg, A., Maehr, W., Nordqvist, T.: UbiComp and Toys The Difficulties of Integrating Technology into the Everyday Object. In: 2th Scandinavian Student Interaction Design Research Conference, Gothenburg, Sweden (2006)
17. Rosales, A.: Collective creation of games using free-play technologies. In: 9th International Conference on Interaction Design and Children, pp. 335–339. ACM, Barcelona (2010)

Designing *Snakey*: A Tangible User Interface Supporting Well Path Planning

John Harris[1], James Young[2], Nicole Sultanum[1], Paul Lapides[1], Ehud Sharlin[1], and Mario Costa Sousa[1]

[1] University of Calgary, Canada
{harrisjj,nbsultan,ehud,smcosta}@ucalgary.ca
paul.lapides@gmail.com
[2] University of Manitoba, Canada
young@cs.umanitoba.ca

Abstract. We present our ongoing design of *Snakey*, a tangible user interface (TUI) dedicated to the field of reservoir geosciences and engineering. Our design of *Snakey* emphasizes intuitive manipulation and interaction with 3D curves, common to underground well path exploration and planning. *Snakey* is not designed to replace the traditional CAD reservoir well path planning tools, but rather to augment them with tangible and collaborative interaction, facilitating spatial exploration during the multi-stage planning processes involved in well path design. Our short paper presents design goals and prototyping solutions relating to *Snakey*'s physicality, sensing technology, input/output mapping, and multi-modal feedback, as well as our findings from a preliminary evaluation of the latest *Snakey* prototype in a design critique session by experienced reservoir engineers.

Keywords: Well path planning, Tangible user interfaces (TUIs), physical interaction, real-time 3D curve manipulation, interactive visualization techniques, reservoir and geosciences engineering.

1 Introduction

The petroleum industry has been exploring and proposing different solutions for well planning tasks [1]. Well planning is a fundamental task present in different stages of oil/gas field development from early exploration to production. Traditionally, well planning is an iterative process, involving teams of multi-disciplinary experts with specific constraints and goals. The 3D well trajectory is often laid out only in 2D maps and cross sections, making it difficult to understand the exact spatial relationships between the 3D reservoir model (synthesized from geological survey data) and the planned well-bore. Desktop-based solutions for well planning have been adopted by the industry, typically with WIMP interfaces, with wells trajectories being represented on the screen as 3D curves located inside a 3D digital model of the reservoir geology and its flow properties. The task of positioning well trajectories using the current WIMP interfaces involves a large number of control-point manipulations applied along each curve representing the well, leading to a time consuming and non-intuitive operation.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 45–53, 2011.

Well planning tasks have benefited from the use of various virtual reality (VR) interaction systems which allowed designers increased immersion in the path exploration task [1,2]. However, these VR solutions are not physical and do not provide tangible affordances. *Snakey* (Fig. 1) is being designed in order to match some of the unique properties of tangible user interfaces (TUIs) to the inherent physical and collaborative nature of 3D well-path planning in reservoir geosciences and engineering (G&E). The *Snakey* interface consists of 1) a graspable, semi-rigid physical device (similar to a long plastic "snake") which mirrors the shape of an underground well path, 2) a visual display directly behind the hanging physical device which reflects the users' physical manipulations onto a virtual model of a well path and its surrounding geological survey data, and 3) output elements (LED lights and vibration motors) embedded in the physical device that provide direct haptic and visual feedback to the user.

The *Snakey* TUI attempts to leverage users' natural spatial perception and manipulation skills, mapping the task of virtual 3D well path onto a meaningful physical representation [3]. *Snakey* directly couples the 3D well path perception and action spaces into a single physical object, lowering the cognitive load required for interaction and improving the spatial mapping and ease of use [4]. *Snakey* is not designed to replace traditional CAD well path planning tools, but is an exploratory application of the tangible, collaborative, and immediate nature of TUIs to the complex, multi-stage planning processes involved in well path design. It is aimed at the reservoir engineering stage, when fluid flow simulations are performed and wells are manipulated to allow for different simulation runs and better understanding on the flow behavior across the reservoir.

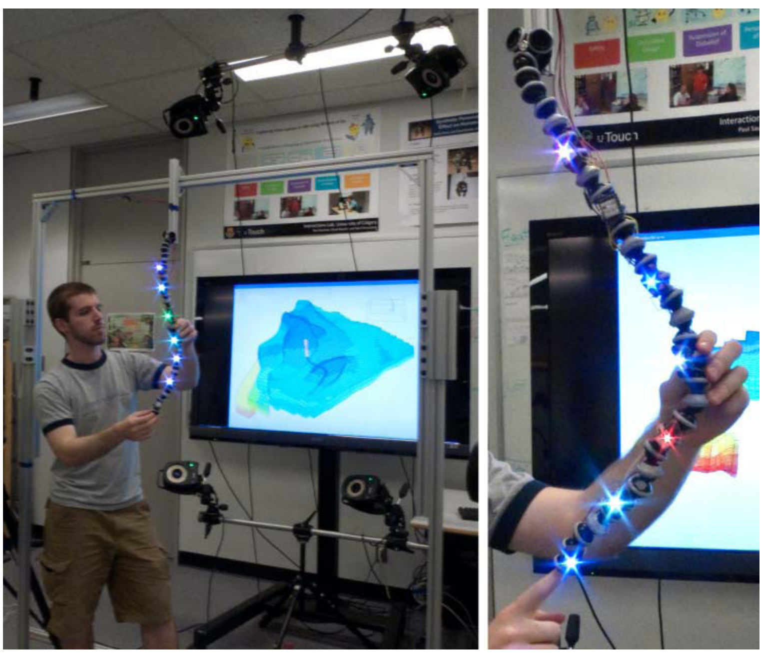


Fig. 1. The *Snakey* interface, visual display, and Vicon mocap camera array

2 Related Work

Physical manipulation of 3D curves was one of the early focuses of TUI research. In 1999, Balakrishnan et al. [5] presented their work on using Measurand's ShapeTape™, an input device that can sense its 3D bend and twist, for interactive manipulation of 3D curves and surfaces. ShapeTape is a rubber ribbon which integrates internal tracking via embedded fiber optic bend and twist sensors. Balakrishnan et al. mentioned attaching spring steel to ShapeTape to provide more physical constraints to the rubber core [5], however to our knowledge there is no published work on integrating ShapeTape into an interface that can physically hold its shape consistently in a manner required by our design goals.

AR-Jig [6] provides interactive physical mapping to 2D curves using a handheld pin-array, which can be used to control 3D surfaces. However, the pins provide a relatively small, and physically discrete spatial control area that we believe would be difficult to meaningfully map to direct interaction with large 3D curves.

Closer to our applied domain of reservoir G&E, GeoTUI [7] provides a tabletop TUI for interaction with geological data, however the interface does not support 3D physical interaction and spatial manipulation of the 3D curves of well paths. Tateosian et al. [8] allow users to manipulate and explore detailed geological tasks using physical 3D surface sculpting, however their interface does not allow physical interaction with a consistent, shape holding 3D curve that can be mapped to spatial explorations of wells in reservoir G&E.

Like others [7], we believe that the reservoir G&E domain can benefit from the exploration of new physical interfaces that will afford a more intuitive and natural interaction with the well path planning task.

3 *Snakey*'s Design Goals and Implementation

The primary design goal of *Snakey* was selecting the TUI's properties to create a valid interaction metaphor for the design and planning of 3D well paths in reservoir G&E visualization tasks. Derived from this goal, the device should afford simple and intuitive physical manipulation of 3D curves in real time *while maintaining* its 3D shape throughout the interaction session, unless directly manipulated by the user. We envisioned reservoir engineers, a single user or a collaborative group, being able to relate to the 3D physical *Snakey* as a constant and consistent representation of the virtual well path. Devising a means for *Snakey* to persistently represent the well, while at the same time allowing users to manipulate and interact with its curve, was our main design challenge.

Another design goal was to allow flexible and easy view and physical access to *Snakey*, enabling reservoir engineers to visualize the overall 3D trajectory of the well from different viewpoints in relation to the reservoir model. We also wanted to support modularity, reconfigurability and branching, as wells are often planned with junctions and forks, establishing multilateral 3D configurations (Fig. 3).

Furthermore we investigated the integration of other physical modalities and properties, beyond the spatial, into *Snakey*. Embedded lighting indicators could highlight sections of the well path that needed further attention. Embedding haptic feedback could allow users to have direct insight relating to some of the well properties, such as fluid flow through the well path or its vicinity, or providing indications of having reached one of the well's or the reservoir's physical constraints.

Early in our implementation we chose the Vicon™ 3D motion capture camera system [9] for tracking the 3D spatial layout of *Snakey*. With this system a number of retro-reflective markers are placed within a given tracking volume and an encompassing array of infrared emitters and cameras detect their 3D positions. The system uses triangulation algorithms to collectively reconstruct the positions of the markers in 3D space and in real-time. For the *Snakey* TUI, the discreet 3D positions of the tracking markers were then interpolated to yield the continuous 3D curve path.

The Vicon 3D camera system provided a large interaction volume for the *Snakey* device, adaptability (e.g. multi-prong curves), and practically unlimited curve lengths within our task by adding more markers. The visual tracking system introduced the problem of temporary occlusions such as when users cover a marker with their hands during manual manipulation, a problem we partly solved through the use of smoothing and interpolation algorithms. However, occlusions remain one of the *Snakey* prototype's reoccurring problems.

The TUI is based on Joby's GorillaPod™ [10]: flexible camera tripods which consist of a set of modular interlocking plastic orbs (a series of ball and socket joints) which are exceptionally light weight and are both easily manipulated and able to hold rigid shapes even when under significant loads. The *Snakey* prototype, using the mid-range *GorillaPod SLR-Zoom* model, can form curves with a minimum bend radius of 3.5cm and when used to form a single, continuous chain it can support the weight of approximately 25 additional orbs arranged in a horizontal line. (Fig. 2)

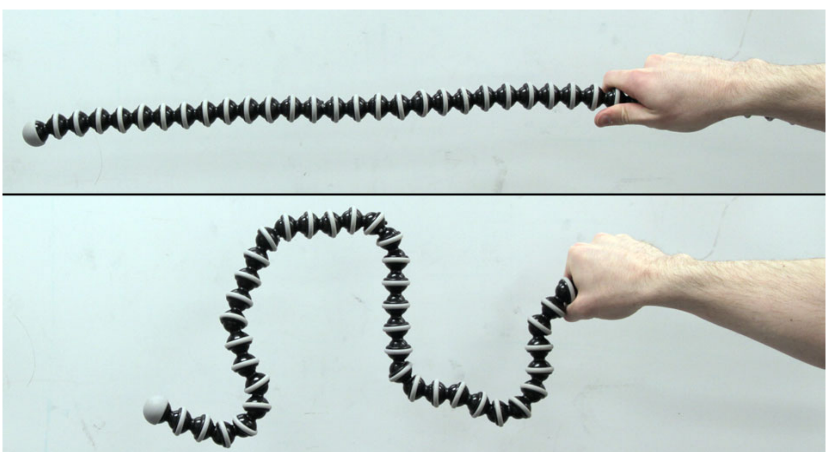


Fig. 2. The "GorillaPod" rigidity, even when heavily cantilevered (top), and its ability to be easily manipulated into complex shapes (bottom)

In terms of modularity, sections of the orb links can be manually pulled apart with sufficient force and then easily popped/pushed back together. Combined with special branching nodes, this provided a quick and easy "building blocks" style method of creating multi-lateral/branching well path formations; an important component of most reservoir drilling operations. (Fig. 3) While the GorillaPod links were physically capable of forming "loops" and other more complex formations, our applied task of well path exploration precludes our need to consider them (as they do not make practical sense for the current *Snakey* task) and as such were not explored in detail.

Additional visual and haptic interactive modalities (in parallel with the on-screen display) were also directly integrated into the physical *Snakey* device. Visual feedback was provided by a set of programmable tri-colour, wide-angle LED lights. Outputting approximately 8000 millicandelas of light intensity (sufficiently bright that they can be painful to look at directly at full power), the LEDs could each be individually programmed to take on a range of colours and intensities using a daisy-chained, two-wire communication bus. Haptic feedback was provided by a set of 10mm diameter, button-style vibration motors, each controlled in parallel (along with the LED lights) by a central microcontroller. Both the LED lights and vibration motors were small enough to be mounted directly onto individual GorillaPod orbs without interfering with *Snakey*'s flexibility of motion. (Fig. 1)

Snakey was mounted to a rigid, reconfigurable aluminum frame allowing the "fixed end" of the *Snakey* chain to be position above, below, or to the side of the working volume, as necessary. (Fig. 1) This approach provides a fixed 3D position in space by which to correlate the physical object with its virtual, on-screen counterpart. It also helps create an open workspace around the physical object wherein multiple collaborating users can interact with, and manipulate the *Snakey* device simultaneously.

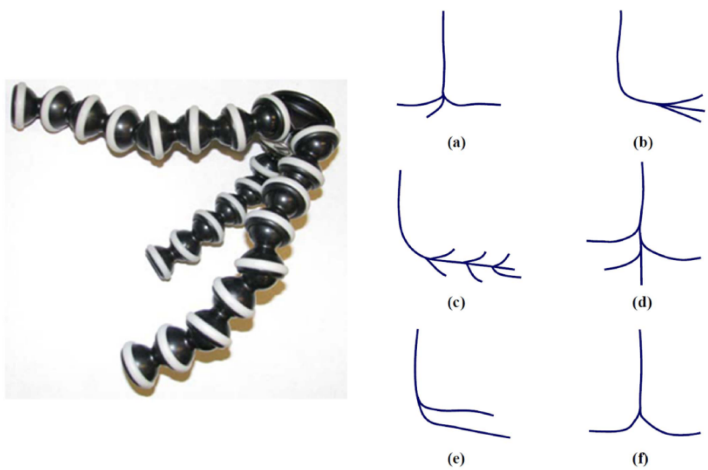


Fig. 3. (Left) A branched configuration of the *Snakey* object. (Right) Common multilateral well path configurations (a) Multibranched, (b) Forked, (c) Laterals into horizontal hole; (d) Laterals into vertical hole; (e) Stacked laterals; (f) Dual-opposing laterals.

Snakey's mounting frame was positioned directly in front of a large, high resolution display which, with the help of the 3D position data provided by the Vicon tracking system, superimposes the reconstructed virtual 3D curve onto the 3D reservoir model (Fig. 1). The displayed virtual model describes both the geology and flow behavior as simulated by high-end systems used in industry, e.g. [11]. As the user manipulates the *Snakey* device in the physical space, its virtual representation updates within the reservoir model in real time. This facilitates a direct input/output mapping between the *Snakey* object's physical shape and its virtual representation. Users can explore the virtual reservoir model immediately and easily by grasping and manipulating the physical *Snakey* device. The interface's physical constraints are expected to translate directly into similar constraints for the virtual model: for example, limited bend radius and avoidance of self-intersection are inherent restrictions both in real-world well paths scenarios, as well as in the *Snakey* TUI.

During interaction the *Snakey* TUI can augment the visual display with further real-time feedback which is physically embedded into the interface. For example by indicating through the rumble intensity of the attached vibration motors the different fluid flow rates in different sections of the well, or by indicating that the well path intersects with a dangerous terrain volume using glow of a section of the integrated LED lights.

4 Preliminary Evaluation of *Snakey*

The design of the *Snakey* prototype was participatory, with close involvement of reservoir engineers and domain experts in all the design iterations. In this section we include the findings, criticism and limitations of *Snakey*, as reflected via a design critique session which was conducted with the help of two experienced reservoir engineers. Our preliminary evaluation revealed that the experts were inspired by the concept of having a graspable physical manifestation of their three-dimensional well path plan. They highlighted that *Snakey*, once we tackle some of its current limitations, can signify a major step forward in terms of usability and interaction with 3D well paths. Below, we provide a collection of important observations gathered from our users. Additionally, we feel that the abundance of ideas and enthusiasm which our users provided reflects the impact that a more refined interface could have on the well path planning process.

While the users were able to easily view the physical *Snakey* device from any angle simply by walking around it, they were unhappy with the current need to rotate and move the geological data on the visual display using a desktop mouse and keyboard, taking attention away from the TUI and introducing a critical disconnect between the mental models of the physical device and the on-screen representation of the virtual data. In order to mitigate the physical inconvenience of having to shift between the *Snakey* device and a desktop mouse, it was proposed that a viewpoint controller would be mounted to the frame holding the *Snakey* device.

Our expert users also asked for added functionality that would allow them to control interface attributes beyond the 3D well curvature and path. For example, indicating where a specific operation takes place along the 3D well by physically squeezing a region of the *Snakey* chain, or sliding one's hand along a region of the TUI to mark an area in the geological data.

Another major design feature that was requested is the ability to store the shape of the physical device and to facilitate comparison of stored alternative well paths. Once the shape had been stored, users would have liked to physically "load" the well shape back onto *Snakey*. Supporting such a task may be accomplished if a future *Snakey* becomes mechanically actuated, in essence turning it into a robotic interface capable of taking on its own formation (e.g. [12]). Simpler physical synchronization schemes were proposed, requiring the user to be actively involved in the physical well "loading" operation, for example by providing indications of when the physical device has approximately matched its digital parallel by showing green (match) or varying shades of yellow-orange-red (no match) using Snakey's sectional LED lights.

Another finding relates to *Snakey's* physical scale limitation. In practice, wells can be many hundreds of meters long, with each and every part of their huge extent often being carefully designed. Currently, the physical *Snakey* affords only a one-to-one mapping to the entire length of the well. However, because the distance between adjacent GorillaPod orb joints is fixed (i.e. each link is a solid piece of plastic), users cannot specify curves any smaller than a single link. A possible solution is to support the design of wells in various stages of scale and granularity. The first stage would map the interface to the well as a whole singular entity; establishing the well's basic position and shape. The next stage would map *Snakey* to only a portion of the complete well path, allowing a subsection of the well to be designed with higher precision. This "zooming in and out" approach revisits the aforementioned need for the physical interface to match previous, stored curves.

The experts were very positive regarding the inclusion of the LED lights and vibration motors as additional multimodal feedback integrated into *Snakey*. They considered scenarios such as having the visual display show pipeline curvature tolerances while the vibration motors indicate fluid flow and the LED lights display rock densities. Following, *Snakey* could arguably allow the user to perceive a fusion of properties that are interrelated, and would have been difficult to visualize simultaneously using a single, traditional display.

The reservoir engineers were also particularly enthused about the inherently collaborative nature of *Snakey*. Any user within a team, regardless of professional training or project role, could instantly step in and manipulate the physical curve, provide meaningful input, and have the result be immediately visible and apparent to everyone involved.

The experts also proposed additions to the capabilities of the virtual screen model that took advantage of the unique *Snakey* interface characteristics but would otherwise be difficult to control/express using a traditional WIMP interface. For example, while the resulting 3D curve can show one-dimensional geological information along its length (both virtually on-screen and physically using the attached LED lights), it was proposed that a 2D cross-section view of the virtual data would also reveal more information. This could be accomplished by treating the planar projection of the *Snakey* chain's 3D form as a curtain that would "cut" the reservoir voxel data at the location of the curve, providing a richer visualization that shows the internal, "under the surface" data to the user. Alternative visualization schemes might include a translucent "tunnel" of surrounding voxel information that is shown within a set radius of the 3D curve.

Finally, the experts saw the immediacy and intuitiveness of the *Snakey* interface as an opportunity to employ more scalable reservoir simulation solutions and provide much faster design feedback than is currently found industry. For example, instead of using traditional CAD software to simulate a complete well path using extremely detail physical parameters and finite-element methods (a process which can take days to compute), a proposed *Snakey* well path could be simulated at a more general level with fewer parameters, having results quickly visualized and reviewed by the design group and allowing *Snakey* TUI to be immediately adjusted for another simulation cycle. With this, engineers can interactively "zero in" on better well path designs in a matter of seconds before investing the time to run a traditional full-scale simulation.

5 Conclusion

We presented *Snakey*, a tangible user interface supporting physical manipulation and planning of 3D well paths. Our current *Snakey* prototype allows users to interact with and manipulate a physical curve-like artifact which is tracked in real-time and spatially mapped to a virtual 3D curve. In turn, this allows users to intuitively explore complex 3D geological data and collaborate in well path planning tasks with a multi-disciplinary team. *Snakey* affords flexible physical manipulation and its main advantage is its ability to maintain its shape even when not being held by a user. Other features explored in the *Snakey* TUI are integrated haptic feedback and dynamic visual cues within the physical interface itself, as well as dynamically changing *Snakey's* topology as afforded by its modular nature. Preliminary evaluation emphasized many of the prototype's limitations, outlined some short-term goals for improvement and also highlighted the promise *Snakey*-like TUIs hold for the domain of reservoir geosciences and engineering, and the task of 3D well path planning.

Acknowledgements. We would like to thank our colleagues and anonymous reviewers for their valuable input. This research was supported by the Alberta Innovates Academy (AITF) / Foundation CMG Industrial Research Chair in Scalable Reservoir Visualization.

References

1. Lidal, E.M., Langeland, T., Giertsen, C., Grimsgaard, J., Helland, R.: A Decade of Increased Oil Recovery in Virtual Reality. IEEE Computer Graphics and Applications (2007)
2. Loftin, R.B., Bavinger, B.A., LeRoy, S.D., Nelson Jr., H.R.: Advanced Visualization Techniques for Exploration and Production. In: Offshore Technology Conference (1997)
3. Ullmer, B., Ishii, H.: Emerging Frameworks for Tangible User Interfaces. In: Carroll, J.M. (ed.) Human-Computer Interaction in the New Millenium, pp. 579–601 (2001)
4. Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., Itoh, Y.: On Tangible User Interfaces, Humans, and Spatiality. Personal Ubiquitous Computing, 338–346 (2004)
5. Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., Singh, K.: Exploring Interactive Curve and Surface Manipulationg Using a Bend and Twist Sensitive Input Strip. In: Proc. of the 1999 Symposium on Interactive 3D Graphics, Atlanta, Georgia, United States (April 26-29, 1999)
6. Anabuki, M., Ishii, H.: AR-Jig: A Handheld Tangible User Interface for Modification of 3D Digital Form via 2D Physical Curve. In: Proc. of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (2007)

7. Couture, N., Rivière, G., Reuter, P.: GeoTUI: a tangible user interface for geosciences. In: Proc. of the 2nd International Conference on Tangible and Embedded Interaction, TEI 2008, Bonn, Germany, February 18 - 20, pp. 89–96. ACM, New York (2008)
8. Tateosian, L.G., Mitasova, H., Harmon, B., Fogleman, B., Weaver, K., Harmon, R.S.: TanGeoMS: Tangible geospatial modeling system. IEEE Transactions on Visualization and Computer Graphics, TVCG (2010)
9. Vicon Motion Capture System, http://vicon.com/
10. Joby Gorilla Pods, http://joby.com/
11. Computer Modeling Group Ltd, http://www.cmgroup.com/
12. Ota, T., Degani, A., Schwartzman, D., Zubiate, B., McGarvey, J., Choset, H., Zenati, M.: A Highly Articulated Robotic Surgical System for Minimally Invasive Surgery. Annals of Thoracic Surgery 87(4), 1253–1256 (2009)

OP: A Novel Programming Model for Integrated Design and Prototyping of Mixed Objects

Céline Coutrix and Laurence Nigay

CNRS, Université Joseph Fourier Grenoble 1
Laboratoire d'Informatique de Grenoble LIG UMR 5217, Grenoble, F-38041, France
{Celine.Coutrix,Laurence.Nigay}@imag.fr

Abstract. In the context of mixed systems that seek to smoothly merge physical and digital worlds, designing and prototyping interaction involves physical and digital aspects of mixed objects. However, even though mixed objects are recurrent in the literature, none of the existing prototyping tools explicitly supports this object level. Moreover, designers have to use distinct tools, on the one hand, tools for designing ideas and on the other hand tools for prototyping them: this makes the design process difficult. To help alleviate these two problems, we present OP (Object Prototyping), a toolkit that provides a new programming model focusing on mixed objects and allows us to seamlessly go back and forth from conceptual ideas to functional physical prototypes, making the iterative design process smooth and integrated. Indeed, OP is explicitly based on an existing conceptual design model, namely the Mixed Interaction Model that has been shown to be useful for exploring the design space of mixed objects. Our user studies show that, despite its threshold, designers and developers using OP can rapidly prototype functional physical objects as part of a design process deeply intertwining conceptual design with prototyping activities.

Keywords: Prototyping, Toolkit, Mixed Systems, Mixed Objects, Augmented Reality, Physical User Interfaces, Tangible User Interfaces, Design.

1 Introduction

Mixed interactive systems seek to smoothly merge physical and digital worlds. Examples include tangible user interfaces, augmented reality, augmented virtuality, physical interfaces and embodied interfaces. In mixed systems, users interact with objects existing in both the physical and digital worlds. These objects are depicted in the literature as augmented objects, physical-digital objects or mixed objects.

Nowadays, designers either work alone or work with developers to design such objects. In the first case, except if they have coding skills, they cannot produce interactive mockups. Some tools provide partial contribution *towards* this problem, like Intuino [26]. In the second case, they can either provide a description to a developer who will then develop a prototype (however this significantly slows the design process) or they can also work together with a developer in designing and prototyping. This latter case is the work practice that we target in our study. In this

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 54–72, 2011.

context, facilitating the systematic exploration of design solutions is a very important problem for designer-developer pairs, as their aim is to generate solutions in the exploration phase [4][16].

In this paper, we introduce OP - Object Prototyping - a toolkit for the prototyping of such objects involved in mixed systems. OP offers two main benefits for designer-developer pairs: (1) it offers a new programming model, based on interacting objects and (2) it provides an integrated design approach, thanks to its strong link with an interaction model that was proven useful for exploring the design space. These two key aspects of OP are motivated by the following problems.

First, several studies of mixed systems developed in the literature [25][7][12] underline that the concept of a mixed object is central and recurrent in the design of mixed systems. A mixed object can be either used as a tool by the user to complete her task like the paper button “fill” in the drawing scenario of the seminal Digital Desk [27], or as the object focus of a task like the drawn house whose roof is to be filled in the drawing scenario of the Digital Desk. Unless they explicitly focus on objects in mixed systems, prototyping toolkits would not provide an optimal abstraction level for better design flexibility. Moreover reuse could be improved, as a mixed object can be an elementary block reused in different applications, like the pucks of [21], a generic mixed object.

Second, as the aim of early design is to explore as many ideas as possible [16][4], researchers and practitioners have proposed: (1) Interaction models, *e.g.* [14][7][9], in order to help to systematically explore the design space, (2) Prototyping toolkits, *e.g.* [13][11], in order to help materialize design ideas and support active thinking.

The use of both types of tools has been shown as being effective [7][11]. Moreover conceptual design and prototyping activities are inextricable [16]. Both activities inform each other: designer-developer pairs are going back and forth from conceptual ideas to practical realizations. An optimal practice of intertwined conceptual design and prototyping would be to explore the design space based on a model while prototyping the design alternatives. However, interaction models and prototyping toolkits are hardly used together. These two types of tools mainly remain unrelated, making their simultaneous use difficult. By using either one of these types of tools, designer-developer pairs therefore end up with either (a) a good covering of the design space, but cannot experience their design ideas with prototypes, or (b) prototypes that they can reflect on or show to stakeholders but cannot be sure that they explored the design space in a satisfactory way and might have forgotten to consider interesting solutions. As pointed in [2], after the presentation of an interaction model for GUI, “operationalizing the design of interaction requires appropriate tools and frameworks.”

To help alleviate both of the two above problems in the design of mixed physical-digital systems, OP is directly based on a conceptual interaction model that focuses on mixed objects. As OP tightly integrates the widely used Qt [23], the Mixed Interaction Model and existing hardware tools to offer a structuring of the prototype, it allows designer-developer pairs to easily go back and forth from ideas to prototypes of a mixed object.

The remainder of this paper is organized as follows: we begin by recalling the key elements of the Mixed Interaction Model, that is the background of this work. We then explain to which extent existing tools address this practice of intertwined conceptual design and prototyping of mixed objects. We next present the OP prototyping toolkit and we end by describing experiences of designer-developer pairs using our tool.

2 Background: The Mixed Interaction Model

The Mixed Interaction Model (MIM) is a conceptual design model that focuses on hybrid physical-digital objects or mixed objects. Some interaction models for GUI [2] also identified this intermediary level between hardware resources and applications that is called the object level. The key contribution of MIM is the structuring of a mixed object: MIM defines building blocks of a mixed object at a higher level of abstraction than an encapsulation of a piece of hardware. The MIM structuring of a mixed object has been shown to be useful for exploring the design space of mixed objects in a systematic way [7].

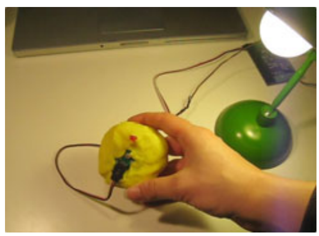
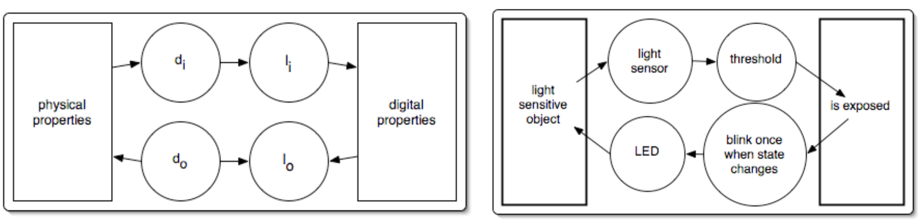


Fig. 1. A mixed physical-digital object sensitive to light that has been built with OP



(a) (b)

Fig. 2. (a) Model of a mixed object. (b) Modeling of a simple light sensitive object (Fig.1).

Based on MIM, a mixed object is defined by its physical and digital properties as well as the link between these two sets of properties. The link between the physical and the digital parts of an object is defined by linking modalities. The definition of a linking modality is based on that of an interaction modality [3]: Given that d is a physical device that acquires or delivers information, and l is an interaction language that defines a set of well-formed expressions that convey meaning, an interaction modality is a pair *(d,l)*, such as *(camera, computer vision)* or *(microphone, pseudo natural language)*. These two levels of abstraction, device and language, are reused. But as opposed to interaction modalities used by the user to interact with mixed environments, the modalities that define the link between physical and digital properties of an object are of lower abstraction level and therefore called *linking modalities*. Fig. 2a shows the two types of linking modalities that compose a mixed object: An input linking modality (d_i,l_i) is responsible for (1) acquiring a subset of *physical properties*, using a device d_i (input device), (2) interpreting these acquired physical data in terms of *digital properties*, using a language l_i (input language). An output linking modality is in charge of (1) generating data based on the set of *digital properties*, using a language l_o (output language), (2) translating these generated physical data into perceivable *physical properties* thanks to a device d_o (output device).

To illustrate the different parts that compose a mixed object, the modeling of a simple object, a light sensitive object shown in Fig. 1, is provided in Fig. 2b. It embeds a light sensor and a LED on its surface. Each time the user sets the object too close to a light source, the LED blinks once. Then, if the user moves the object far enough from the light source, the LED blinks once again. Fig. 2b shows the modeling of this object: The light sensor (input linking device) captures the level of exposure. If this level is above a threshold (input linking language), then the digital property, `is exposed`, is set to true. If the level falls under the threshold, the digital property is set to false. Each time the state of the digital property changes, an output linking language `beep` is activated. This output linking language triggers the LED to blink once.

3 Related Work

The existing toolkits for mixed systems could be analyzed based on their type of underlying language (compiled vs. interpreted, textual vs. graphical, etc.), their threshold and ceiling [18], their popularity in use, etc. We take a different viewpoint on existing tools by analyzing (1) how these tools support the structuring of the prototype based on a mixed object and (2) to which extent they support a systematic exploration of the design space by relying on the ability of an interaction model to help designers create new designs [2].

3.1 The Mixed Object Programming Model

Phidgets [22], Arduino (http://arduino.cc/), ARToolKit [1], Intuino [26], BOXES [13], MaxMSP [17] and PureData (http://puredata.info/) do not imply a code structuring based on a mixed object in their resulting code. For instance, Fig. 3 shows a Phidgets C code for the light-sensitive object of Fig. 1 and 2b. Compared with the MIM modeling of Fig. 2b, the code corresponding to the linking input language is scattered at lines 3, 4, 6, 8 and 10, interwoven with code for input and output devices and the output language. As a consequence, it is very difficult to localize all the lines of code corresponding to a particular element of a mixed object (principle of separation of concerns), as identified in the MIM modeling of Fig. 2b. Arduino has a similar approach with setup() and loop() functions resulting in a code structured in a different way. ARToolKit offers a similar loose structuring of the code, resolving difficulties of programming with camera and computer vision techniques, and not difficulties of programming with sensors.

```
#include <phidget21.h>
   CPhidgetInterfaceKitHandle Kit;
   int OldValue = 0;
   int SensorChangeHandler(CPhidgetInterfaceKitHandle h, void *p, int i, int v)
   {
    if ((v < 500 && OldValue >= 500) || (v > 500 && OldValue <= 500)
    {CPhidgetInterfaceKit_setOutputState(Kit, 0, 1);}
    else
    {CPhidgetInterfaceKit_setOutputState(Kit, 0, 0);}
    OldValue = v;
    return 0;
   }
   int main (int argc, char * const argv[]) {
    CPhidgetInterfaceKit_create(&Kit);
    CPhidgetInterfaceKit_set_OnSensorChange_Handler (Kit, SensorChangeHandler,
   NULL);
    CPhidget_open((CPhidgetHandle)Kit, -1);
    int result;
    const char* err;
    if((result = CPhidget_waitForAttachment((CPhidgetHandle)Kit, 10000)))
    {CPhidget_getErrorDescription(result, &err);
     printf("Problem waiting for attachment: %s\n", err);}
    while (true) {}
   }
```

Fig. 3. Phidgets C code of the light sensitive object presented in Fig. 1 and modeled in Fig. 2b

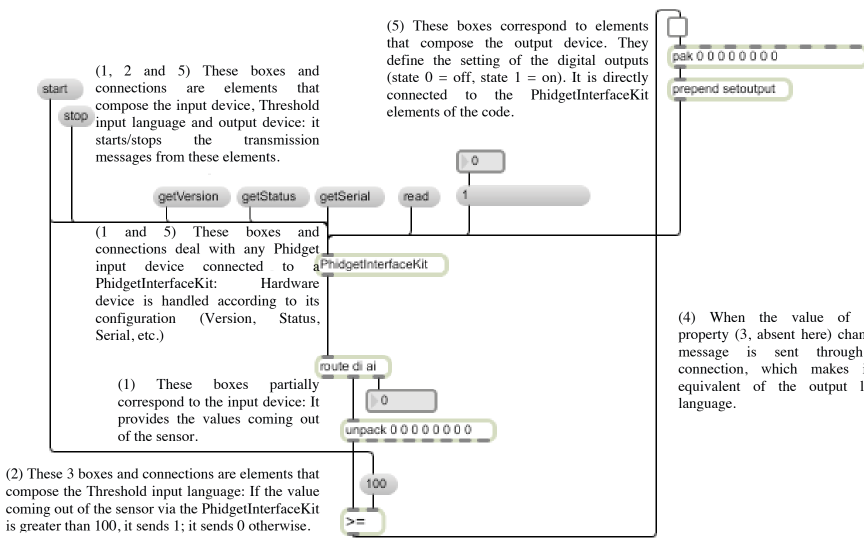


Fig. 4. MAX/MSP patch of the light sensitive object presented in Fig. 1 and modeled in Fig. 2b

Unlike Phidgets, well-known in the research community, MaxMSP is on the contrary widely used in the art/design community. Like its open-source counterpart PureData, it is a graphical programming language (example in Fig. 4). However, as such, it does not necessarily guide the resulting prototype to follow a particular structure. The patch shown in Fig. 4 is one code out of numerous possibilities for the light sensitive mixed object (Fig. 1). We designed this solution to be as simple as possible but also as close as possible to the MIM modeling of Fig. 2b. However we observe that the portions of the code corresponding to the linking input device (1 in Fig. 4), the linking input language (2 in Fig. 4) and the linking output device (5 in Fig. 4) are mixed together. Moreover, while some elements of a MIM modeling, like the ones mentioned above, are composed of boxes (objects: boxes with borders, messages: boxes without borders) and connectors, others like the linking output language (4 in Fig. 4) is composed of a single connector, therefore making it difficult to visually localize the blocks composing the code in regard to a MIM modeling.

ICON [8], ICARE [3], OpenInterface [20] and Context toolkit [24] identify the object level, but at a higher level of abstraction that does not allow its design. Indeed, with those toolkits, the object is a predefined brick used in the design of an entire application. Moreover, they mostly focus on interaction from the user to the application and hardly consider the feedback towards the user.

Papier-mâché [15] and d.tools [11] are the closest to the object programming model. On the one hand, papier-mâché allows the designer to manipulate the physical part (called Phob) and the digital part (called noun or verb), but does not use a detailed definition of the association between those two parts. Moreover, it allows connecting a physical part to "noun", equivalent to digital properties, but also to "verb" equivalent to commands. The command associated with a tool is application-dependant and is at a higher level of abstraction than the object level. On the other hand, d.tools implies objects loosely structured according to physically stuck elements rather than conceptual objects. This approach can possibly combine an object that is a tool with a task object, as a unique entity. Instead of focusing on the design of mixed objects, it considers the design of a mixed system as a whole.

We now examine to which extent existing tools support a systematic exploration of the design space and are related to a conceptual model.

3.2 The Systematic Exploration as a Design Approach

Several toolkits have no explicit link to an interaction model. Among those, Phidgets [22], Arduino, ARToolKit [1], MaxMSP [17], PureData, Context toolkit [24] and ICON [8] are based on existing developers' practices and solve technological problems, not design problems. They either provide textual coding (Phidgets, Arduino, ARToolKit, Context toolkit) or graphical user interface (GUI) for coding (MaxMSP, PureData, ICON) but do not promote exploration of alternatives according to an interaction model. For instance, ARToolKit and Phidgets use callback functions for hardware resources. As a consequence, coding/development and design practices are not close enough to allow designers to easily go back and forth from design to prototyping with these toolkits without restructuring their ideas to fit the programming model.

BOXES [13], d.tools [11] and Intuino [26] are not based on developers' practices, but on non-instrumented design practices. Intuino provides a GUI to better fine-tune sensors and actuators signals used with Arduino. The tool BOXES provides a direct link between physical material and an existing GUI. As opposed to BOXES, the toolkit d.tools provides indirect prototyping of the interaction through state transition diagrams, where states define an on-screen graphical representation of the connected tangible object to be prototyped. However, it does not encourage the systematic model-based exploration of alternatives other than the opportunistic ideas that a designer might have.

Papier-mâché [15], ICARE [3] and OpenInterface [20] toolkits have an explicit link to existing interaction models. They support an exploratory design process by proposing a structuring of the code, to be entered either textually like Papier-mâché or graphically like ICARE and OpenInterface. For instance, Papier-mâché introduces associations between physical Phobs and digital elements based on name (object) or verb (command) and as such the tool can be closely related to Fishkin's taxonomy of tangible user interfaces (TUI) [10]. ICARE and OpenInterface are related to a model of multimodal interaction [3]. However, the conceptual models behind these toolkits are limited to explore the design space of mixed systems [7]. For instance, the concepts of *name* and *verb* convey respectively the concepts of *task object* and *tool* [9] and correspond to two different types of metaphors that can be applied [10], but leave apart all the possibilities for linking modalities.

Table 1. Related works summary

Underlying interaction model \ Object level	No	Partial	Yes
No	Phidgets, Arduino, ARToolKit, MaxMSP, PureData, Intuino, BOXES	Context toolkit, ICON, d.tools	
Partial		Papier-mâché, ICARE, OpenInterface	
Yes			

Table 1 shows a summary of existing toolkits according to our two axes of analysis: (1) prototyping at the object level and (2) systematic exploration of the design space based on an interaction model.

In order to both support an object programming model as well as a systematic exploration of the design space, we introduce the OP toolkit that is based on the Mixed Interaction Model and capitalizes on these existing tools when possible. By adopting this approach, our purpose is first not to reduce the technological difficulties encountered when building mixed objects. As explained previously, there are toolkits that answer these problems such as computer vision toolkits (ARToolKit, Papier-Mâché) or hardware toolkits (Phidgets, Arduino). Our toolkit has to be built upon them and be able to integrate the ones resolving technological challenges. OP consequently provides lower thresholds than low-level toolkits on top of which OP is built on. Second the OP toolkit focuses on mixed objects only. OP is not intended as a tool to build an entire application but only mixed objects that are the focus of the design process, as opposed to d.tools, ICON, ICARE and OpenInterface. The OP toolkit could then be included in such tools.

4 OP Toolkit

The OP toolkit offers an extensible library of different types of components based on the Mixed Interaction Model [7]. The toolkit includes around 8000 lines of code providing input/output linking device components, input/output linking language components, composition components for combining devices or languages, as well as digital property components. By doing so, the OP toolkit provides: (1) Modularity at the mixed object level in order to make the interface *flexible* and to make mixed objects *reusable* for other interaction contexts; (2) Modularity at the linking modality level – modules for input and output devices, languages, and compositions – in order to make mixed objects *flexible* and to make linking modality components *reusable* for other mixed objects; (3) Extensibility in order to make the toolkit itself *flexible*. The developer should be able to add new building blocks and extend to new technologies. In its current version, OP still requires some basic notions of programming for assembling the various parts of a mixed object and for connecting objects to applications. Thus, as explained in the introduction section, the target end-users are designer-developer pairs, developers or designers with some programming skills.

4.1 Building Mixed Objects

OP currently includes the following components for building mixed objects.

The linking device components are based on three different existing toolkits:

- `VideoInputDevice`, based on the ARToolKit: It captures the video input. Its outputs are images from the input video.
- `MIDIDevice`: This component corresponds to either an input or an output device, and captures/delivers data from/to Interface-Z MIDI sensors/actuators (http://www.interface-z.com/).
- `PhidgetInterfaceKitDevice`: Either an input or an output device, this component captures/delivers data from/to sensors/actuators plugged to a Phidget Interface Kit.

Commonly used devices like speakers or screens are directly supported by the toolkit by linking language components, since computers already support them. So no device components are provided for these standard devices in the current version.

For linking languages, OP offers 10 components. Components amongst them are:

- `IdentityLanguage`: This component does not deform the input values coming from the sensor. A property of the component allows opposing of the output values to the input values. For example, if a sensor provides values between 0 and 999, the output values are then from 999 to 0.
- `RampLanguage`: This component generalizes the `IdentityLanguage` component and implements a deformation of the input values according to a ramp function (see Fig. 5a).
- `ThresholdLanguage`: This component delivers a boolean value true/false if the input integer value is above the threshold and false/true otherwise. Properties of the component are the threshold and a property specifying whether the output value is true above or below the threshold.
- `RepeatLanguage`: Either an input or an output language, this component repeats its inputs several times at specified intervals. Properties are the number of repeats and the interval between repeats.
- `BeepOutputLanguage`: This component corresponds to an impulse function that triggers a beep if used with a sound file or a blink if connected to a LED.

The other five linking language components are documented at http://iihm.imag.fr/demo/op/. In order to prototype more complex linking languages, these linking language components can be connected in series. For example, for making a LED blink twice, we connect a `RepeatLanguage` to a `BeepOutputLanguage`. Moreover for combining linking modalities, we developed a `Complementarity` component that can combine data coming either from device or language components. For example, for combining the data coming from two accelerometers (Fig. 9g), we use the `Complementarity` component. This component draws directly upon ICARE components [3] for fusion of interaction modalities in multimodal interaction.

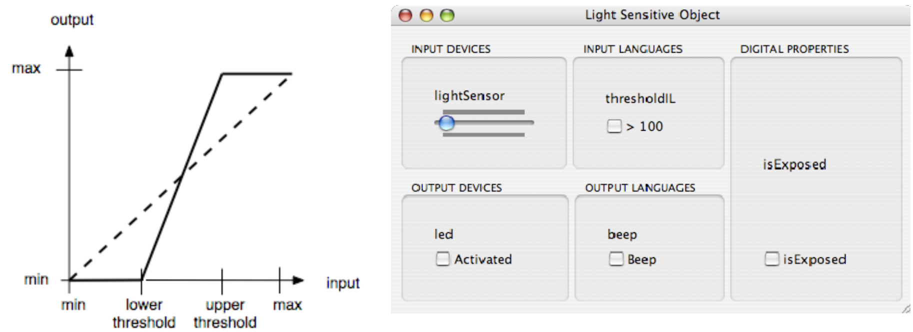
For digital properties, OP includes a `DigitalProperty` component that can handle any type of digital property, based on the generic type QVariant from Qt [23].

All components, regardless of type, can be connected through the signal/slot mechanism of Qt [23]: A component output is a signal, and a component input is a slot. We connect an input to an output thanks to a line of code: `connect(Component1, SIGNAL, Component2, SLOT)`. OP components have predefined `updated` signals emitted each time its output value is modified, and `update` slots to be connected to the former signals.

Fig. 5c illustrates the use of the library for prototyping the simple light sensitive object of Fig. 1. Using OP components described above, the developers/designers can prototype the object described in Fig. 2b with a few lines of code (Fig. 5c). We use a light sensor from Phidgets in line 3. We set the properties of this component: its name is “lightSensor”, its direction is “in”, it is plugged to the first input of the circuit board, and it is an analogue sensor (on the contrary to switches for example). According to the modeling of the object (Fig. 2b), we then use in line 6 a threshold component that outputs true when the value is above 100. Line 7 shows how we declare the `isExposed` digital property, with a boolean type. A `BeepOutputLanguage` is used in line 9, without any sound file. Finally a LED connected to the first digital output on the same Phidgets interface kit is declared at line 11. From line 12 to line 19, we connect these five components together, so that outputs of each component provide inputs to the following one. Compared to Phidgets (Fig. 3) and MaxMSP (Fig. 4), the code structure follows the MIM modeling. From this code of a mixed object, it is therefore easy to make modifications. For example:

- Modification of the output behavior of the object: We can make the LED blink four times in order to provide insistence instead of bare observability. Fig. 6 shows how we added a `RepeatLanguage` component and connected it to the other components.
- Modification of the manipulation of the object: We can change the Phidgets light sensor by a pressure sensor from Interface-Z. Fig. 7 shows how we changed the Phidgets light sensor by an Interface-Z MIDI pressure sensor. In this latter case, there is no need to reconnect components, because we only changed the type of the device component, which has no influence on connections.

For fluid design as pointed for the Phidgets or in [16], the toolkit also provides an optional graphical user interface (GUI) to debug/test the prototyped mixed objects. The spatial structuring of the GUI is explicitly based on the MIM interaction model. Fig. 5b shows the graphical tool for the case of the light sensitive object of Fig. 2b. Boxes represent MIM component types. This interface helps the developer/designer to observe and control the details of the behavior of the object. To check at runtime the behavior of the components, developers either act on physical properties (e.g., put/move the object close/away to/from a light source) or act on widgets (e.g., move the sliders of Fig. 5b using the mouse). This graphical utility is also used as a wizard of oz tool, in the case of non available components. This enables rapid and early corrections of the prototype.



(a) (b)

```
int indexOnBoard = 0;
bool isAnalogue = true;
PhidgetInterfaceKitDevice lightSensor("lightSensor", LinkingComponent::IN,
indexOnBoard, isAnalogue);

int threshold = 100;
bool isTrueAbove = true;
ThresholdInputLanguage thresholdIL("thresholdIL", threshold, isTrueAbove);

DigitalProperty isExposed("isExposed", QVariant::Bool);

char* soundFile = "";
BeepOutputLanguage beep("beep", soundFile);

isAnalogue = false;
PhidgetInterfaceKitDevice led("led", LinkingComponent::OUT, indexOnBoard,
isAnalogue);

QObject::connect(&lightSensor, SIGNAL(updated(int, QTime)), &thresholdIL,
SLOT(update(int, QTime)));
QObject::connect(&thresholdIL, SIGNAL(updated(QVariant, QTime)), &isExposed,
SLOT(updateProperty(QVariant)));
QObject::connect(&isExposed, SIGNAL(PropertyUpdated(QVariant)), &beep,
SLOT(update(void)));
QObject::connect(&beep, SIGNAL(updated(bool)), &led, SLOT(UpdateOutput(bool)));
```

(c)

Fig. 5. (a) Two input linking language components: Identity (dashed line) and Ramp (plain line). (b) Graphical User Interface. (c) Code with inserted picture of a simple object built using OP.

```
int repeatsNb = 4;
float interval = 0.5;
RepeatLanguage repeat(“repeat”, LinkingComponent::OUT, repeatsNb, interval);
QObject::connect(&isExposed, SIGNAL(PropertyUpdated(QVariant)),
                 &repeat, SLOT(update(QVariant)));
QObject::connect(&repeat, SIGNAL(updated(bool)),
                 &beep, SLOT(update(void)));
```

Fig. 6. Making the LED blink four times by adding a `RepeatLanguage` component

```
int indexOnBoard = 0;
bool isAnalogue = true;
int resolution = 7;
MIDIDevice pressureSensor("PressureSensor", LinkingComponent::IN, indexOnBoard,
isAnalogue, resolution);
```

Fig. 7. Changing the light sensor by a pressure sensor: the first lines of code are modified

In this section, we explained how we can build mixed objects using OP components. We now present how we insert such objects into a complete application. We recall that a mixed object in the context of a complete application is either a tool used by the user to perform her/his task or the object that is the focus of the task [7].

4.2 Inserting Mixed Objects into an Application

The toolkit provides modularity at the mixed object level so that an object can be reused and adapted for various interaction contexts or applications. For a mixed object to communicate with an application, the `DigitalProperty` component serves as an interface with the application. The signals and slots of a digital property are connected to the rest of the application: the value of a digital property can be modified by an application through a slot and a digital property can communicate with an application by emitting a signal when its value changes. If the application is written in Qt, connecting the object to the application is straightforward. We only need to write one line of code connecting the signal of the `DigitalProperty` component to the desired slot of the application. A survey in our lab showed that out of 70 responses, 19 persons are familiar with Qt. Therefore, even if Qt is actually used, it is obviously not the only solution in use. For an application not written in Qt, developers only need to define a simple QObject whose slot calls the desired function of the application. This slot must then be connected to the signal of the mixed object. For example, for inserting the tool object of Fig. 1 into Google Earth, we build a simple QObject with the slot presented in Fig. 8.

```
void KeySimulationInteractionLanguage::update(QVariant pMessage){
 if (pMessage.canConvert(QVariant::Bool)){
 if (pMessage.toBool()){ CGPostKeyboardEvent('d',(CGKeyCode)0x02, true);}
 else {CGPostKeyboardEvent('d', (CGKeyCode)0x02, false);}}}
```

Fig. 8. Slot written for the simple light sensitive object in order to work in Google Earth for rotating the earth. CGPostKeyboardEvent is a function from the Apple ApplicationServices framework that allows simulating keyboard events.

When this slot receives true, it simulates a 'd' key press, and when the slot receives false, it simulates a 'd' key release. For this, we use a function from the Apple ApplicationServices framework that allows simulating keyboard events. Equivalent exists for the other platforms. Once this `update` slot is connected to the signal `PropertyUpdated` of the `isExposed` digital property (Fig. 5c), the light sensitive object allows the user to rotate the earth, each time the object is close to a light source. As another example, Fig. 9h shows the cardboard music controller that was used as an illustration of the BOXES toolkit [13], that we rebuilt using OP: the OP cardboard music controller is connected to the iTunes application (play/pause by pressing the button) in the same way as described above. In this stage of our work, the link to an application is done by hand, but as soon as OP is integrated in a UIMS like the one described in [20], this link will be embedded within the integrated environment.

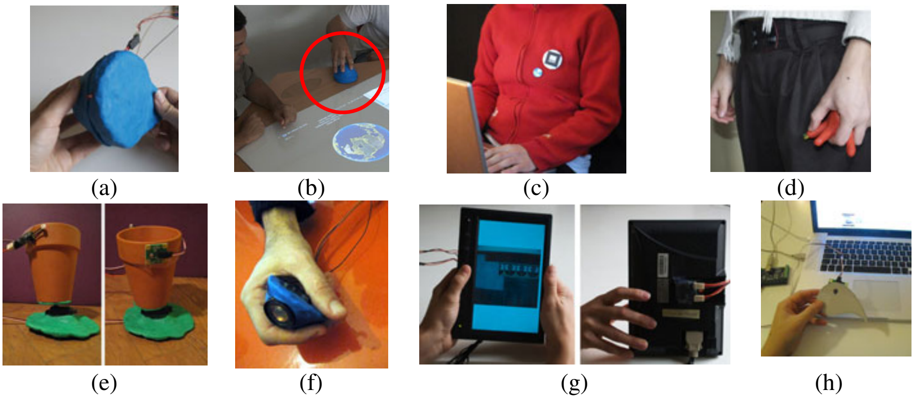


Fig. 9. A selection of mixed objects built with the OP toolkit: (a) Prototype of the ORBIS turnable tool with LED. (b) The ORBIS turnable tool in use on a table for rotating the earth in Google Earth. (c) Prototype of a simple augmented button badge with ARToolKit Marker that detects the presence of a laptop user. (d) “Octopus” tool with foldable tentacles, micro and diodes designed for recording audio [7]. (e) Flowerpot prototype that is oriented with a servomotor according to its most sunny side, thanks to interface-z and Phidgets light sensors. (f) Tool for shuffling pictures in a slideshow, with accelerometers and a loudspeaker: the user shakes it three times and hears a sound as a feedback. (g) Prototype of the ORBIS picture viewer. Pictures are always correctly displayed according to the orientation of the screen: the prototype is based on combined accelerometers fixed on the back side of the screen. (h) The remote control for the music player of [13] that we rebuilt with OP and iTunes.

Fig. 9 shows a variety of objects that have been prototyped using the OP toolkit, as a demonstration like in [22]. These examples cover a wide range of interaction styles that can take part in mixed systems, from augmented reality to augmented virtuality and tangible user interfaces. In the following section we present user studies, including the evaluation of the mixed object programming model, by detailing the case of one of these designed objects (Fig. 9a).

5 Evaluation and Use

Evaluating a software tool is a difficult problem as stated in [19]. As a consequence, the contributions of several software tools like MaxMSP, Arduino, PureData are not evaluated. Other tools, like Phidgets, do not follow requirements advised in [19] for evaluation.

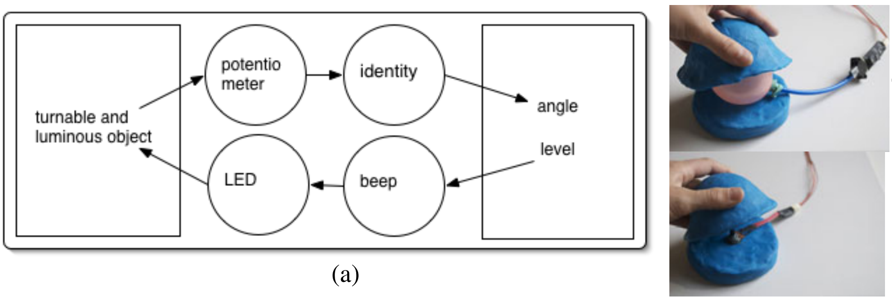
For the OP toolkit, its use in different projects (Fig. 9) is one form of empirical evaluation. Moreover, using OP, it is possible to prototype objects that were developed with other toolkits, like the cardboard music controller [13] (Fig. 9h). To better assess the benefits and limitations of OP, we considered its two key contributions independently. By doing so, the first study aims at evaluating the OP programming model centered on mixed objects. The second study aims at evaluating the expressive match of the toolkit in the context of an integrated use with the Mixed Interaction Model.

5.1 Evaluation of the Mixed Object Programming Model

The OP programming model is based on the definition of a mixed object of the Mixed Interaction Model. This model has shown its benefits for systematic exploration of design alternatives [7]. Therefore, we will not evaluate in this paper the help that the model provides to explore the design space. Nevertheless to further evaluate the OP programming model, we consider the practical problem of designing a physical-digital object that serves as a tool in an application called ORBIS [6]. Consistent with the target users of OP, a typical pair of designer-developer collaborated to design and prototype ORBIS. The system was designed in order to provide new ways to access and enjoy personal pictures, music and videos. ORBIS is to be used in a private, personal and mobile context. In the first version of ORBIS, we only considered tasks related to a list of pictures and Fig. 9g shows a prototype. In this paper, we present the design of a mixed object (i.e., a tool) for navigating the list of pictures in ORBIS. Besides prototyping and testing the OP toolkit, our aim was the joint exploration of design dimensions like appearance and interaction within a systematic approach at the early stage of the design. Before explaining the design steps using OP and the benefits of its programming model, we present the resulting designed prototype, a turnable tool presented in Fig. 9a and modeled in Fig. 10a. For navigating in the list of pictures, the user rotates the tool. The physical angle updates an `angle` digital property through a (potentiometer, identity) linking modality. Another digital property, `level`, is materialized through a (beep, LED) linking modality. Within the ORBIS application, the `angle` digital property is used to compute the index of the current displayed picture. When this index is changed by ORBIS, it modifies the `level` digital property of the tool as a feedback. Since ORBIS has been developed with Qt, the connection between the tool and the ORBIS list of pictures is straightforward. We now explain how the design space has been explored in a systematic way using OP, for obtaining this prototyped tool.

Physical Properties. The design space of the physical shape of the mixed object depends on the appearance (e.g., size, color) and interaction (e.g., affordances). For prototyping the object, we considered the interaction and the appearance in concert, but we only discuss the interaction here. Amongst others, we designed the shape presented in Fig. 9a. From this shape, we studied intrinsic affordances: we found that this shape afforded pressing and turning. Hence we decided that the possible sensed physical properties are pressure and angle.

Input Linking Modality. Because the shape affords pressing and rotating, we prototyped the tool with linking devices that sense angle and pressure. Prototypes with Interface-Z MIDI atmospheric pressure sensor mounted on a balloon and potentiometer are presented in Fig. 11b and c. Since these devices can be connected to the same plug of the Interface-Z circuit board, we did not need to modify the OP linking device component. We experimentally found that pressing was tiring more rapidly than rotating. This draws attention to the fact that even if the design space allows exploration of possibilities, prototyping is essential to assessing a design choice.



(a) (b) (c)

Fig. 10. Modeling of the ORBIS turnable tool presented in Fig. 9a (a) and prototyping the object with an input linking modality sensing (b) pressure or (c) rotation

The design space of the input linking language is restricted by the choice of the potentiometer and the digital property `angle`. Yet, we envisioned two possibilities: the linking language can take the full circle defined by the shape of the physical object into account. But the gesture for completely rotating the object is not easy to perform. On the contrary, the linking language can only take a subinterval of the full circle into account. We prototyped these two design solutions with different linking languages (Fig. 11): the first solution using an identity language component and the second one using a ramp language component. Tuning the language component using the provided OP components only implies 3 lines of code: the component declaration (the code of Fig. 11a versus the code of Fig. 11b) and the new connection since the name of the component is changed. The same change in an Arduino code costs more, because the programming is not at the object level (Fig. 12a). First, it is not as easy to locate the proper place in the code to insert the equivalent of the input language. Then, one has to add 11 lines of code including the hand-computed transformation (Fig. 12b). The abstraction level of OP is a benefit in this design situation.

(a)

```
IdentityInputLanguage identity("identity", isOpposite, min, max);
```

(b)

```
int lowerThreshold = 42;
int upperThreshold = 84;
RampInputLanguage ramp("ramp", isOpposite, min, lowerThreshold,
upperThreshold, max);
```

Fig. 11. Prototyping the ORBIS object with an input linking language: (a) identity or (b) ramp

While testing the two resulting OP prototypes, we noticed that even if the ramp language enables users to navigate the range in a small and single movement, it was not easily understood. Hence we chose the identity transformation, even if users have to make several movements in order to turn the tool from 0 to 360 degrees.

Output Linking Modality. In order to materialize the `level` digital property that provides feedback, we designed an output linking modality. We wanted this feedback to be peripheral since it defines the reaction of the tool, while the user focuses on the list of pictures. Towards this purpose, a visual or audio beep was considered as sufficient in comparison with displaying the index of the current picture on top of the picture. As a corresponding device for the chosen language, we could use a LED or a loudspeaker. We developed the two solutions (Fig. 13 and Fig. 14 respectively). The difference between the two pieces of code is that the loudspeaker is a standard device with no need to be declared as opposed to the LED, which is an OP device component to be declared and connected to the language component. After testing with both OP prototypes, we found that a LED was less obtrusive than a loudspeaker and chose the LED as an output linking device.

(a)

```
int level;                    // digital property level
int level_old;                // previous value of the property
int angle;                    // digital property level

void setup() {
  pinMode(13, OUTPUT);
}

void feedback() {
  digitalWrite(13, HIGH);  // set de LED on
  delay(1000);             // wait for a second
  digitalWrite(13, LOW);   // set the LED off
}

void loop() {
  int angle = analogRead(A0); // read the value of "angle" from the sensor

  if (level!=level_old) {     // if "level" changed
    feedback();
  }
}
```

(b)

```
  int lowerBound = 42;   // lower bound of the ramp transformation
  int upperBound = 600;  // upper bound of the ramp transformation
  int valueMax = 1023;   // maximum value from the sensor

  if (angle < lowerBound) {
  angle = 0;
  } else if (angle > upperBound) {
  angle = valueMax;
  } else {
  int a = valueMax / (upperBound - lowerBound);
  int b = -lowerBound * a;
  angle = a * angle + b;
}
```

Fig. 12. Prototyping the ORBIS object with Arduino (a) with an identity input linking language and (b) the lines of code to be added for using a ramp transformation

Through simple examples taken from our design experience in collaboration with a product designer, we illustrated the object level of the OP toolkit and its benefit towards flexibility of the evolving prototypes. Only a few lines of code, easily located, needed to be changed when ideas evolved. Moreover while collaborating with the designer, we clearly observed the usefulness of focusing on both physical appearance and interaction of the object at the same time - such joint activities being possible thanks to the OP toolkit.

```
DigitalProperty level("level", QVariant::Int);
char* soundFile = "";
BeepOutputLanguage beep("beep", soundFile);
QObject::connect(&level, SIGNAL(PropertyUpdated(QVariant)),
&beep, SLOT(update(void)));
isAnalogue = false;
PhidgetInterfaceKitDevice led("led", LinkingComponent::OUT, indexOnBoard,
isAnalogue);
QObject::connect(&beep, SIGNAL(updated(bool)),
                 &led, SLOT(UpdateOutput(bool)));
```

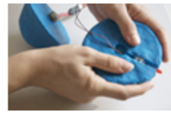


Fig. 13. ORBIS object: Prototyping the output linking modality with a red LED

```
DigitalProperty level("level", QVariant::Int);
char* soundFile = "./Pop.aiff";
BeepOutputLanguage beep("beep", soundFile);
QObject::connect(&level, SIGNAL(PropertyUpdated(QVariant)), &beep,
SLOT(update(void)));
```

Fig. 14. ORBIS object: Prototyping the output linking modality with a loudspeaker

The object level also promotes the reusability of objects: we actually reused the designed mixed object in another application context. We easily integrated this turnable object into Google Earth. Fig. 9b shows the prototyped tool in use for rotating the earth. This was the starting point for exploring design solutions of a tool for Google Earth. The mixed object must then be adapted and tuned for this new context of use. This shows the reusability at the mixed object level supported by the toolkit.

5.2 Evaluation of the Integrated Design Approach

Olsen [19] lists 9 possible claims of a user interface system. Three of them aim at *reducing solution viscosity* (reducing the effort to explore many possible design solutions): *flexibility* (rapid changes), *expressive leverage* (accomplish more by expressing less) or *expressive match* (closeness between the means for expressing design choices and the problem being solved). As our claim is the *expressive match* between the toolkit and an interaction model, thus targeting the reduction of *solution viscosity*, we chose to evaluate the toolkit regarding this claim. We therefore applied Olsen's framework for evaluation of *expressive match*:

- Explain the target *situations, tasks and users*,
- State the importance of the problem,
- Then demonstrate the expressive match (i) by measuring time to create a design or express a set of choices, *or* (ii) by challenging subjects to correct a design flaw and by reporting time, errors, difficulties and/or success rates. In this section, we report the second method of evaluation (ii).

The study gathered 4 participants who were developers but novices with the OP toolkit. Pairs of participants were asked to make two different randomly chosen modifications in the code of the ORBIS turnable tool (Fig. 10a). They first carried out the modifications of the output linking device of Fig. 13 and 15, followed by the modifications presented in Fig. 11. Modifications to be done were presented as MIM descriptions (Fig. 10a) without text. Making them work in pairs enabled them to talk more naturally. We provided the documentation of the toolkit for this exercise. We chose to ask them to perform these realistic modifications because we wanted to evaluate if they could figure out how to make a modification that can occur in real design settings. Indeed these modifications come from our prior experience with designers. After completing these two exercises by pairs, we conducted a discussion altogether about their difficulties as well as the identified benefits and drawbacks of the toolkit. During all the experiment, the participants were not given any help or information about the MIM modeling and the toolkit.

We identified one problem: they expected to find a linking device component in OP for the loudspeaker device (exercise 3 - modifications presented in Fig. 13 and 15). Indeed, it was unnatural for them to change a parameter in the linking language component when asked to change the linking device. This illustrates the relevance of our approach. Even though OP is an improvement towards expressive match between MIM and a prototyping tool, the toolkit should go even further and literally follow the MIM outline, bypassing existing software. The Identity language component is one contribution towards this goal, and we are already working on this identified problem related to standard devices (e.g., screen and loudspeaker).

Encouraging results also came up: During an exercise, participants had to write pseudo-code from a MIM modeling or a text description, and then explain an OP code. One of the participants who was given the MIM modeling of Fig. 2a drew a similar description for explaining the code in the second exercise. This shows that it can be straightforward to go back and forth from the interaction model to an OP code. Moreover after the experiment, one participant said that using OP was even not code writing for him, showing that he found it very easy working with the pair of tools. Finally they all suggested a graphical user interface for the toolkit. We actually plan on providing a tangible interface for OP as a counterbalance to on-screen prototyping.

6 Conclusion

To address the challenge of fluid design of mixed systems, we have presented the OP toolkit for prototyping mixed objects.

OP introduces the object level in prototyping mixed systems, a level not supported by existing toolkits. On the one hand, the OP toolkit is built on top of low-level technological toolkits, but still requires, in its current version, some basic notions of programming for assembling the parts of a mixed object. Its scope includes the scope of Phidgets [22] or BOXES [13] amongst others. On the other hand, the OP toolkit presents a high ceiling by enabling various mixed objects to be connected to applications. We are currently examining the integration of the OP mixed object library into User Interface Management Systems (UIMS) like the one described in [20]. Indeed OP provides support to prototype tools and task objects, so that a UIMS can use them as building blocks for the development of the entire application.

OP provides a set of components for rapidly building functional physical objects that are based on the Mixed Interaction Model [7], a conceptual design model that has been shown to be useful for exploring the design space of mixed objects. The tool allows its users to explore at the same time the physical forms of the object with various materials (foam, play dough, cardboard, etc.) as well as the interaction with the object via seamless conceptual reasoning and practical prototyping.

In the future, we would like to empower new users, namely designers with no programming skill but not necessarily working together with a developer. Toward this aim, we will be exploring a tangible interface for the toolkit. Indeed, prototyping is often participatory and tangible user interfaces suit group work, in contrast to on-screen interfaces. OP could provide tangible blocks for each abstraction level of the mixed object, to be embedded in a rapidly shaped physical prototype. In this way, re-arrangements and iteration directly on the physical prototypes could be done by each member of the group. In addition, a further open challenge we would like to address is in defining a tool based on the OP toolkit for letting the end-users define at runtime a mixed object by linking physical and digital properties. For example in [5] three types of coupling are defined: [personal, universal, transient] coupling between physical and digital properties of a mixed object that is defined dynamically.

Acknowledgments. We thank N. Mandran (LIG) for her support on evaluations.

References

1. ARTooKit, `http://www.hitl.washington.edu/artoolkit/`
2. Beaudoin-Lafon: Designing Interaction, not Interfaces. In: AVI 2004, pp. 15–22 (2004)
3. Bouchet, et al.: ICARE Software Components for Rapidly Developing Multimodal Interfaces. In: ICMI 2004, pp. 251–258 (2004)
4. Buxton: Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann Publishers Inc., San Francisco (2007)
5. Carvey, et al.: Rubber Shark as User Interface. In: CHI 2006, pp. 634–639 (2006)
6. Coutrix, Nigay: Balancing Physical and Digital Properties in Mixed Objects. In: AVI 2008, pp. 305–308 (2008)
7. Coutrix, Nigay: An Integrating Framework for Mixed Systems. In: The Engineering of Mixed Reality Systems Book, vol. 9, Springer, Heidelberg (2009)
8. Dragicevic, Fekete: Input Device Selection and Interaction Configuration with ICON. In: IHM-HCI 2001, pp. 543–448 (2001)
9. Dubois, Gray: A Design-Oriented Information-Flow Refinement of the ASUR Interaction Model. In: EHCI-HCSE-DSVIS 2007 (2007)
10. Fishkin: A taxonomy for and analysis of tangible interfaces. Personal Ubiquitous Computing 8(5), 347–358 (2004)
11. Hartmann, et al.: Reflective Physical Prototyping through Integrated Design, Test, and Analysis. In: UIST 2006, pp. 299–308 (2006)
12. Holmquist, et al.: Token-based access to digital information. In: HUC 1999, pp. 234–245 (1999)
13. Hudson, Mankoff: Rapid Construction of Functioning Physical Interfaces from cardboard, Thumbtack, Tin Foil and Masking tape. In: UIST 2006, pp. 289–298 (2006)
14. Jacob, et al.: Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In: CHI 2008, pp. 201–210 (2008)
15. Klemmer, et al.: Papier-Mâché: Tookit Support for Tangible Input. In: CHI 2004, pp. 399–406 (2004)

16. Lim, et al.: The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. ACM TOCHI 15(2)
17. MaxMSP, http://cycling74.com/
18. Myers, et al.: Past, Present and Future of User Interface Software Tools. ACM TOCHI 7(1)
19. Olsen: Evaluating user interface systems research. In: UIST 2007, pp. 251–258 (2007)
20. OpenInterface Platform, http://www.oi-project.org
21. Patten, Ishii: Mechanical Constraints as Computational Constraints in Tabletop Tangible Interfaces. In: CHI 2007, pp. 809–818 (2007)
22. Phidgets, http://www.phidgets.com/
23. Qt, http://trolltech.com/products/qt
24. Salber, et al.: The Context Toolkit: Aiding the development of Context-Enabled Applications. In: CHI 1999, pp. 434–441 (1999)
25. Underkoffler, et al.: A luminous-tangible workbench for urban planning and design. In: CHI 1999, pp. 386–393 (1999)
26. Wakita, Anezaki: Intuino: an authoring tool for supporting the prototyping of organic interfaces. In: DIS 2010, pp. 179–188 (2010).
27. Wellner: Interacting with paper on the DigitalDesk. CACM 36(7), 87–96 (1993)

A Personal Approach: The *Persona* Technique in a Companion's Design Lifecycle

Joana Campos and Ana Paiva

Instituto Superior Tecnico - UTL and INESC-ID, Portugal
joana.campos@ist.utl.pt, ana.paiva@inesc-id.pt

Abstract. *Artificial companions* are a new type of technology that is changing the way people interact with intelligent systems, by trying to build and sustain long term relationships with users. To evaluate such systems the typical usability methods for system evaluation are not enough, due to the need of accessing aspects such as social behaviour, emotional sensitivity and personalized interaction over long periods of time, with very specific users. In this paper, we describe the (full) design cycle of a companion system, arguing that a user-centred approach is the more appropriate, taking into account the referred specific issues. As such, to help us focus on developing a companion system for the prospective users, we based our design in the archetype of actual users – *a persona*. Furthermore, we brought this same concept into the evaluation phase in order to access the companion performance in a long-term interaction.

Keywords: Companion design, persona, companion evaluation, long-term.

1 Introduction

During the past few years we have witnessed a new boost in the area of intelligent agents applied to human-computer interaction (HCI). A new kind of technology is gaining ground as it tries to establish "social relationships" with users, rather than simply interacting with them [2]. Embodied as robots, 3D synthetic characters, or even interactive toys, an *artificial companion* should be able not only to promote more pleasing and natural relationships with users, but also to build and sustain such relations over long periods of time.

Companion systems are designed to assist or help users in a specific task and at the same time to act socially in an appropriate fashion [7]. Furthermore, this technology attempts to achieve long-term robustness, and as a result it should encompass some capabilities, such as affect sensitivity, memory or learning [7]. The Senior Companion system [8], for example, helps elders to organize their life memories, while it provides comments and tries to establish an emotional connection. Another suitable example is the Health and Fitness Companion [20], which tries to persuade the user to engage in a different and healthier lifestyle. It does so by providing advice and guidance throughout the day about combinations of diet and exercise.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 73–90, 2011.

Indeed, *companion systems* are a new type of technology that allows one to manage his/her life and personalize the interaction accordingly. Furthermore, these systems are task specific and user focused, whereupon to address design issues it is crucial to adopt a methodology that iteratively involves the user in its design process. Therefore, we argue that user-centred design methods convey this process in a dynamic way, wherein the user participation is essential throughout the whole process both from the requirements analysis to the evaluation phases. However, when users are not present, there are still a lot of decisions to be made that should follow the same paradigm.

To overcome this issue and to help us focus on developing a companion system for the prospective users, we based the design of our *companion* on an archetype of actual cohorts – a *persona*. This technique relies on a character that is built upon information from potential users who participated in early stages of design. The underlying idea of *personas* is to increase user focus and awareness, as they are specific assumptions about a set of people and also to make the decision criteria explicit [18][24].

In this paper we describe the design cycle of a *companion* system (fig.1) focusing on the essential processes of user-centred design (UCD): plan and identify context of use, gather requirements, prototype design and evaluation [22]. We would also like to emphasise that the use of UCD enables a shorter development lifecycle by guiding the implementation process.

However, when evaluating *companion* systems there are two key issues to consider [2]. First, relationships are long-term built over many interactions and secondly relationships are persistent, emotional and personalized. Undoubtedly, the memory is an essential component for the agent to show its capabilities according to the two previous issues. As such, it should be primarily populated with a considerable number of interactions. Yet, this issue would turn the user-centred design for *companions* a very expensive and extremely long process. That is related to the fact that long-term evaluations between participants and companions are costly and hard to manage.

As [3] suggested companions technologies require new models for evaluation. For that reason, we suggest to use the concept of *persona* on the early evaluation stages. As the *persona* is a complete and carefully defined character it would be possible to fill the *companion's* memory with the tailored *personal* experiences. We believe that by following this paradigm we can simulate previous interactions and access the agent's performance, whereas we are simulating a long-term interaction in a short duration evaluation.

2 Companion Context of Use and Design Overview

Throughout this paper we will describe the development of a companion system – MAY (my Memories Are Yours) – created to assist a teenager user on self-reflection and daily companionship about what happens in his/her life [5]. This system tries to preserve the user's expressivity in analogy with a traditional diary.

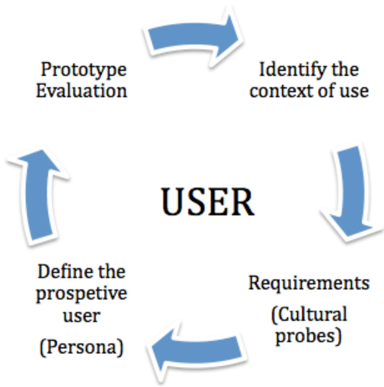


Fig. 1. The companion design cycle

To try to achieve that goal it is necessary to take into account that language is the currency of most human social processes [10], so the interaction is based on that issue. Moreover, the main component of this *companion* is its memory, as we believe to be not only an essential aspect to improve intelligent systems in general, but also an important feature to help our *companion* behave more naturally in a social setting.

When planning the design process, we used a set of (what we believe to be) suitable methods for guiding a *companion* development focusing on a UCD perspective. First, we used the cultural probe method for gathering information about user's tasks and to analyse their needs in a less obtrusive way. From that data we were able to identify and formally define a *persona*, which guided the subsequent stages of the *companion* development. With the gathered information, we were able to build a companion prototype, which was also evaluated. This time, we used the *persona* as a central element of the process. The role of the potential users that participated in this phase was to classify whether the memory allowed the agent to produce appropriate responses according to its content (persona's experiences), or not.

3 Personas

Persona is an archetype of actual users with well-defined characteristics and it was first introduced in HCI community by Cooper [11]. Their use helps designers focus on prospective users of a system or interface, taking into account their needs and goals [9]. These imaginary characters must be defined according to data retrieved from initial investigation of the target user.

Personas have various attributes such as name, age, gender, educational achievements, profession, possessions, likenesses, family, occupation, friends, pets, tasks, life stories, goals, among others [19]. They are "imaginary" users that designers consider while developing a product. This technique arises from the difficulty of continuously work closely with potential users on a possible new product, especially in large organizations [19].

The use of personas has recently become popular worldwide as a design method. For instance, it was used for the development of MSN Explorer products [24]. This success can be attributed to its potential of engaging designers during the prototyping phase. Researchers argue that people more easily remember stories and narratives than technical reports. Consequently, *personas* with their stories can persist on the designers mind set. Furthermore, personas act as a mechanism to enhance attention and help on the organization of data [19] carried out during the design phase.

This idea is also prevalent even in the design of TV series. It is frequent to see people believing that TV characters are real and see their life as realistic. Additionally, people are keen in discussing details and make predictions about characters behaviour. These predictions are made according to a set of characteristics, which they have gathered from seen episodes. On the other side, the actor builds his character and exercises a story where the character may fit. Certainly, the actor adds some details to help him behave more naturally [19], as well as, the designers make assumptions about the created persona.

Some studies [9][24] evaluated if the presence of the *persona* factor had a positive effect on the final design comparatively to the absence of it. Their studies underlined that the use of personas helped designers on picturing the user and get to a consensus. However, *personas* should be used for a particular effort and should not substitute other techniques for user-centred design. Instead, they should complement them by enhancing user focus. On the whole, this is a good approach for complementing and guiding not only the design process, but the test phase as well [19].

If the persona is efficient in engaging designers to focus on the users, it should be equally effective on engaging users, when they perform tests with some system or interface. That issue was already raised [18] as the author used personas to test the usability of a web prototype.

4 Gathering Information about the Users

When starting the design process, our initial purpose was to uncover the practices through which people experience memory sharing and discover how they perform activities that the system tries to assist. But how could we analyse ‘real life’ without losing personal expressivity?

To achieve a ‘close view’ of daily life without blurring the essential aspects (the one’s that we had not thought of) we moved away from the controlled environment found in a lab, sending a probe to potential users and waiting for its return.

Cultural probes are an interesting collecting process, wherein people agree to share personal aspects of their lives. The probe exercise gives necessary detachment from the designer, and provides information about how people use the given materials to express themselves in an unobtrusive way [13]. Probes do not directly lead to designs, as we could experience, but make us aware of details, which are valuable in inspiring design ideas [16]. We used this approach as a starting point to satisfy our intentions.

Our experiment was focused on teenagers, mainly ones that have just started their academic life in the university or were about to start it. We wanted to disclose what they are willing to tell to a companion, which things are important in their life and which are of their concern. How 'stories' of their life are told, and consequently organized in words, and which are their routines. Further, we wanted to analyse how people used the provided materials.

4.1 The Probes

We gave to the participants an envelope (Fig. 2) containing ten tasks to perform, a blue pen, coloured pens, post-its, stickers with emotions and a notebook – their *companion*.

Each task had clear instructions of what was needed to do, yet not restrictive about how to perform them. The first task aimed at creating a first bond with the companion requiring an image of it/she/he. The following tasks, tried to 'extract' ways of interaction and sharing experiences (memories) process. People had liberty to deal with the requests as they preferred.

Fig. 2. The probe pack

The 14 participants (seven boys and seven girls) from ages 16 to 18 were first-year students of different universities and high school finalists. They took the probes and returned them 3 weeks later.

Probes should not be used to produce scientific analysable data [16], therefore, it is common to give participants ambiguous and absurd tasks to create detachment from preconceived definitions of the world. Although our designed tasks (for more information please see [6]) were relatively straightforward, they called to a subjective interpretation from us, and a deeper thought from the participants. However, we could experience that the most structured questions/tasks were not the ones that produced more accurate responses and that give us a clue on how to analyse the probe material. Tasks that ask them to describe episodes of their lives had a lot of (more) valuable information.

The most relevant data extracted concerns the companion image, what users consider important in their lives, as well as, the process of remembering described in a diary form.

4.2 What Was Probed?

This experiment provided an engagement with the participants, allowing us a ‘close view’ of their lives. The ‘storytelling tone’ was constant in their responses, as they literally wrote a diary. However, a direct interaction with a companion was not left out, and during the experiment some subjects really ‘talked’ to MAY. In the light of this, we established that the communication between user and companion would be by text to preserve the expressivity we found in the data. Then, the central point of the interaction would be dialogue, which facilitates the sharing of knowledge and individual’s information.

Moreover, the probes enhanced a conversational behaviour as suggested by [14]. There is a focus on talking about personal experiences and relationships, as well as the emotional responses involved in those interactions. The personal experiences fit in one of the four important dimensions of younger’s life: love, sport, leisure and school. The way in which the given materials were used did not surprise us, but stood out the necessity of a feature to enable attaching objects, like images or sounds, to the written text. Still, the probes oriented the design of the interface and a first prototype was modelled.

5 Defining the Prospective Users

The collected answers from the cultural probes, revealed to be an insight of each participant and its combination brought up a “conceptual model” of the prospective group of users and how the system should behave.

Following the *persona* design technique, we created a fictional user, named Amy (Table 1 and 2). The data extracted from one probe was our starting point to create that fictional person, which behaviour was interactively built based on real data from the rest of the potential users, who have participated on the probe study. The created character aggregated a set of characteristics from several people (participants in the probe study) and not from just one. Our aim at creating Amy was to guide the companion’s design both its interface and implementation. Further, we would like to extend the use of personas to help us surpass the long-term issue, based on the engagement that this tool can provide. More details on this matter are explained on section 7.

The persona Amy is 18 years old and is on her first year of Computer Science in the university. She likes summer time and being with her friends whenever is possible. April is her best friend, and they usually go to the cinema together and out for dinner once in a while. As many teenagers one of her goals is passing the driving test and of course be successful in her studies at the university.

Table 1. Persona details and life stories. The enumerated characteristics were suggested by Grudin [19] and completed with the probes' information.

Persona Definition		
Ethnography	**Name**	AMY
	Age	18
	Gender	Female
	Socioeconomic Status	Middle Class
	University	Blind review
Details	**Family**	Amy lives with her parents and sister. She always have fun with them and they are always there to back her up
	Friends	Friends are important in Amy's life, and them are in her list of concerns. Her best friend is April, they used to go to the cinema, to go shopping or have dinner/lunch once in a while. They try see each other every Tuesday or Friday.
	Pets	She has a dog to take care.
	Sport	Every time she plays volleyball, she plays with her heart. She totally loves it and for her it's a great escape from school.
	Likes	Amy is a summer person. She also likes listening music, travelling, camping, school (why not), eat good food, going out at night with her friends. She prefers places like Hard Rock Café, but Bairro Alto is one of her choices too.
	Dislikes	Wake up early, public transports (school-home), domestic tasks at home.
	Goals	Pass the driving test, get good grades at university subjects, to be in good shape, wake up on time for school
	Tasks	Take care of the dog; at Tuesdays and Fridays is her turn to clean up the kitchen; clean up her room weekly; go to volleyball practices; go to drive lessons

6 The Scenario

As it was described before, the companion MAY, is an agent created to assist a teenager user on self-reflection about what happens in his/her life. The interaction

Table 2. Persona Life Stories. (continuation of table 1)

Persona Definition	
Life Stories	"In Summer 2008 I went on a camping trip to Açores with some people from school. Against all odds it rained a lot! The worse came later… At the middle of the night some tents collapsed and we have to go our 'neighbors' tens to spend the night. In the morning all campsite had been flooded and our clothes and stuff were soaking."
	"Today is my 16 birthday and I got my first car – a Microcar. I am so happy and excited. It is like a dream come true. I cannot wait to show it to all my friends."
	"Today was a normal day at school. Slightly boring… don't really want to work. But I went to one beach for lunch with friends, really good moreover."

between them is through dialogue, in chat like interface, which allows the agent to collect the user's experiences and save them in a diary form (or a timeline). The memories (or events) stored in the companion's memory constitute a kind of "shared memory" between the two [5], which can be compared to an affective diary [25].

To produce an adequate response, the agent starts by searching its memory for anything appropriate to say. It looks for active goals, past events with some relevant information for the current situation or even to go beyond the present and infer future plans. This allows increasing the agent's responsiveness, that is, listeners make empathetic and contextually grounded responses to what the speaker is saying. This process is considered not only to enhance believability of conversations, but also serve engagement and intimacy in relations [4] making it an enjoyable two-way interaction.

7 Persona and Companion Early Evaluation

As noted earlier, new systems that try to change how people interact with technology making them develop behaviours towards relationship building [2], require new methods for evaluation. In the context of the COMPANIONS[1] project, for example, researchers outlined a new evaluation paradigm to measure their companions' conversation quality [3]. They suggest the appropriateness annotation as a measure to evaluate the companions' conversational performance, along with objective and subjective measures. Others [26] have explored the affect and emotions in interactions as part of user experience as the interest for emotional design [27] has gained interest.

In our work, we are particularly interested if people can identify cues in dialogue that may elicit some kind of behaviour that could demonstrate social capabilities in the agent's discourse. Furthermore, we wanted to see if users perceive the agent as intentional, based on the result of previous interactions stored in a timeline. As such, a User-Centred method [3] seemed appropriate for our study. This allowed us to acquire

[1] http://www.companions-project.org

subjective impressions and opinions by the participants collected in Likert based surveys [3].

For a companion system showing user acquaintance during a dialogue interaction it is necessary many interactions to capture and store information. However, in early stages of the design, such kind of evaluation is costly and hard to manage.

So, we brought the concept of *persona* into the evaluation stage in order to assess the companion's memory performance in a long-term interaction, but in a short-time period. As such, we created a *persona*, called Amy (already described in section 5).

As an attempt to simulate the long-term issue, we filled the agent's memory with enough information to perform in a real situation. We introduced memories of experiences into the system, taking into account the life story we had created for Amy, supported by the idea that models of fictional persons can be engaging as real people [18]. This process was done with a help of an authoring tool to define a persona according the specified parameters, which was built by us. Using this tool, it was possible to populate large "slots" of the memory, according to activities or events that we thought to be acceptable with respect to the persona's definition.

In the companion evaluation described in this section, we tried to verify if the presence of previously shared memories between the user and the agent enabled the development of specific kind of relationship between the two, in comparison with the absence of such element.

Thus, we evaluated if *an agent that is capable of indexing user's experiences and using that temporally grounded information in a social interaction, will make users to establish a stronger companionship relation with the agent based on what the agent knows.*

7.1 Evaluation Procedure

A total of 90 participants (72 male, 18 female, aged 18-24) took part in the experiment. All of them were undergraduate and the study was available through an online questionnaire. The questionnaire conducted the participants through three stages (see fig. 3). In the first two stages the subjects had to witness an interaction between Amy and MAY, which they should observe carefully. After each interaction they would be asked about MAY's behaviour. The last stage aimed at measuring how MAY's attitudes induced the relationship between Amy and MAY. For this last stage we applied an established friendship questionnaire. In each stage, when inquired, the subjects had to classify statements in agreement with a five-point Likert scale (1 - I strongly disagree; 3 - I don't agree or disagree; 5 - I totally agree), coupled with a justification to validate the response.

Notice that none of the subjects knew what the system did, neither if it should act as a companion nor an agent that saves personal facts about the user through shared memories. Plus, the interactions between Amy and MAY reflected some of the agent's capabilities of using information collected during previous interactions with the system. The full questionnaire and interactions between Amy and MAY are presented in [6].

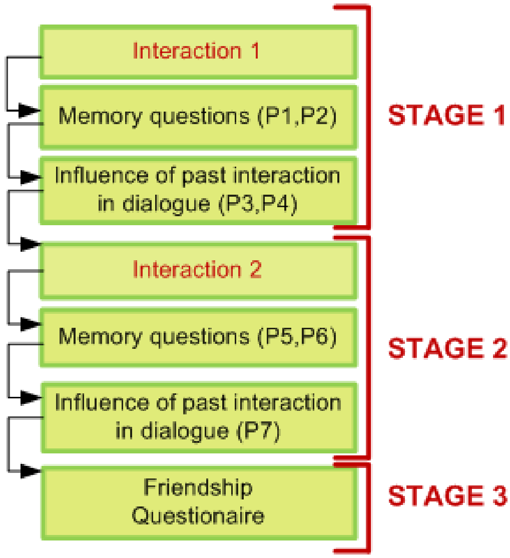


Fig. 3. Evaluation Flow

7.2 Manipulation

We conducted a between-groups experiment to evaluate the raised hypothesis. Thus, one group of participants was exposed to the memory condition and the other group to the no memory condition. The online questionnaire automatically randomized the participants to one of the conditions, which express the following situations:

- *memory* – there is no modification in MAY's architecture. There is a link to the Agent's Cognitive Features Module. This module is responsible for bringing back relevant information for the current situation.
- *no memory* – MAY's architecture was manipulated and there is no access to the Agent's Cognitive Features Module.
  - However, in this condition, MAY still knows the user's name and maintains the pro-active behaviour based on the design of the dialogue system. That is, it still 'cares' about the user by asking intentionally generic questions to get information about user's day.

7.3 Results

The collected data was analysed using the Mann-Whitney test to compare the differences between the two conditions.

For clarification, on the very right-hand column of each table (from stage 1 onwards) we provided r as a measure of the effect size, i.e. how significantly different the two conditions are. Although r may be interpreted as a correlation, in this study we use it to measure the strength of the experimental effect [25], which in this case was the presence of memory in the companion's responsiveness in each of the variables (questions 1 to 7).

The test-statistics (U) was converted into a z-score by SPSS and the *r* was calculated using the following formula: r = z-score/square(N), wherein N is the size of the sample. Consequently if z-score has a negative value the *r* will have a negative value as well. One may argue that means that the experimental group was worse off on these areas. But it does not, as it is not considered as a correlation factor. The issue is that variables were inversely encoded in SPSS. However, it is irrelevant in this case, as we want the absolute value of *r*.

Stage 1. In the first stage of the experiment the users watched an interaction between Amy and MAY, where the agent shows that he knows some predefined goals in Amy's life. Below are some examples of the type of interaction (memory condition)[2]:

```
11   MAY >> How was the Cranberries concert last night?
     Amy >> It was awesome! She sang all the songs!
...
25   MAY >> I'll be waiting.
             Did you go out for dinner two days ago?
27   Amy >> Yes. It was my sister's birthday. ...
```

Rather, in no memory condition such acquaintance is not verified, but the agent displays an empathetic behaviour embedded in the dialogue system:

```
11   MAY >> How was your day yesterday?
     Amy >> It was awesome! I went to the Cranberries'
     concert!
...
25   MAY >> I'll be waiting.
             Tell me your last and important memories.
27   Amy >> Saturday I went out for dinner. It was my
     sister's birthday. (...)
```

Figure 4 depicts a summary of the results obtained for the different questions at this stage of evaluation and exhibits the clear differences between the two conditions (see also Fig.6 for a graphical visualization of the descriptive statistics).

As the results show, users recognised the presence of knowledge about goals (Q1, Q2) and the results between conditions were significant ($p < 0.001$). These results may seem obvious, but we would like to emphasise that the users didn't know anything abut the system nor how it should behave. Nevertheless, the most relevant data refers to Q4, where the user recognised the information depicted in the dialogue as temporally grounded.

[2] The full interaction is depicted in [6].

However, no memory condition had a wide discrepancy of opinions that ranged within the five points (Mdn[Quartiles]=3.5[1,5]), when the participants were asked about the influence of the past (Q4). This dispersion of values is due to the agent knowing the user's name. For some participants that fact was a sufficient (in the absence of other) for the existence of prior interactions. As one user justified "MAY knew to whom was speaking, as she referred as Amy".

Cognitive Features	Id	Question Statement	Descriptive Statistics *memory (N=46)*	*no memory(N=44)*	Mann-Whitney differences between conditions
			Mdn[Quartiles]	Mdn[Quartiles]	
Tracking Goals	Q1	MAY knew about the event "go to the cranberries' concert"	5[4,5]	2[1,3]	U = 129.000 $p < 0.001$ r = -0.777
	Q2	MAY knew about the event "go out have dinner"	4.5[4,5]	2[1,3]	U = 232.500 $p < 0.001$ r = -0.68
Influence of past Interactions	Q3	MAY knew about the event "go out have dinner" because Amy had told it in any past interaction	5[4,5]	1[1,3]	U = 248.500 $p < 0.001$ r = -0.679
	Q4	Past interactions have influenced the current one	5[4,5]	3.5[1,5]	U = 527.500 $p < 0.001$ r = -0.452

Fig. 4. Mann-Whitney statistics for the first set of questions

Stage 2. In the interaction portrayed in the second phase of the evaluation, the memory condition enhanced the agent's acquaintance about some details of common or regular events in Amy's life.

```
12  Amy >> Yes. In a while I'm going to walk Teddy.
13  MAY >> Are you going to the park?
...
16  Amy >> This evening I'm going to the cinema
17  MAY >> Are you going with April, Amy? ...
```

Oppositely, in the "no memory" condition the agent keeps answering to Amy with general sentences that could fit in most situations, using the pattern-matching mechanism offered by the dialogue system.

```
12  Amy >> Yes. In a while I'm going to walk Teddy.
13  MAY >> Sounds good to me.
...
16  Amy >> This evening I'm going to the cinema
17  MAY >> I would like to go too ...
```

Cognitive Features	Question Id	Question Statement	Descriptive Statistics memory (N=46)	Descriptive Statistics no memory(N=44)	Mann-Whitney differences between conditions
Virtual Sensing	Q5	MAY knew something about the action "walk Teddy"	4[3,5]	3[1,4]	U = 589,000 $p < 0.001$ r = -0.370
	Q6	MAY knew something about the action "go to the cinema"	4[3,5]	4[2,4]	U = 722.500 $p < 0,05$ r = -0.256
Influence of past Interactions	Q7	Past interactions have influenced the current one	5[4,5]	4[4,5]	U = 797,500 $p < 0.051$ r = -0.206

Fig. 5. Mann-Whitney statistics for the second set of questions

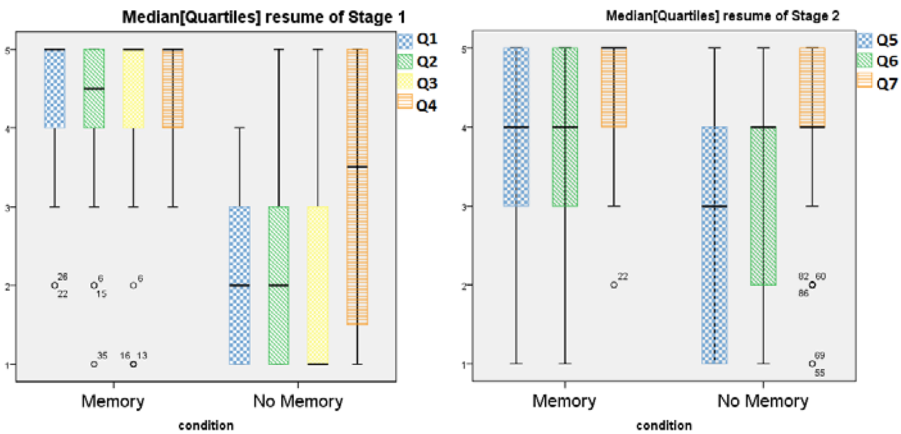


Fig. 6. Descriptive statistics for Stage 1 and 2

Once again the results in this stage may seem obvious, but users did not know anything about the system or how it should behave.

In both stages (1 and 2), specific elements in the dialogue from the stored information were recognized as intentional and as a result of past interactions. Yet, at stage 2, such statement is not so obvious as the medians of both conditions are equal or very close to each other. The explanation for significant results in this stage is based on the theory behind Mann-Whitney test. Mann-Whitney is not a median test, but a rank test. That is, it calculates the difference between two distributions of data. There is a difference between the distributions, which is supported by the quartiles (as depicted in Fig. 5).

Nonetheless, the *r* describes whether such significance is observed by chance or not. It tells us that our experimental condition had a small to medium effect size. These results may have happened for two reasons: 1) the agent's replies may have lead to a misunderstanding by the users. The responses make them feel that the agent 'knew' something about some characteristics of the activity; 2) by watching the previous interaction at stage one.

The participants probably had expectations about the second interaction, which made them think that the agent took conclusions similar to their own assumptions. According to user's justifications "MAY recognizes cinema as a fun activity because he replied that he would like to go"(referring to his answer to Q6).

Stage 3

Measures

One of the aspects that we were interested was to identify if the social mechanisms of memory used in this companion system lead to some level of companionship between user and agent. With MAY we wanted to build a companion that stays with the user for a very long time, developing a relationship and getting to 'know' the user's preferences and wishes [8].

In human relationships it is possible identify a type of relationship that fits in this definition – friendship. Therefore, we employed the McGill Friendship Questionnaire (MFQ) [26], which measures six dimensions of friendship between people. We used this questionnaire to classify the relation between MAY and user (Amy). Mendelson has identified six different functions of friendship conceptually distinct:

1. *Stimulating Companionship* – engagement in enjoyable, fun and exciting activities;
2. *Help* – guidance, aid and other types of help;
3. *Intimacy* – being attentive to other's states and needs and open to honest expressions about thoughts, feelings and personal information;
4. *Reliable Alliance* – keeping available and loyal;
5. *Self-validation* – comfort, encourage and help the friend on keeping positive confidence in himself; and
6. *Emotional Security* – providing comfort and confidence in novel or threatening situations.

According to Allan [1] friendship is "a relational term which signifies something about the quality and character of the relationship involved". Similarly, we wanted to measure the type of the relationship that the agent is capable of developing with the user based on what it knows about him/her. With MFQ we can somehow get a glimpse of the quality of the relationship between Amy and MAY. In particular, we were interested to examine if positive aspects of the relationship [12] prevail in memory condition in comparison with the no memory condition. The used assertions of MFQ were manipulated to fit into our experiment.

That adaptation was based on the results obtained by [21] in a online survey, in which the users had to associate a set of assertions with the dimensions of MFQ. Our experiment and interactions within both conditions are presented in [6]. Notice that the Emotional Security dimension was not measured in this study. The main reason resides on the fact that in order to recognise this dimension one needs much more than two interactions for the users to understand how MAY can provide comfort to them, as Leite et al [21] pointed in their study. Plus, the agent is only prepared to recognise situations that need a concerned response at the initial state of the dialogue and does not recognize emotional states during the interaction.

Results

At this stage of the evaluation, we applied the MFQ to measure the quality and the characteristics of the relationship. The dimensions *Intimacy* and *Companionship* were the ones that differ significantly ($p < 0.001$) in the extent of both conditions (see fig. 7). The dimensions Help, Self-Validation, Reliable Alliance are embedded in the dialogue structure, so the pro-active feature or 'attitude' of the agent contributes to the high level of agreement with the correspondent statements, in both versions of the questionnaire. The agent encourages the user, gives advice and has an empathetic behaviour that helps in the maintenance of the relationship. As Fig.7 shows, that differences were not significant for those conditions. However, looking at the participants justifications it was evident that they understood the agent's notion of time and that events talked about were theme of conversation between Amy and MAY in some past interaction(s). For example, some users' justifications were "Probably that information was mentioned before"; "Amy told MAY before"; "MAY knew that Amy usually goes to the cinema with April"; "MAY thought that Amy was going with April".

Having or not having memory can explain 11% of the total variability in M7 and 26% of the total variability in M2 (intimacy and companionship dimensions respectively) that suggests a wide effect of our experimental condition in these two dimensions, which we believe to be equally important in the maintenance of a relationship and enough to satisfy hypotheses. Some studies, such as [12] verified that when friends are seen as a primary source of social support, companionship and intimacy are quite important factors that make those relationships endure.

8 Discussion

The overall goal of this last evaluation was to measure the user's perception of the relationship that the system might have developed with the user, based on its memory functions. The experiment was done in three stages using the persona Amy. Users had to classify the relationship between Amy and MAY, after the agent show some knowledge about her life.

The positive results on *companionship* and *intimacy* dimensions were interesting results for a first prototype. Intimacy is a dimension that is strongly related to long-term, as it develops over several interactions between two people. This may suggest that use of the *persona* Amy allowed us to simulate a long-term interaction in a short duration evaluation, letting us to glimpse, which dimensions would have emerged over time. However, a longitudinal study may reveal other effects on the user than laboratory studies. Certainly, the use of *personas* must not undermine other types of UCD evaluation methods, but rather complement them.It is clear that the described paradigm is adapted to a niche: *companion* systems, which depend on memory and also on several interactions with a user to perform properly. However, is such situation, the use of *personas* seems adequate due to its specific definition and properties, allowing to simulate informed previous interactions with the system.

Friendship Dimensions	Id	Question Statement	Descriptive Statistics *memory (N=46)*	Descriptive Statistics *no memory (N=44)*	Mann-Whitney differences between conditions
			Mdn[Qaurtiles]	Mdn[Quartiles]	
Companionship	M1	MAY is friendly	5[4,5]	5[4,5]	U = 842.000 $p < 0.05$(1-tailed) r = -0.17
	M2	MAY was companion during interaction	4[4,5]	3[2,4]	U = 433.500 $p < 0,001$ r = -0.512
Help	M3	MAY's comments are useful	3[2,4]	3[2,4]	U = 955.500 *ns*
Self-Validation	M4	MAY does Amy feel that she can do things well	4[3,4]	4[3,4]	U = 975.500 *ns*
	M5	MAY encourage Amy	5[4,5]	5[4,5]	U = 995.500 *ns*
Intimacy	M6	MAY have interest about Amy's life	5[4,5]	5[4,5]	U = 949.000 *ns*
	M7	MAY knows Amy well	4[3,4]	3[3,4]	U = 639.000 $p < 0.05$ r = -0.332
	M8	MAY knows facts about Amy's life	4.5[4,5]	4[2,4]	U = 582.000 $p < 0.001$ r = -0.383
Reliable Alliance	M9	MAY woul still Amy's friend even if a month pass	5[4,5]	5[4,5]	U = 982.500 *ns*

Fig. 7. Mann-Whitney statistics for the agent's characteristics

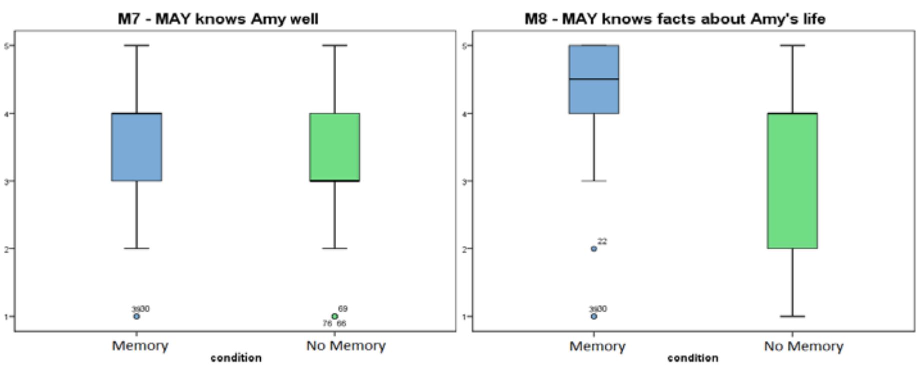


Fig. 8. Differences between the two conditions - *Intimacy* dimension

9 Conclusion

User-centered design (UCD) practices are employed for accessing a clear understanding of the user and task requirements, as well as, giving the user an active role in software development [22]. It is clear that when developing *companion* systems the design must drawn in potential users, as the system should be engaging over long periods of time.

In this paper we described the design cycle of a *companion* system focusing on the essential processes of UCD using the archetype *persona* to guide design decisions. Furthermore, we suggest using the concept of *persona* to the early evaluation stages. As the *persona* is a complete and carefully defined character it would be possible to fill the *companion's* memory with the tailored *personal* experiences. We believe that by following this paradigm we can simulate previous interactions and access the agent's performance, whereas we are simulating a long-term interaction in a short duration evaluation.

Acknowlegments. This work is partially supported by the European Community (EC) and is currently funded by the EU FP7 ICT-215554 project LIREC (LIving with Robots and IntEractive Companions), and FCT (INESC-ID multiannual funding) through the PIDDAC Program funds.

References

1. Allan, G.: Friendship: Developing a Sociological Perspective. Westview Press, Boulder (1989)
2. Benyon, D., Mival, O.: Landscaping personification technologies: from interactions to relationships. In: CHI 2008 extended abstracts on Human factors in computing systems (2008)
3. Benyon, D., Bradley, J., Mival, O., Webb, N.: Evaluation Paradigm
4. Bluck, S.: Autobiographical memory: Exploring its functions in everyday life. Memory 11(2) (2003)
5. Campos, J., Paiva, A.: MAY: My Memories Are Yours. In: 10th International Conference on Intelligent Virtual Agents, pp. 406–412. Springer, Heidelberg (2010)
6. Campos, J.: MAY: My Memories Are Yours. An interactive companion that saves the user's memories. Master Thesis (2010)
7. Castellano, G., Aylett, R., Dautenhahn, K., et al.: Long-term affect sensitive and socially interactive companions. In: Proceedings of the 4th International Workshop on Human-Computer Conversation (2008)
8. Catizone, R., Worgan, S., Wilks, Y., Cheng, W.: A multimodal conversational companion for reminiscing about images. In: Fourth International Workshop on Human-Computer Conversation, Bellagio, Italy (2008)
9. Chang, Y., Lim, Y., Stolterman, E.: Personas: from theory to practices. In: NordiCHI 2008: Proceedings of the 5th Nordic Conference on Human-Computer Interaction, pp. 439–442. ACM, New York (2008)
10. Chung, C., Pennebaker, J.: The psychological functions of function words. In: Social communication: Frontiers of social psychology, pp. 343–359. Psychology Press, San Diego (2007)
11. Cooper, A.: The Inmates are running the Asylum. In: Software-Ergonomie 1999, Design von Informationswelten, Gemeinsame Fachtagung des German Chapter of the ACM, der Gesellschaft für Informatik (GI) und der SAP AG, 17– Teubner (1999)
12. de Souza, L.K.: Amizade em adultos: adaptação e validação dos questionários McGill e um estudo de diferenças de género. Universidade Federal do Rio Grande do Sul (2006)

13. Dix, A., Finlay, J., Abowd, G.D., Beale, R.: Human-Computer Interaction, 3rd edn. Prentice Hall, Englewood Cliffs (2004)
14. Dunbar, R.I.M., Marriott, A., Duncan, N.D.C.: Human conversational behavior. Human Nature 8(3), 231–246 (1997)
15. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural probes. Interactions 6(1) (1999)
16. Gomes, D.R., Pereira, C.A., Petreca, T.M., Aquino, Jr., P.: Teste com usuário baseados em personas em beneficio da prototipaçao em sistemas web. In: IADIS Ibero-Americana (2008)
17. Grudin, J.: Why Personas Work: The psychological evidence. In: The Persona Lifecycle, keeping people in mind through- out product design, pp. 643–663. Elsevier, Amsterdam (2006)
18. Grudin, J., Pruitt, J.: Participatory design and product development: An infrastructure for engagement (2002)
19. Hakulinen, J., Turunen, M., Smith, C., Cavazza, M., Charlton, D.: A model for flexible interoperability between dialogue management and domain reasoning for conversational spoken dialogue systems. In: Fourth International Workshop on Human-Computer Conversation, Bellagio, Italy (2008)
20. Leite, I., Mascarenhas, S., Pereira, A., Martinho, C., Prada, R., Paiva, A.: Why can we be friends? an empathetic game companion for long- term interaction. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents (2010)
21. Maguire, M.: Methods to support human-centred design. International Journal of Human-Computer Studies 55(4), 587–634 (2001)
22. Picard, R.W.: Affective Computing. MIT Press, Cambridge (2000)
23. Pruitt, J., Grudin, J.: Personas: practice and theory. In: DUX 2003: Proceedings of the 2003 Conference on Designing for User Experiences, New York, NY, USA, pp. 1–15 (2003)
24. Mendelson, M.J.: Measuring friendship quality in late adolescents and young adults: Mcgill friendship questionnaires. Canadian Journal of Behavioural Science (November 11, 2009)
25. Field, A., Hole, G.: How to Design and Report Experiments (2003)
26. Mahlke, S.: Understanding users' experience of interaction. In: Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics, University of Athens (2005)
27. Norman, D.A.: Emotional Design: why we love (or hate) everyday things. Basic Book, NY (2004)

Emotive Expression through the Movement of Interactive Robotic Vehicles

Eric Kryski and Ehud Sharlin

University of Calgary
{epkryski,ehud}@ucalgary.ca

Abstract. In this paper, we discuss our on going design of interactive personal vehicles that exhibit behavioral constructs expressed through motion in order to improve the user's commuting experience. The behavioral, personality-like traits demonstrated by the interactive vehicles are intended to be useful and helpful, as well as to stretch beyond the effectiveness into affect and emotion, creating an overall more satisfying experience for the user. This shortpaper presents our design goals and approach, describes the evolution of the implementation of our personal vehicle prototypes, and outlines our current preliminary design critique evaluation findings.

Keywords: Social human-robot interaction (HRI), Vehicle personality, Emotional expression through motion.

1 Introduction

Vehicle personality may seem like an abstract concept but in reality many people already treat their vehicles as social entities which possess the ability to understand and respond to their requests. Humans can have distinct emotional connections to inanimate objects, in the same way that they may have an emotional connection to a pet or child. Emotions that are evoked by these relationships can range from anger to fear to pleasure, and can play a fundamental role in the way we operate our vehicles [1].

Furthermore, the interactions we have with our vehicles can be viewed as a reminiscence of the symbiotic relationship between humanity and its classic modes of transportation, for example, the close relationships between handlers of camels or horses. However, although we can argue that there is some emotional connection between owners and vehicles, a connection strong enough, for example, to sometime persuade owners to name their vehicle, we believe that modern interactive vehicles could elevate this emotional connection by using synthetic personality and emotive expression. Arguably, if we can bring this heightened level of emotional connection to the way we interact with our vehicles it could potentially greatly enhance our commuting experience while simultaneously solidifying humanity's 'love affair' with our vehicles [2].

The purpose of our research is to design an interactive personal vehicle that exhibits emotional expression through motion, and to discover how humans will react

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 91–99, 2011.

to such a vehicle. That is, more specifically, what behavior types exhibited by the vehicle motion and movement would be acceptable to humans? Which displays of behavior are beneficial and which ones pose a negative impact on human-vehicle interaction? Our work was inspired by previous efforts in the domain, for example it has been shown that matching a driver's emotion to the perceived emotion of a simulated car voice can improve the driver performance and safety [3]. Our work however is an attempt to explore the emotional connection between human and vehicle through a more basic and entirely non-verbal layer: emotional expression as expressed via the vehicle motion. Our intent is not to replace other modalities such as visual or auditory, but rather to discern what physical emotive cues could supplement these other modalities. In this paper we examine three different classes of behavioral constructs (*semi-autonomous*, *cooperative*, and *training or taming*), which we argue are applicable to emotive vehicle expression through motion. We present the personal vehicle testbed prototypes we designed and implemented in order to probe these behaviors and their effect on users, and present our preliminary evaluation of one of these classes, as a simple test case of emotive expression exhibited through motion.

2 Background

Media and technology provide many examples of convincing displays of personality conveyed by seemingly inanimate objects, for example in computer games, film, and robotics. In film, viewers have been persuaded to emotionally reflect on a caring teapot, an inquisitive lamp, or a heartbroken robot. Even if these objects cannot talk or do not possess any facial features audiences were still able to associate them with personality and emotive expression. Reeves and Nass suggested that associating human characteristics to objects is an inherently human trait [4], and Heider and Simmel in their seminal research [5] outlined that people can relate emotion, gender, and motivation solely by perceiving an objects motion. Since movement is fundamental to robotic interfaces the above findings can provide crucial insight on how to design socially acceptable robots.

Based on this previous research, recent efforts are pursuing possible mapping between motion and emotion. In Saerbeck and Bartneck's research it was suggested that humans perceive a robot's movements in different ways, depending on the acceleration value and curvature [8]. Another effort presented a preliminary study about how abstract emotive actuation, when exhibited by a robot, is perceived by people [6], showing that people made conclusions as to what an abstract object was feeling, its intent and even its gender, based upon how fast the object moved, in which directions, and how quickly it accelerated.

3 Classes of Behavioral Expression

Since interpersonal communication relies heavily upon body language [5] it is plausible to believe that motion is a primary contributing factor to people's perception of emotion and intent. It seems realistic to think that in order to simulate human emotions robotic vehicles should apply behavioral contexts that are clear and apparent to humans. This is the intent of our research: to discover which behavioral expressions, conveyed via motion, are acceptable for a vehicle, and why. We argue that physical interaction can be more subtle, but also more effective, than either visual or auditory methods. We also believe that humans are acquainted with subtleties, and that interfaces that can express these subtleties can ultimately be more effective, and enhance the interaction experience, without necessarily overruling other, less subtle, modalities. Below we introduce a proposed set of applicable behavior classes that a vehicle may be able to exhibit during emotive movement.

Semi-autonomous Behavior. During the semi-autonomous behavior the vehicle navigates or transports the user where they wish or are required to go. It is merely a reactionary behavior where the primary purpose is to transport the user but also to alarm the user, for example via abrupt motion or vibration, of any information that is of importance. This personality profile allows the user to focus on other tasks as opposed to navigating, but still maintains situational awareness via subtle motion-based expression.

Taming/Training Behavior. In this behavioral state the vehicle initially begins as a blank slate that may or may not behave the way the user wishes. Similar to a puppy dog or a horse, the vehicle must be "broken" or trained to respond the way the user would like them to. The vehicle may, for example, move too fast for the user who prefers a more cautious style of movement. Conversely, the user may want the vehicle to move in a more aggressive manner with quick accelerations and more abrupt stops. The main goal of the taming/training behavioral type is not to be reckless or unsafe, but rather to afford the user with a sense of customization, companionship and pride of ownership akin to if one were to train a horse or a dog. It should be noted that similarly to training an animal, the training of an emotively expressive vehicle would also be performed in a secure environment before the vehicle is publicly exposed.

Cooperative Behavior. During this behavioral state the user and the vehicle collaborate to enhance the user's experience, and help the user fulfill a predefined set of tasks (such as reaching her meetings on time). The vehicle may exhibit portions of the semi-autonomous state or may be completely passive until the user wishes to perform some action that is not in their best interest according to the vehicle's HRI awareness [9]. The primary goal of the cooperative behavior type is to perform every action in the best interest of the user. The vehicle accomplishes this goal by utilizing a set of physical gestures or movements that subtly, or explicitly direct the user to predefined areas to complete predetermined tasks or physically removing the user from an unwanted dangerous situation.

4 Prototype Design

In order to practically evaluate our emotive vehicle expression through motion research question, and to examine the three behavioral types we outlined above, we designed two prototype personal robotic vehicles.

Our initial inspiration came from Segway™ but we ended up opting to use another widely used personal vehicle; an electric wheelchair. However, we did not want to limit the discussion scope to wheelchair bound individuals. As a result, we developed two prototypes based upon an electric wheelchair platform, one to be operated in the seated position and one in the standing position (Fig. 1A, 1B). In order to support rapid prototyping and critical evaluation of our concept, rather than a sophisticated technical contribution, we decided to use the 'Wizard of Oz' prototyping approach [10] as it greatly reduced the vehicle complexity and allowed us to develop a functional prototype in a short period of time. The shortcoming of our approach is that we deferred many of the actual complex implementation challenges associated with our interaction design, and focused our contribution only on the fundamental interaction concept.

4.1 Mechanics

We have designed two prototype vehicles, a sitting (Fig. 1A) and a standing platform (Fig. 1B), both of which are based upon an electric wheelchair structure. We chose an electric wheelchair because it is already viewed as a personal vehicle, it was easy to modify for our needs, and electric wheelchairs have the ability to operate in many different environments. That is, they are not constricted by legal issues in the same way that a Segway™ or an automobile is.

(a)

(b)

Fig. 1. (a) Seated Platform (b) Standing Platform

In order to support our 'Wizard of Oz' prototyping approach both vehicles use a Nintendo Wiimote with Nunchuk extension as input for both the user and the wizard. Each controller is connected via Bluetooth to an onboard vehicle PC. In order to get the desired wireless range, so that the wizard is not detected by the user, we used a generic class 1 bluetooth dongle, which provided an approximate operating range of 50 meters indoors.

After receiving information from the Wiimotes, the PC communicates with the vehicles' Dimension Engineering Sabertooth 2X25A motor driver via a USB to TTL cable. For safety reasons both prototypes include a momentary "on" push button kill switch, in between the battery ground and the motor driver ground, a button that the user must keep constantly pressed using their foot in order for the vehicle to operate.

4.2 Wizard-User Control

Control of our prototypes were based on a "Wizard of Oz" puppeteering method, wherein the user was unaware that the vehicle was at times controlled by another human using a second Nintendo Wiimote. The intent was to allow the 'wizard' to provide input on behalf of the vehicle, deceiving the user into thinking that it is actually the vehicle that is deciding and acting out the movements. As previously mentioned we opted for the 'Wizard of Oz' approach because it allowed us to perform rapid prototyping and it greatly reduced the complexity of the vehicle while still maintaining the ability to get user feedback [10] regarding how a vehicle's motion impacts the user's perception of a vehicle's intent and emotional behavior. A negative aspect of using this technique is that the wizard needs to deceive the user, which requires considerable awareness of the user intentions and puppeteering skills, both prone to human error. This can be especially problematic when dealing with a larger robot that is responsible for transporting a human, however this risk was mitigated in part by the user's on vehicle kill switch as well as by the wizard's ability to cease the vehicle locomotion at any point.

Upon initial implementation a decision was made to give the user full control until the wizard pressed a button on their Wiimote and overrode the user's control. At first glance this seemed appropriate but it quickly became apparent that when the user had no control at all they thought that there was a malfunction or they completely gave up on controlling the machine.

As an alternative to this solution we decided to, when the wizard interjects, give the user half the power of the wizard. This way the user can, if they so choose, "fight" the vehicle to an extent, which is controlled by the wizard, but the wizard will still be able to guide the vehicle in the direction that they intend. We believe that this solution will be more readily embraced by the user as it still leaves them with some sense of control over the vehicle regardless of its emotive expression.

5 Evaluation

For the design critique we chose to evaluate the cooperative behavior class, primarily because this behavioral expression seemed the most practical to implement and we hypothesized that users would be able to more readily grasp the concept of a "personal assistant" or "helping robotic vehicle". As well, for the prototyping effort we believed that it was the simplest to execute and the easiest to evaluate. As such, presented are two selected possible scenarios which inspired our design, both demonstrate how the cooperative behavior type may be beneficial to a user:

Scenario 1. *The user is at an airport and their flight has been delayed. They have decided to take their personal vehicle to peruse the shops in the airport in order to pass the time. Being preoccupied, the user is unaware of an announcement that their departure time has been moved up. The vehicle however, could be aware of the flight schedule and may begin to subtly physically nudge the user towards the direction of the departure gate. The user may decide to correct the vehicle and return to the shop but as it gets closer to the departure time the vehicle may become more persistent and assertive. Alternatively, if the user is continually correcting the vehicle's behavior, it may simply give up on its attempts and silently obey.*

Scenario 2. *The user is operating the vehicle inside a corporate building and there is an explosion. Alarms are going off, inaudible announcements are being made over the speaker system and smoke is beginning to spread. The user is now distressed and there is public panic. Since some people have reduced rational decision making ability during times of emergency or high stress [8] the vehicle would assume an authority role and immediately guide the user to safety without requiring user guidance or input. The vehicle may appear protective, brave, or confident by moving in an abrupt direct manner, thus instilling confidence in the user.*

During these two very different scenarios the motive displays of emotion by the vehicle, such as being brave or helpful, will vary situationally between passive and aggressive. Consequently, even though we are dealing with the cooperative behavior class the assertiveness or passiveness of the vehicle's movement is an important factor in determining the appropriate motive emotion for each situation. As such, this is something that we took into consideration when selecting the cooperative behavior class for our prototypes.

For the preliminary evaluation of our prototypes we performed a design critique using some of our design team members acting as users. The obvious limitation of our approach was that the users were informed of our prototyping approach. However, we still believe that our findings, outlined in this section, provide important insight on how to proceed pursuing our research question.

For our preliminary design critique evaluation we decided to use the cooperative personality type in lab setting, and accordingly opted to develop a scenario where it makes sense that the vehicle would exhibit a cooperative personality. As such, we devised a scenario where the user is required to complete a certain set of pre-defined daily lab tasks, such as picking up a print out from the printer room, in an allotted period of time. We constructed this scenario based on what we thought would be a meaningful and relatively validtask for most participants, that is, many people perform similar chores on a daily basis. It should be noted however, that our participants did not integrate these vehicles into their daily lives. Our rationale for selecting the cooperative behavior class was that it required the least amount of human learning and we hypothesized that users would be able to more readily associate with a cooperative behavior. The design critique was conducted with two members of our research lab and one engineering student external to our lab, using the seated platform. All of the participants were involved and familiar with our research and hence these findings can be viewed only as a design critique, not a user study.

5.1 Prototype Findings

Overall the prototypes operated quite well and both vehicles can be operated smoothly or in a jerking manner. On the mechanical end, the seated platform is quite stable, even at high speeds, but the standing platform can be quite tippy in the lateral direction. A larger base, lower center of gravity, or wider wheels will increase the force it takes to tip the standing vehicle. Taking this safety concern into consideration is why we opted to use the seated platform for the design critique.

5.2 Interaction Findings

After the initial trial we found that the participants were often unaware of the vehicle's intention, which caused them to become confused, frustrated or fearful. As such, we opted to revise the design critique scenario to get the user to perform a series of tasks in the same fashion, except that some key information required to complete some of the tasks, would be missing. Since the vehicle's intended behavior was to be cooperative, we hypothesized that having a few missing items provided a more perceivable rational motivation for why the vehicle would interject. Thus, if the vehicle's cooperative behavior is performed convincingly the participant could potentially increase their level of trust in the vehicle, and may be more in tune to its gestures. Although this approach provides a reasonable way for preliminary reflection on usability and influence of the perceived vehicle behavior on the user, it also highlighted some shortcomings. We had found that it was imperative that the task-related items and their locations be clearly marked and visible. When they were not, the participants became easily confused or frustrated, regardless of the vehicle's attempt to help. Also, we have found that our users overall reflection was that the seated platform was sometimes inhibiting, with participants often wishing to get out of the vehicle to look for missing items.

Furthermore, the wizard had to be extremely skilled with the joystick control and adapting to the participant's actions. The wizard must be very aware of when the right time to interject is and how much force to apply. If the timing is incorrect or the force is too strong the participant may think that there is a vehicle malfunction. Ultimately there is the very real possibility that the participant will not even understand that the vehicle is trying to hint at anything as opposed to just behaving abnormally.

Finally, we discovered that without an initial process of gaining trust in the vehicle the participant was generally unaware of what the movements could mean. There is some learning by the participant that should take place before the vehicle's interjection movements can become meaningful. The participant must be able to recognize that the movement by the vehicle is intentional, and discern what its intent is. Initially this is unclear and for some participants we discovered that the process is almost bound to fail, unless there is a mechanism that allows them to gain trust in the vehicle.

6 Conclusion

We think that the new emerging area of human vehicle interaction can benefit from the introduction of emotive layers of interaction between the user and the vehicle. We believe that our work on emotive expression through movement can potentially introduce a subtle dimension to how we interact with our vehicles, and might lead to further research that can help strengthen the bond we have with our transportation means. Humans have had personal connections with their horses or camels for centuries because of the simple fact these modes of transportation are alive, and behave as such. The animal exhibits personality, has a mind of its own and expresses emotions, but respects the power of its master. The human is responsible for the well-being of the animal so in turn they both serve each other functionally, as well as, to some extent, emotionally. If we can emulate that same experience with future vehicles we believe that humans could start to develop stronger, more affectionate bonds with their vehicles, similar to those that exist historically with other zoological modes of transportation. We believe that such bonds can improve the commuting experience, make it more interactive, more effective, and ultimately more enjoyable.

The implementation effort we have reported in this short paper is preliminary, and needs to be improved. In the long run we would like to replace the Wizard of Oz with an autonomous vehicle. Furthermore, it would be beneficial to further explore the other proposed behavioral classes we envisioned: *semi-autonomous* and *training or taming*.

We believe that trust will play an inherently important role in interaction with vehicular robots. When a machine is responsible for safely transporting users, taking control of their movement through space, trust becomes essential. Thus, in order to build trust while expressing behavior or emotion through movement, the vehicle must be able to convey the intent of a movement to the user. We believe that users would then be inclined to assign some emotion to the movement of the vehicle, whether it be angry, lazy or confused as long as they are not thinking the vehicle is simply malfunctioning. Unfortunately, especially without familiarity, an unforeseen movement is generally interpreted as an error. We believe that based on our preliminary findings the main challenge facing the design of emotive expression through motion in robotic vehicles is getting over the trust barrier. However, like any relationship this may simply be a matter of time and patience, and repeated exposure to the vehicle may be beneficial to users as it would increase familiarity and hopefully eventually increase trust as well.

References

1. Sheller, M.: Automotive Emotions: Feeling the Car. Theory Culture Society 21, 221 (2004)
2. Sachs, A.: For Love of the Automobile: Looking Back into the History of Our Desires, Berkeley (2002)
3. Nass, et al.: Improving automotive safety by pairing driver emotion and car voice emotion. In: CHI 2005 (2005)
4. Reeves, B., Nass, C.: The Media Equation. Cambridge University Press, New York (1996)
5. Heider, F., Simmel, M.: An Experimental Study of Apparent Behaviour. American Journal of Psychology 57, 243–259 (1944)
6. Harris, J., Sharlin, E.: Exploring Emotive Actuation and Its Role in Human-Robot Interaction. In: HRI 2010 (2010)

7. Kowalski-Trakofler, K.M., Vaught, C., Scharf, T.: Judgment and Decision Making Under Stress: An Overview for Emergency Managers. International Journal Emergency Management 1, 278–289 (2003)
8. Saerbeck, M., Bartneck, C.: Perception of Affect Elicited by Robot Motion. In: HRI 2010 (2010)
9. Drury, J.L., et al.: Awareness in human-robot interactions. IEEE Systems, Man and Cybernetics (2003)
10. Dahlback, N., et al.: Wizard of Oz Studies - Why and How. In: Proceedings of IUI 1993, pp. 193–200. ACM, New York (1993)

Evaluation of an Integrated Paper and Digital Document Management System

Matthew Jervis and Masood Masoodian

Department of Computer Science, The University of Waikato
Hamilton, New Zealand
{mjervis,masood}@cs.waikato.ac.nz

Abstract. Paper documents are still an integral component of our everyday working lives, and are likely to remain so for the foreseeable future. Fortunately, advances in electronic technology in recent years have made it possible to develop digital systems that better integrate paper and digital document management. However, as far as we are aware, none of these integrated document management systems have been evaluated to demonstrate whether the users would indeed find them valuable for managing paper documents. In this paper we present a user study of one such system, called SOPHYA, that provides a means of linking the management of real world document artefacts (e.g., folders) with their electronic counterparts, so that document management activities such as locating, retrieving, and filing documents can be better supported, especially for large collections. The study is in the form of a laboratory-based evaluation, in which participants with experience in document management used the system to perform several tasks, such as locating and retrieving folders, and provided feedback based on their own experience. The results of this study show that users are able to apply software functionality they use for storage and retrieval of electronic documents to effectively manage paper documents as well. Our study has also identified a range of other potential benefits that the users envisage for such a technology in their diverse work environments.

Keywords: Document management, paper documents, electronic documents, digital documents, tangible interfaces, evaluation.

1 Introduction

Paper documents in their various forms, whether as individual pages or as bound books, etc., are used in our work as well as ordinary daily lives. Paper as a medium of communication has a range of affordances that are not easily provided by its digital alternatives. Amongst these affordances are paper's tangibility, flexible navigation, ease of annotation, the possibility of reading and comparing multiple documents at once, and many others that have been identified by Sellen and Harper [11].

Electronic alternatives to paper documents, on the other hand, clearly have their own advantages, including their ease of storage, organisation, search and retrieval. It

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 100–116, 2011.

is therefore not surprising that, at present, paper and electronic documents co-exist in most workplaces. As such, the need for this co-existence has been realized by many, and systems have in recent years been developed to allow some form of integration between management of paper and electronic documents.

The development of most of these demonstrative prototype systems has been made possible by the availability of trackable RFID tags, which can even be incorporated into paper itself. Examples of RFID-based document tracking systems can be found in [1, 2, 6, 8, 9].

Passive wireless technologies, such as RFID, while able to detect when a tag is in the proximity of a reader are limited in the accuracy of their location determination. Also, there are limits to the power that can be drawn by the tags, thus limiting the functionality such a system could provide. Alternative systems rely on using semi-passive or active tags that require a battery, or wired technology. An example of such a wired technology is the SOPHYA paper document management system [3, 5, 10].

Despite the development of these paper document management systems over the past decade, to the best of our knowledge, none of them has been tested by users, either in the context of a usability evaluation or in a real-world work environment. This is perhaps due to the fact that almost all of these systems are demonstrative in nature and do not provide a full range of functionality that can be tested in a realistic setting.

The overall aim of our research has been to evaluate whether an electronic paper document management system can be of potential benefit to users who deal with paper documents as part of their work, and to identify whether such users can manage to use such a system with the same level of ease with which they use digital document management systems.

We have chosen the aforementioned SOPHYA system for this evaluation because it provides for a full range of functionality in terms of paper document storage, search and retrieval, that are comparable to those provided by digital document management systems. The type of study that we have conducted is a laboratory-based evaluation, in which participants with a document management background perform tasks using SOPHYA, and provide feedback based on their experience.

In the next section we will provide a brief overview of SOPHYA, and those of its capabilities that form the basis of our user evaluation. The rest of this paper will then describe the user study we have conducted, along with its findings, and their implications for the design of other integrated paper and digital document management systems. We realise that our laboratory-based user study is not as comprehensive as a longitudinal study of the deployment of SOPHYA in a real world office environment. The aim of our study, however, was to evaluate whether users who are experienced in managing paper documents in conventional offices can easily learn to use SOPHYA to effectively perform the kind of tasks they currently perform without the support of technology. We were also interested in getting feedback from experienced office workers that would guide the future development of technology such as SOPHYA, as well as their deployment in real work settings.

2 Overview of the SOPHYA Technology

Although the architecture of SOPHYA has been designed to allow integrated management of physical and digital artefacts in general, its current prototype implementation is specific to the management of paper and digital documents. SOPHYA has a modular architecture [4] consisting of five layers, three of them (hardware, firmware and middleware) are core components of SOPHYA, while the other two (digital document management system server, and the clients) are application specific.

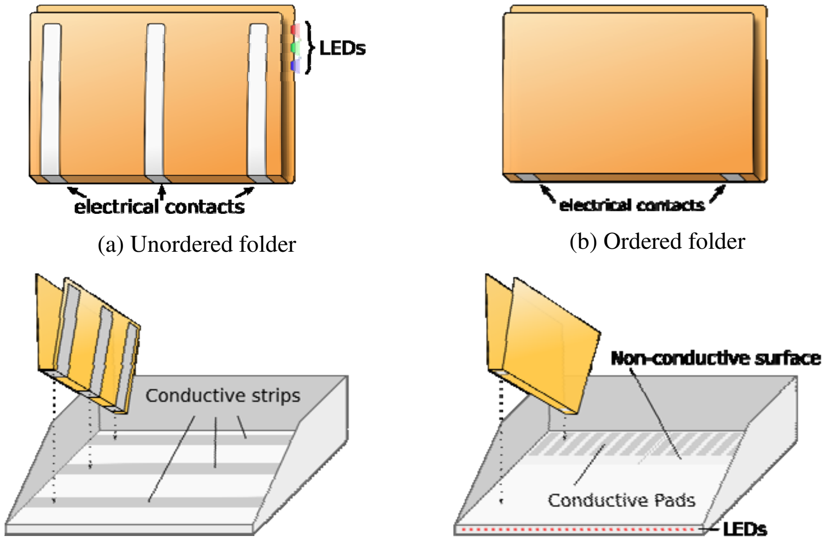
The hardware component of SOPHYA handles the management of individual containers, each of which can hold a collection of documents (e.g. folder, archival box, book, etc.). Containers can in turn be placed in physical storage locations (e.g. filing cabinets, shelves, in-trays, desktop, etc.). Currently there are two different, but fully integrated, implementations of the SOPHYA hardware technology. One of these implementations does not determine the order of the documents placed in physical storage location (referred to as the unordered SOPHYA [3]), while the other provides support for collections that are ordered in a particular manner in each physical storage location (referred to as the ordered SOPHYA [5]).

These two implementations serve complementary purposes. The ordered system is better suited to long-term archival type storage, where there is a need for some natural order in storage of documents (e.g. chronological, alphabetical, etc.). Conversely, the unordered system is able to provide more functionality on the containers (e.g., user interface elements such as LEDs) and supports stacking of folders, making it more suited for everyday use, where documents are moved around and it is much more important to identify their location in temporary storage than to keep some kind of order in their placement.

The containers and physical storage locations are augmented with different electronic components, depending on the type of SOPHYA technology used (e.g. for unordered and ordered see Figure 1). In both implementations, however, each container has a unique ID, and may have optional user interface components such as LEDs. Firmware embedded in the physical storage locations is able to communicate with the containers (e.g. read their IDs and control their user interface components, etc.). This firmware also communicates with the middleware, which in turn is responsible for dealing with data coming from different physical storage locations, and sending the require information to any application specific document management system server using SOPHYA.

These three core layers of SOPHYA are separated, and abstracted away, from the application specific components that can be developed independently to utilise the SOPHYA technology (for details of how this can be done see [4]).

SOPHYA provides a two-way interaction between the physical and digital worlds. From the physical to the digital world, SOPHYA is able to communicate which collection a given document container currently resides in. Conversely, from digital to physical, SOPHYA can, for example, use LEDs on the container or storage location to give visual indication of a container's current location in the physical collections. Figure 2 depicts the levels through which communication passes, from the document management software in the digital world down to the document containers being managed in the physical world, and vice versa.



(a) Unordered folder (b) Ordered folder

(c) Unordered shelf (d) Ordered shelf

Fig. 1. Comparison of unordered and ordered variants of SOPHYA. The ordered variant is able to determine the position of folders on the shelf, whereas the unordered variant can only determine their presence.

For the purpose of the user evaluation reported in this paper, we used an integrated combination of both ordered and unordered implementations of SOPHYA. We also developed a specific client application suitable for the task scenario of our study which utilized SOPHYA. The details of our setup and client application are described more fully later in this paper.

3 User Study of SOPHYA

As mentioned earlier, a growing number of systems are being developed to manage paper documents electronically, with some of them even attempting to integrate the management of paper documents with that of digital documents. Most of these systems tend to replicate the type of functionality that computer systems have for search and retrieval of digital documents to allow similar capabilities for management of paper documents.

There are, however, no substantial user studies of these systems to provide any empirical evidence for their effectiveness in allowing users to manage the storage, search, and retrieval of paper documents using techniques adopted from systems developed for management of electronic documents.

We have therefore conducted a user evaluation of SOPHYA, which has a comprehensive set of functionality for management of paper documents, to demonstrate the effectiveness of a range of document management capabilities it provides for.

Our user study of SOPHYA was in the form of a laboratory-based evaluation. This type of study is clearly not sufficient for demonstrating whether a system such as SOPHYA will in fact be adopted in a real world environment and used long-term. However, the objective of our study was to evaluate the usability of document storage, search, and retrieval capabilities of SOPHYA, and gain feedback from experienced office workers as to whether they envisage potential benefits in adopting it as a tool in their work environment.

3.1 Methodology

Our user study consisted of three parts. It began with the participants following a tutorial in order to familiarise themselves with SOPHYA. This was followed by a set of five tasks—based on the scenario described below—that they had to complete, after each of which the participants were required to fill out a questionnaire. Finally, after the completion of the study tasks and questionnaires, we conducted a semi-structured interview with each of the participants.

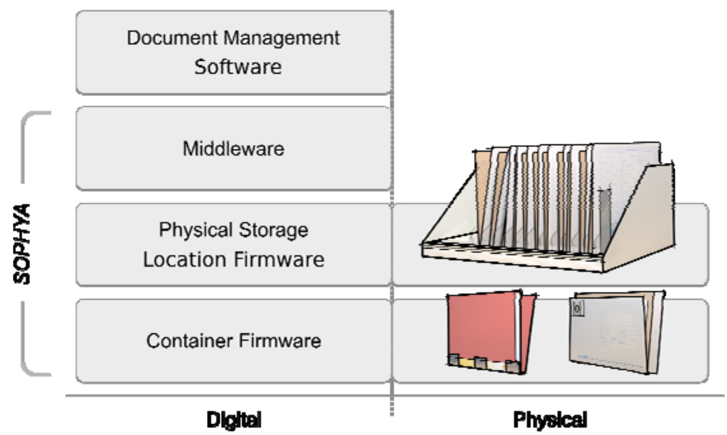


Fig. 2. Integrating the physical and digital document management systems. Physical document containers and storage locations are augmented with SOPHYA circuitry that gives them a digital element.

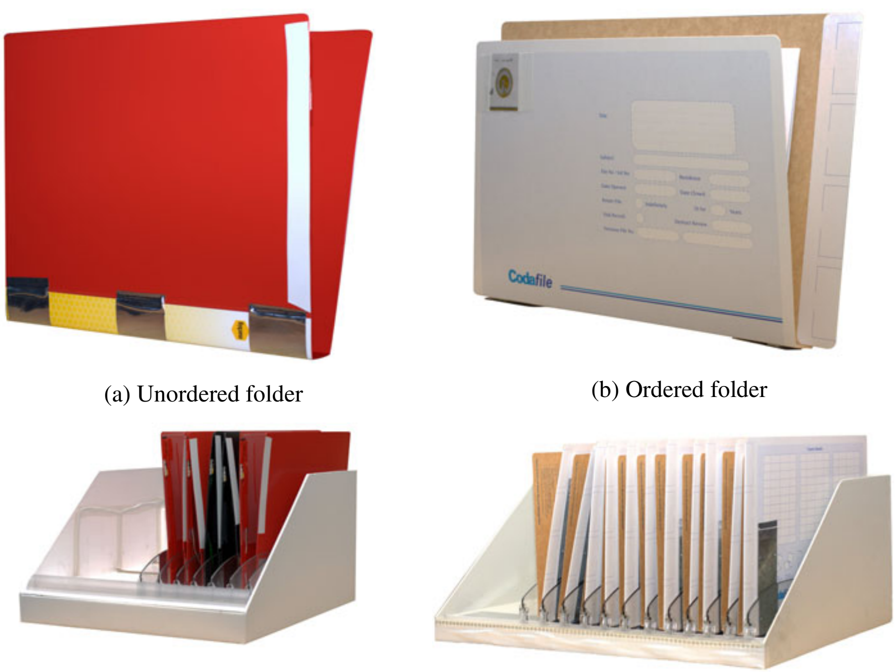
Task Scenario. In order to simulate a real-world office environment we developed our tutorial and study tasks based on a scenario of a business office, which we envisaged as being a potential environment for the application of the SOPHYA technology. The business office we simulated was that of a fictitious landscape design firm, in which the participants played the role of a personal assistant to the manager of the firm. Management of documents in our fictional landscape firm was job-centric. This meant that each design job was assigned a folder in the SOPHYA system, which integrated the electronically-augmented physical folder of the job with its digital representation in the document management system (see below).

Experimental Setup. We conducted the study in a usability laboratory at our university. A desk in the laboratory was set up with a laptop computer running the

SOPHYA-based client application that we developed for this study. The SOPHYA system we used consisted of two augmented shelves of folders, one unordered and one ordered, as well as an RFID tag-reader. Additionally four unaugmented conventional document trays were used for organising documents as part of the study tasks. The desk layout of our experimental setup is shown in Figure 3.

The two variants of SOPHYA (i.e. unordered and ordered) were used in this study for different purposes. Folders of the unordered type (Figure 4(a)) were used to hold documents related to the design jobs that were in-progress in the scenario, while ordered (Figure 4(b)) folders were used for the completed jobs. As mentioned earlier, we envisaged that the unordered version of SOPHYA would be more suitable for managing "piles" of folders and documents that are used regularly while they contain jobs that are in-progress, whereas the ordered version of SOPHYA would be more suitable for managing archival type folders of jobs that have been completed.

Fig. 3. The office setup for the user study



(a) Unordered folder (b) Ordered folder

(c) Unordered shelf (d) Ordered shelf

Fig. 4. SOPHYA folders and shelves used in the user study

Client Software. As previously stated, SOPHYA does not include an electronic document management software, because the requirements of such software would depend on the type of application environment in which it is deployed. Therefore, in order to conduct a user study of SOPHYA we had to develop a software client using the API provided by the middleware component of SOPHYA.

It should be noted that the aim of our study was to focus on the functionalities provided by SOPHYA, and not the usability of our client application. Therefore, we made the client application itself very simple in terms of its design, using it to provide the type of functionality we needed as part of our study, and tailoring it specifically to the task scenario we have described. We also used the client application to collect user study data (e.g. task completion time, etc.), as well as for presenting the task questionnaires to the participants and collecting their responses.

Perhaps the most important functionality that the client application provided was to allow the study participants to search for information about the design job folders and the documents they contained based on a number of different search criteria (e.g. keywords, dates, job types, etc.). Figure 5 shows the search screen of our client application. Once a user had completed a search activity, the client allowed the user to locate the relevant folders and documents by visually identifying them in the folder shelves (e.g. by turning their LED on to a specific colour, etc.).

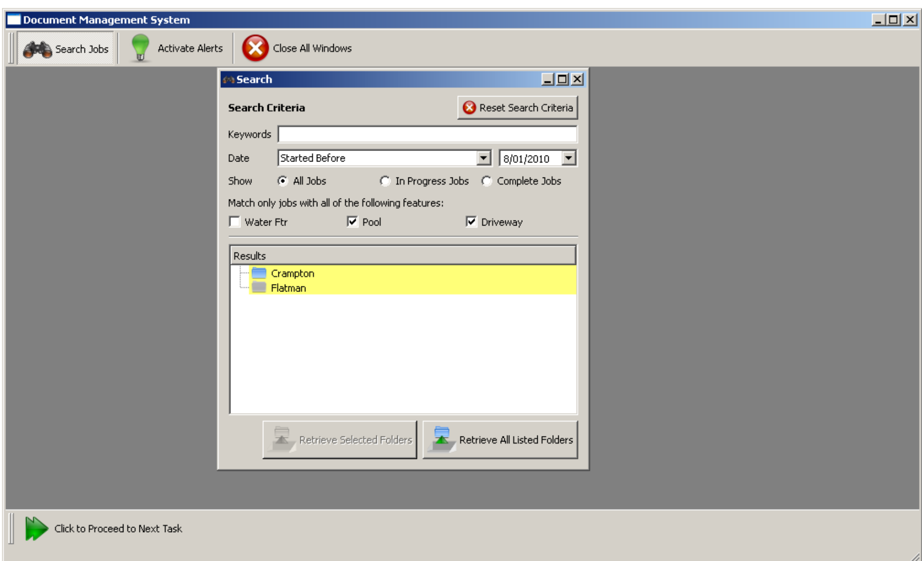


Fig. 5. The client application developed for the user study, showing the search dialog

Tutorial. The study participants completed a short, simple tutorial before starting on the main study tasks. The tutorial described the fictitious design firm, identified different components of the experimental setup, and showed the basic functionality of the client application, which included searching for folders, locating and retrieving folders, and returning folders to one of the two shelves depending on whether they were jobs in-progress or completed. The tutorial on average took around 11 minutes to complete.

3.2 Tasks

The participants were asked to perform five tasks, each consisting of several steps. The aims of these tasks were to get the users to utilise SOPHYA to perform a range of *filing* activities which are normally carried out in a typical office environment. These tasks are described below.

Retrieving Specific Folders. The first two tasks aimed to demonstrate the link between electronic search and physical retrieval. For each of these tasks, the participants were required to retrieve folders matching a given date criteria. The first task was to retrieve folders of two jobs in-progress (from the unordered shelf), while the second task was to retrieve the folders of two completed jobs (from the ordered shelf).

Collating Specific Documents from Different Folders. The third task required the participants to retrieve several completed job folders based on various criteria and find a specific document in each of them. The participants then needed to place these documents into a new folder, and return the original folders back to their shelf. Although the task sheet provided clear steps to achieve this task, the task required some cognitive effort on the part of the participants as they had to retrieve the folders in multiple steps.

Returning the folders to the shelf required using SOPHYA to guide the placement of folders on the shelf. The folders used in this task were for completed jobs (from the ordered shelf), and also had RFID tags attached to them. Placing these folders over the RFID tag reader instructed SOPHYA to turn on LEDs in the position where the folders belonged on the shelf.

Combining Paper and Digital Documents. The aim of the fourth task was to demonstrate how SOPHYA could be used to integrate the management and use of both digital and paper documents. This task required the participants to browse the electronic contents of a given folder to find a specific digital text document. They then had to retrieve certain paper documents from their physical folder, and compare these paper documents against a list of changes in the digital text document.

Generating Alerts. The final task aimed to demonstrate how the folder LEDs could be used to provide visual alerts when certain conditions were met. Two types of alerts were demonstrated, a time-triggered alert, and an action-triggered alert. As this task was intended as a demonstration of the alert feature it did not require the participant to actually setup alerts, but rather asked them to make use of alerts that had been previously setup.

The time-triggered alert caused the LEDs of all the in-progress job folders that had not been accessed in the past seven days to be turned on. The event-triggered alert, on the other hand, turned on the LED of a completed job folder when a document that had to be returned to its folder was placed on the RFID tag-reader on the desktop.

3.3 Questionnaires

After the completion of each task the participants were automatically presented with a questionnaire in the client application, which they had to fill out before being allowed to proceed to the next task.

The questionnaire consisted of five questions, which were repeated for each of the five tasks the participants performed. Each question had to be answered in two parts: a rating part on a scale 1–7 (1 being very difficult and 7 being very easy) and a why part, asking for the reason for the rating. The five questions of the questionnaire are shown in Table 1.

Table 1. Tasks questionnaire

No.	Question
Q1.	How easy was it to understand the task you have just performed?
Q2.	How easy was it to perform the task using SOPHYA?
Q3.	How easy would it be to perform this task again using SOPHYA?
Q4.	How helpful was SOPHYA in assisting you to perform this task?
Q5.	How easy would it be to perform this task manually (without SOPHYA)?

3.4 Participants

Selection of the participants for this study was carried out very carefully to make sure that the chosen participants had sufficient real office experience of dealing with paper documents and were familiar with the process of filing, search, and retrieval of folders and documents. The main reason for these selection criteria was to have study participants who could provide us with some level of comparison between paper document management tasks performed in our study using SOPHYA and their real-life experience of doing such tasks in their normal work environments.

In total 16 participants were selected for this study, of whom 10 were female and 6 male. The participants were mostly from various units across our university. Thirteen of the participants worked in offices where they dealt with paper documents and files on a daily basis, one participant had previously worked as a librarian, one was a researcher in the field of knowledge management, and one worked solely with digital documents.

3.5 Results

Table 2 shows the results of task accuracy and average completion time for each of the five study tasks. Considering the fact that the participants were not familiar with the organisation of the documents and folder, and had to rely totally on SOPHYA to perform the tasks, the average times taken to do the tasks were between 2 to 6 minutes.

The accuracy of tasks completions were high for most tasks, except for Task 2. In fact most of the errors in completing the tasks were caused by what could be considered "clerical" errors rather than being caused by SOPHYA.

Table 2. Percentage of participants who completed each task accurately (e.g. retrieved the correct folders) and average time taken to complete each task

Task	Accuracy (%)	Time (mm:ss)	
		Avg.	Std. dev
T1.	94	02:35	01:38
T2.	50	02:55	00:27
T3.	81	05:46	01:18
T4.	94	04:23	01:15
T5.	100	03:04	01:41

For instance in the case of Task 2, the task asked the participants to retrieve folders for jobs completed before a given date in 2009. However, 10 of the 16 participants missed this detail given in the task specification and entered criteria for the correct day and month, but in 2010. Thus, rather than the search resulting in two folders, 12 folders were returned. Upon seeing such an unexpectedly large number of results, 3 of the participants re-read the instructions and re-entered the criteria, while seven proceeded to retrieve all 12 folders.

For Task 3, the encountered difficulties were mainly due to the participants either not realising that two separate searches were required to retrieve the required folders, or not clearing the criteria in the search form between searches.

Table 3 shows the average ratings for each of the questions of the questionnaire for each of the five study tasks. The participants rated SOPHYA very highly in relation to ease-of-use, re-use, and the assistance it provided (Questions 1-4). They also felt that based on their experience performing the study tasks would be reasonably difficult to do without SOPHYA (Question 5). The main reasons that participants gave for their ratings to Question 5 were:

- SOPHYA saves time over manually looking through all the files;
- SOPHYA provides the ability to search the database for folders matching given criteria; and
- it is easier to find folders with SOPHYA when unfamiliar with the filing system.

Table 3. Median (left column) and inter-quartile range (right column) of ratings given for the questions of each task questionnaire

	Questionnaire Ratings									
	Q1.		Q2.		Q3.		Q4.		Q5.	
T1.	6.0	*0.640*	7.0	*0.063*	6.5	*2.000*	6.0	*1.300*	7.0	*1.000*
T2.	6.0	*1.300*	7.0	*0.047*	7.0	*1.000*	7.0	*1.300*	7.0	*1.000*
T3.	7.0	*1.000*	7.0	*0.000*	7.0	*0.480*	7.0	*1.000*	7.0	*0.000*
T4.	6.0	*2.300*	7.0	*0.480*	7.0	*1.000*	7.0	*1.000*	6.5	*1.000*
T5.	4.0	*2.300*	3.5	*3.000*	3.5	*2.000*	4.0	*2.000*	4.0	*2.000*

3.6 User Interviews

Although the results of the task completion time and accuracy, as well as the ratings of the participants were very positive, perhaps the most important findings of our study came from the interviews conducted with the participants after the study tasks were completed.

The interviews we conducted with our study participants were semi-structured. To guide our interviews we used a set of open-ended questions that we followed broadly. These questions (see Table 4) can be divided into three groups: those relating to the background of the participants and their previous office experience, those relating to the usefulness of SOPHYA, and those relating to its adoption in real world.

Table 4. Interview questions

Question
Background
Do/have you worked with paper documents in your work?
How did/do you manage them?
How large was the collection of documents?
Would others use the documents/files as well?
How would you collaborate/share the documents/files?
Did you deal with sensitive files?
How was security implemented? How did you control access to files?
Usefulness of SOPHYA
Based on your experience, do you think a system such as SOPHYA would be useful in the offices you have worked in?
Why/why not?
Can you see any other situations where it might be useful?
Barriers to adoption of SOPHYA
Can you see any barriers to using such a system?

In the following sections we have grouped the feedback we got from the participants into a number of categories.

Ease of Use. Most of the comments we got from the participants supported their positive ratings of SOPHYA's ease of use. The participants commented on how easy or intuitive it was to use SOPHYA. While, the simplicity of the tasks and the design of the client application specifically for this evaluation (as opposed to general purpose document management software) may have had some bearing on the perceived ease of use, it is likely that many of the features of SOPHYA would remain easy to use even in a more complex setting. Some of the feedback related to ease of use included:

– *"With that [SOPHYA] you just find where everything is and where it should go."*
– *"Because the light was flashing it completely removed any kind of thinking process and seemed to speed up the file retrieval process hugely."*
– *"It seems to me very simple, find it, there goes a light, swipe it, put it back; it's pretty intuitive in that regard."*
– *"It's very easy to use once you've got the hang of it. And the more you used it the easier it is."*

Usefulness of SOPHYA in Offices. Six of the 13 participants who work in an office with paper documents on a daily basis, said that SOPHYA would be useful in their current or previous office. The seven other participants that did not think SOPHYA would be useful in their office provided a range of reasons for this. They generally felt that their office was not suited to using SOPHYA or that their current system was serving them well enough. For instance they said that their office was too small to need such a system, or that they already knew their current system well enough that there would be no gain to using SOPHYA in their office. However, one participant who said that he knew his own system well enough also stated that SOPHYA would be useful when retrieving documents from his co-workers if they used SOPHYA.

The study participants were also asked if there were any other types of offices in which they thought SOPHYA technology would be useful. The most common suggestion was that it would be useful for large paper-based offices, such as legal offices and accounting firms. There are also other situations where there are large, unwieldy documents such as diagrams or maps, which are not suitable to have in digital format. Additionally, it was suggested that SOPHYA would be useful in offices where there are many casual or temporary staff, as SOPHYA does not require familiarity with the filing system in order to use it.

One of the participants had experience as a librarian and pointed out that SOPHYA would be very useful in a library. She mentioned that *"once [a book] is misfiled it's useless, so if this can take out misfiling things it is very appealing"*. This confirms the research findings of others [7] who have noted that library patrons often have difficulty locating books. A system such as SOPHYA, designed to integrate management of physical and electronic artefacts, would be useful in a library. Its use would also simplify the process of shelving, reducing the training required, and allowing skilled librarians to concentrate on other tasks.

Another suggestion, though not specifically related to an office environment, was that a system such as SOPHYA would be valuable in the aviation industry. The participant described a system used for storing spare parts, where the parts are allocated randomly to bins (rather than in sections, e.g. screws, bolts, etc.) in order to reduce the chance of the wrong part being retrieved, as there are many similar parts. He stated that SOPHYA would be perfect in this scenario, as it would allow a search for the part to be performed electronically, and then a light could indicate the appropriate bin where the part could be retrieved from. We have already stated that SOPHYA has in fact been designed for management of physical artefacts in general, and therefore this type of application would be suitable for it.

Potential Benefits of SOPHYA. One of the major benefits mentioned by the participants was that SOPHYA would save time when retrieving or returning documents. For some it was an alternative to *"fumbling through"* their folders in order to find the one they needed. Having the LED to display the folders to be retrieved, especially if it could be coupled with tagging of documents, would save time.

The integration of searching with display of the search results on the folder LEDs was also mentioned by participants as a benefit. Several participants pointed out, simply looking up the position of a single folder may not be of great benefit, but integrating it with search makes it a useful tool. As one participant said: *"If it was just 'oh where should I put this file' then it's probably not really going to be that successful, people are going to say 'it's just as quick to find it alphabetically,' but when it's connected to searching for certain things it is really good."*

Another potential benefit mentioned is ensuring that folders are returned to the correct place once people have finished using them. This is especially pertinent in a larger office where multiple people share files and the filing schemes may be complex. Two of the participants described cases in which the people in their office had returned folders to the wrong place. *"Often with that amount of files, go to look for a file and someone has used it and has put it just a little bit out of place; because either they've transposed the numbers in their mind as they've put it back or put it according to the numbers, but then we have special files that are not filed by that numerical number, they have their own separate shelf as well because they are special projects."* With traditional systems once the folder is misfiled someone has to manually scan through the folders in order to find it rather than just relying on, for instance, the numerical or alphabetical ordering. Similarly, some of the participants considered SOPHYA as technology to prevent document loss. It was also suggested by one participant that by making the task of filing easier it would change people's filing behaviour so that they would be more organised.

Another participants talked about wanting to keep her documents in order, so as to feel in control of them; *"however, with so many people having access to them that it was not possible to control them. But having a system like SOPHYA would remove the need to keep them in order as the LEDs could show their position."*

The ability to locate a file that has been removed from the collection (e.g. filing cabinet) was seen as another benefit of SOPHYA. *"There's always emails going around 'has anyone seen the file for such-and-such'." "With the files I deal with sometimes I'll have to, say for example, get sign off from another manager. So I will give the files to them. In my tracking spreadsheet I will put down—so I know where it is—'have put on bla-bla-bla's desk'."* One participant suggested that requiring the user to swipe their ID card when retrieving a folder from SOPHYA would mean that the system could keep track of who has taken which file.

Making it easier for people to use a filing system when they were unfamiliar with it was another benefit mentioned for SOPHYA. The fact that people become very familiar with the files they work with everyday is well known. This is something that is demonstrated when people try to use a filing system they are unfamiliar with: *"You really notice how complicated your filing system is or isn't when you have new staff. They just have no idea, can't find anything."*

The alert feature of SOPHYA was also considered valuable by a number of participants. One participant described the alert system in use in their office, effectively an electronic listing of things to do during the week. However, while this lists what needs to be done in the coming week, if they are busy they may not get to them. The participant mentioned that she would be more inclined to pay attention to flashing lights on the folders than electronic reminders: *"if there were flashing lights on my rotascan I would so deal to those quickly before anyone else saw them. You're sitting there and your boss comes in and sees lots of flashing lights on these files, they're gonna know you've been slack."*

SOPHYA alerts are in fact more peripheral than electronic alerts on computers. Alerts on the computer are considered more *"in-your-face"* and could be easily closed; whereas, it was suggested that if one was required to walk over to the folders in order to acknowledge the alert, then *"I might as well pull the file out now that I have to go all the way over and touch it."*

The ability to identify folders that have not been used in a certain period of time was also seen as being useful. For example: *"sometimes you're so busy with other students, ones that are knocking on the door all the time, you forget about the ones that don't even bother to turn up to anything. Having some kind of system like that so you can keep on top of everything would be awesome."*

It was also pointed out that SOPHYA could reduce duplication, because if it was easier to find documents that are already in a shared archive then people may no longer feel the need to keep their own personal copies. *"It was so much of a pain to find the contracts I ended up printing them myself. Every year it must have been over a 1000 pages of contracts and other stuff that was all in the files but it was just such a performance to find it some of the times that it was just easier to print it out."*

Barriers to Adoption of SOPHYA. In order to get some idea of the kind of problems that may be encountered when attempting to deploy a system such as SOPHYA, participants were asked what barriers to implementation they could foresee. The participants provided us with a range of responses that are summarised below.

Possibly the biggest barrier described by the participants was the cost. The cost/benefit ratio would have to make it worthwhile to deploy such a system, balancing the setup and maintenance costs with the savings in time. As one participant noted: *"being inefficient might prove to be cheaper, even though it might be more frustrating and annoying."* For instance, if a large number of documents have to be tagged with RFID tags for use with SOPHYA, then this could incur a large overhead. *"You'd have to see if doing it this way uses less energy overall compared to your traditional style of office maintenance which is put it in a pile somewhere then physically search through it later."*

It was also pointed out, that if the organisation in which SOPHYA is deployed is not itself organised then just adding new technology is not necessarily going to solve the deeper problem. *"But that's like any technology. If you're not good at running meetings then doing teleconferences isn't going to help you. Your technology emphasises what you're good or bad at rather than fixing something necessarily. But for those ones that do have systems and do have policies around that sort of stuff I'd imagine they could almost use it straight away."*

Another potential problem that was mentioned is that this type of automated system is only as good as its data entry. If information about documents is entered incorrectly into the system then it would make it more difficult to find documents.

Additionally, the staff who have to use the system may resist it if it threatens them: *"with something new, depending on how it was introduced to the staff that had to run it [may be a barrier]."* It was suggested that if the staff do not like technical advancements, or do not want change they may become a barrier to its adoption. Similarly, training is another overhead that would be required. Although this study has shown that learning to use SOPHYA is not likely to be very difficult or time consuming.

The risk of the system breaking would be yet another barrier. The system would need to be reliable, especially if using such a system resulted in staff with less filing training using it: *"then if something went wrong they may not notice a problem or not be able to deal with it."*

Finally, personalisation was also seen as a potential pitfall: *"Everyone might have a different way of dealing with their files. That could be a barrier in terms of trying to find a way to personalise it to every person."*

4 Discussion

Our study has provided us with many valuable insights into the effectiveness of SOPHYA as an integrated paper and digital document management system. The study has indicated that a system such as SOPHYA would be beneficial if deployed in real offices. In particular, the offices that would benefit most from SOPHYA would be larger offices, with more complex filing and document management requirements. Examples of such offices include those with a large number of paper documents, or a large number of staff who share access to the paper document management system. Systems such as SOPHYA would also benefit offices with a lot of staff turnover, inexperienced or temporary staff who would need to access documents without having to get considerable amount of training on the organisation of the manual document management system. However, in such cases where people accessing the system are less experienced they are likely to be much more dependant on the system, and thus would require very high levels of reliability from the system to be able to perform their document management activities.

Beyond the obvious benefit of providing for faster and easier search and retrieval of documents and folders, the potential benefits of systems such as SOPHYA for large and complex offices include the possibility of integration with organisational database systems, providing linkage between online and offline documents, assisting with maintaining consistency and accuracy of document archives (e.g. by helping the users to put folders in the right place, or finding out when they are not in the right place), providing alert mechanisms which could be used to maintain the security of paper documents, as well as allowing users to manage the usage and tracking of documents across the organisation, and so on.

There are also several potential barriers to adoption of paper document management systems such as SOPHYA. Those identified by our study include concerns for the cost of such system, staff resistance to its introduction, the need for personalisation of filing systems, the risk of user dependency and lack of effective access to documents in cases of system break down, and the heavy reliance of such systems on good organisational disciplines for maintaining accurate information about documents in the system. It should, however, be noted that most of these concerns and potential barriers are often expressed before the introduction of any computerised system into organisations that deal with manual or paper-based systems. Clearly unless a system such as SOPHYA is introduced into an organisation and used over a considerable period of time it would be impossible to find out to what extent these kinds of barriers and concerns are serious and cannot be improved upon over time.

One limiting factor to the use of SOPHYA which needs to be addressed is that SOPHYA deals with the management and tracking of folders. Although this level of granularity might be enough for some offices it would not be sufficient for others. Some offices may require document tracking as well as folder tracking, for instance by using RFID-tagged documents in combination with SOPHYA. In such cases, however, in order to keep the cost of tagging down, document tracking may only need to be applied to important or sensitive documents.

There are already several document level tracking systems proposed by others, and referred to earlier in this paper. We envisage that any comprehensive paper document management system would be a combination of wired (e.g. SOPHYA) and RFID-based systems. The main advantage of SOPHYA over purely RFID-based systems is in its ability to provide for a two way interaction—whereas document tracking with RFID is limited to one-way tracking of the location of documents, with SOPHYA there is also the possibility of two-way communication with the folders and storage locations.

5 Conclusions

In this paper we have presented a user evaluation of SOPHYA, which is an integrated paper and digital document management system. Although the study was based on using SOPHYA, many of its findings are equally relevant to other physical document management systems.

We would, however, like to point out that the real success of paper document management systems would be dependant on the range of applications that seamlessly integrate the digital and paper document collections and information systems. At present, due to the lack of reliable underlying technology, such as SOPHYA, the range of client level applications is limited or they are completely lacking. Further research is needed to guide the development of such interactive applications and information visualisation techniques to fully utilise systems like SOPHYA.

Acknowledgments. We would like to gratefully acknowledge the contributions of those who took part in our user study. The user study reported here was approved by the Ethics Committee of the Faculty of Computing and Mathematical Sciences, the University of Waikato.

References

1. AbuSafiya, M., Mazumdar, S.: Accommodating paper in document databases. In: DocEng 2004: Proceedings of the 2004 ACM Symposium on Document Engineering, pp. 155–162. ACM, New York (2004), `http://portal.acm.org/citation.cfm?id=1030428`
2. Arregui, D., Fernstrom, C., Pacull, F., Rondeau, G., Willamowski, J., Crochon, E., Favre-Reguillon, F.: Paper-based communicating objects in the future office. In: SOC 2003 Smart Object Conference (2003), `http://citeseerx.ist.psu.edu/viewdoc/download?doi:10.1.1.2.2920&rep=rep1&type=pdf`
3. Jervis, M., Masoodian, M.: Digital management and retrieval of physical documents. In: TEI 2009: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 47–54. ACM, New York (2009), `http://portal.acm.org/citation.cfm?doid=1517664.1517682`
4. Jervis, M.G., Masoodian, M.: Design of a modular architecture for integrating paper and digital document management systems. In: PaperComp 2010: Proceedings of the 1st International Workshop on Paper Computing, Part of the 12th ACM International Conference on Ubiquitous Computing, pp. 44–47. ACM, New York (2010)

5. Jervis, M., Masoodian, M.: SOPHYA: a system for digital management of ordered physical document collections. In: TEI 2010: Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 33–40 (2010), http://portal.acm.org/citation.cfm?id=1709894
6. Kojima, H., Iwata, K., Nishimura, N.: Document management and information organizing method using RFID tags. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4553, pp. 601–610. Springer, Heidelberg (2007), http://portal.acm.org/citation.cfm?id=1769451.1769523
7. McKay, D., Conyers, B.: Where the streets have no name: how library users get lost in the stacks. In: CHINZ 2010: Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, Auckland, New Zealand, pp. 77–80 (2010)
8. O'Neill, J., Grasso, A., Willamowski, J.: RFID: enhancing paper documents with electronic properties. In: CSCW, Computer Supported Cooperative Work, Workshop Collaborating Over Paper and Digital Documents (2006)
9. Raskar, R., Beardsley, P., Dietz, P., van Baar, J.: Photosensing wireless tags for geometric procedures. Commun. ACM 48(9), 46–51 (2005)
10. Seifried, T., Jervis, M., Haller, M., Masoodian, M., Villar, N.: Integration of virtual and real document organization. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, pp. 81–88. ACM, New York (2008), http://portal.acm.org/citation.cfm?id=1347410
11. Sellen, A.J., Harper, R.H.R.: The Myth of the Paperless Office. MIT Press, Cambridge (2003)

BendFlip: Examining Input Techniques for Electronic Book Readers with Flexible Form Factors

Doug Wightman, Tim Ginn, and Roel Vertegaal

Human Media Lab, Queen's University, Kingston, ON K7L 3N6, Canada
{wightman,ginn,roel}@cs.queensu.ca

Abstract. We present recommendations for the design of flexible electronic book readers, based on an empirical evaluation of form factors and input techniques in a page navigation task. We compared capacitive touch, pressure, and bend sensors between rigid and flexible form factors using a prototype electronic book reader. Results suggest that the time required to perform bend techniques is comparable to button techniques for page navigation on flexible form factors. We also found that a bend technique had fewer errors on flexible form factors. We discuss the physical affordances of flexible e-book form factors, and why they might be preferable to rigid designs. We conclude by presenting recommendations for the design of future flexible e-book readers.

Keywords: Flexible displays, electronic book readers, bimanual input.

1 Introduction

For some time, researchers have advocated for electronic book readers that preserve more of the physical affordances of paper documents [15,21,25]. According to O'Hara et al. [17], paper document navigation supports the development of incidental knowledge of the location of information within documents. Sheets of paper are lightweight displays that can be navigated by physically moving them. Sellen and Harper describe some of the disadvantages of Graphical User Interfaces (GUIs) [21]. When compared to paper documents, GUIs feature input that is indirect, often one-handed, and heavily dependent on visual as opposed to tactile cues. This can make the interleaving of multiple tasks across several displays cumbersome.

In recent years, the development of e-ink displays [6] has solved some of the above issues. Electronic ink displays feature a paper-like high contrast, and low power consumption, in a relatively lightweight form factor. However, most e-ink products still feature form factors that are much thicker and heavier than paper documents. The current generation of e-ink e-book readers, such as the Amazon Kindle [1], cannot yet be seen as a replacement technology for paper documents.

More recently, flexible display technologies have been developed (see Figure 1) [18]. These displays are compelling for future human-computer interaction design because they can provide more of the physical affordances of paper:

1. These displays are potentially as *lightweight* as paper. Users can hold these displays in one hand for extended periods of time without much effort. The Kindle DX, an e-ink e-book reader, by contrast, weighs more than a pound.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 117–133, 2011.

2. These displays are potentially as *thin* as paper. One could stack multiple displays on a real-world desk, allowing for tighter interleaving of tasks across multiple displays, as is seen in interactions with paper [21].
3. Thin, flexible displays can feature *input techniques* that are similar to those used in paper documents. Flexible displays, like paper, bend. This capacity to bend provides an opportunity for explicit input [19,20]. Since bending affects the shape of the device, it provides tactile, kinesthetic as well as visual feedback.
4. Flexible displays are *mobile*. A flexible display can be deformed to fit inside a pocket. A large flexible display might be rolled up to fit in a backpack.
5. The *screen real estate* of a flexible display, like paper, can be *easily varied*. There are benefits to using a device that can have a large screen only part of the time. A paper road map can be opened for viewing large distances and then folded to provide a smaller form factor that contains all the information required for the next couple of hours of driving.

In this paper, we examine how we might provide input to future handheld flexible devices, such as e-book readers. We evaluated form factors and input techniques for flexible devices, with a focus on tasks found in e-book readers. We present an empirical evaluation of input techniques for the most frequently occurring task on e-book readers: single page navigation. In our experiment, we compared the use of capacitive touch, pressure sensors (buttons with varying force), and bend sensors for navigating single pages on an e-book reader with both rigid and flexible form factors. Results indicate that, given a flexible form factor, bends have comparable performance to buttons, and can have fewer errors. The performance of some input techniques is also notably comparable between rigid and flexible form factors. We conclude with recommendations for the design of future flexible e-book readers.

Fig. 1. E-ink Flexible Display

2 Related Work

Balakrishnan et al. [2] discuss how they used ShapeTape, an optical bend sensor, to model curved surfaces on a computer. Flexible input techniques can operate in many dimensions. This makes it possible for users to twist and bend a device in ways that potentially activate more than one sensor at a time. They also warned of an "iron horse effect" when attempting to bring affordances of real world devices into digital systems: in designing a flexible electronic book reader it is important to copy only those properties of physical books that are desirable in some way, and to avoid unnecessarily copying the disadvantages of physical books.

TWEND was a hardware prototype developed by Herkenrath et al. [11] that allowed complex navigations using twists and bends. TWEND was constructed out of 8 optical bend sensors in order to be able to recognize a wide variety of contortions. While they did not provide an evaluation, they postulated that: "New users consider the page flipping and page scrolling gestures more efficient than techniques found in PDAs and mobile phones."

Lee et al. [13] conducted a user study of deformable materials. In their experiment, they asked users to create and perform gestures for 11 different tasks using thin plastic sheets, paper and an elastic cloth. They used paper to take advantage of untrained users examining what gestures they might find interesting for future flexible display input. They found a significant preference and time difference in how long users took to plan their gestures between cloth and plastic with elastic cloth having the highest preference and lowest planning time. They also found the left hand played mainly a supporting role, while the right hand performed most interactions on the middle of the right side of the material. Surprisingly, users were more consistent in the development of their gestures with materials that allowed more freedom, such as elastic cloth, than with materials that allowed less freedom, such as thin plastic.

Wobbrock et al. [27] emphasized the need for mobile devices without styli and proposed the use of back-of-device navigation. Amongst their findings are that the index finger on the back of the device is relatively worse for complex gestures than it is for simple moving and pointing tasks. Thumbs and index fingers performed similarly on the front of the device, but performance of thumbs decreased when moving to the back, while performance of index fingers did not. Additionally, their study showed that for one-handed and two-handed input, error rates mimicked movement time results. The best performance was achieved with the index finger on the front of the device. Some potential benefits of back-of-device navigation include not obscuring the screen or altering natural gripping or holding posture, and allowing users to grasp the bulk of the device rather than simply the lowest third of the device. Some of these benefits could apply to flexible electronic book readers.

Baudisch and Chu [3] investigated back-of-device navigation on very small touch devices. They found high accuracy across small screen sizes (2.4" and smaller), including in situations in which front of device navigation with a touch screen fails due to occlusion. They found that it was important for users to still be able to see part of their hand, so as to predict the location of their fingers on the back of the device. An electronic book reader could take advantage of this same property by having buttons or controls on the back of the device near the edges, where most of the hand remains visible.

Wilson and Landoni [26] evaluated the page navigation features of four electronic book reader applications. Their design recommendations for future electronic book reader applications include simple bookmarking techniques that adhere to the paper book metaphor. Careful design of button and dial placement for page navigation was found to improve performance. In their evaluations, users of devices that employed dials commented that they felt they could read faster. Simple “page forward/page back” buttons were also considered intuitive, but button size mattered.

Harrison et al. [10] describe a user interface for document navigation applications that used pressure sensors in the top left and right corners of the display, to detect flicking gestures. It did not sense bend gestures. Paging through the document was achieved based on the direction and location of presses. Despite simulating the interaction technique of physical books, this technique is not necessarily the most efficient for page navigation. There are a couple of reasons for this: first, the movement required to complete the gesture takes more time than the theoretically most efficient technique, an in-place tap. Second, the pressure sensors they used required a certain amount of force to be triggered, which caused enough resistance for users to report they were not as lightweight as flicking through paper pages.

Chen et al. [5] discuss the use of flipping and fanning navigational techniques for electronic book readers. In their system, two small displays were joined together through a hinge, providing some of the physical affordances of a paper book. Their interface mimicked the turning of physical pages. Flipping was used for small navigations within the document (such as between pages) and fanning was used for large navigations (such as between chapters). However, real paper is extremely lightweight, making physical page turns rather effortless. Their techniques required gross motor movements to fold the entire hinged display. The authors indicated that they might have attempted to borrow “too literally” from book interactions, particularly given the size and weight of their device.

Gummi is one of the few projects to propose actual physical deformation of a display as a navigational technique. Schwesig et al. [19, 20] designed a mockup of a small credit-card sized computer. Navigation was achieved through bending the display, e.g., to zoom in and out of information. Similar in architecture to BendFlip, the interface was implemented using a rigid form factor display and a flexible sheet of acrylic, augmented with pressure sensors. Gummi was not designed as an electronic book reader and the authors did not report on its use for page navigation.

There is limited prior work on page navigation with flexible devices. In PaperWindows, Holman et al. [12] simulated the use of flexible display surfaces for page navigation by augmenting real paper with projected images. This was accomplished by tracking the shape and location of paper documents using computer vision. Pages could be navigated by flipping over a sheet of paper. This action required considerable gross motor activity, thus reducing its efficiency.

Similar in nature, Watanabe et al. [24] discuss Bookisheet, a set of flexible input devices made out of sheets of thin acrylic augmented with bend sensors. Bookisheet could simulate the turn of pages through bends. Bookisheet changes between discrete jumping and continuous scrolling modes based upon the degree of bend between the two pages that compose the device. Bookisheet also features “finger-bookmarking”, which provides a temporary bookmark while a sensor is pressed.

Pressure Sensors and Haptic Feedback. One issue with mobile display devices is that they are often used in environments such as busses or trains that are bumpy. Small buttons can be difficult to press in such environments. Brewster et al. [4] compared standard buttons to buttons with tactile feedback for a text entry task on a mobile device. Tactile feedback was found to increase the amount of text entered, reduce error rates (by 22.4% in the lab and 25.7% when mobile) and cause users to correct more errors (48.3% in the lab, 66.9% when mobile) compared to the same device without tactile feedback. Only rigid devices and input techniques were investigated. Luk et al. [14] investigated lateral skin stretch for mobile applications. Tactile feedback technologies, including lateral skin stretch hardware, could also be applied to flexible devices.

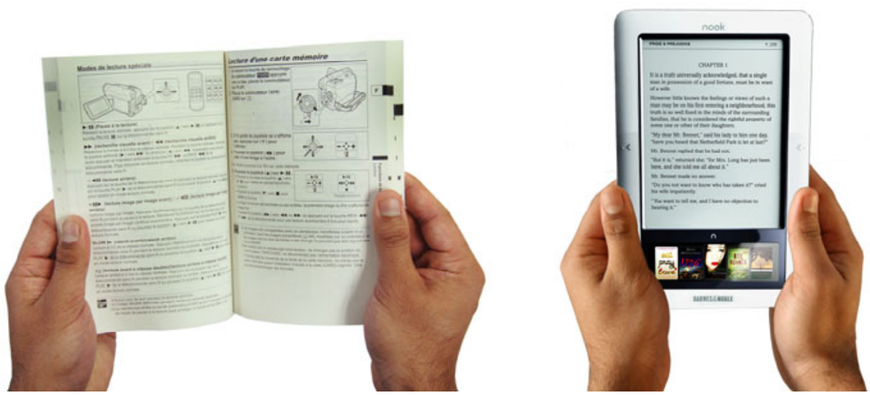


Fig. 2. Example hand positions when holding a paper book and e-book reader

3 Design Rationale

To educate ourselves on the requirements for the design of a flexible e-book reader, we informally observed individuals reading paper books as well as using two commercial e-book readers, the Amazon Kindle [1] and the Barnes and Noble Nook [16]. We were particularly interested to learn the ways in which users navigated single pages, which we found to be the single most frequently performed task while reading. In paper books, page turns are typically accomplished by turning the page through picking up the corner, less frequently by flicking the page up using the thumbs on the side of the page. Both gestures involve gross motor movements of the hand. By contrast, electronic book readers offer only buttons on their bezel for single page navigation. An advantage of these buttons is that they can be used to tap in place and do not require gross motor activity.

From our analysis of the literature and user observations we can identify a number of key design factors of e-book reader form factors that we believe are closely tied to the performance of *flexible* e-book reader designs.

3.1 Size and Portability

Size, of course, is particularly relevant in mobile scenarios, where users should not only be able to carry the device, but also use it in potentially cramped seating; for example, during a rush-hour public transit commute. However, the requirement to be small impacts the ability to read comfortably. While small enough to fit inside a pocket, most smart phone form factors do not allow a comfortable reading experience due to their limited screen real estate. More screen real estate allows proper use of headlines to direct attention in large format newsprints. Just like paper, it appears, different size formats serve different functions. With its 10” screen, the Kindle DX is unsuitable for most mobile tasks, but allows for superior reading experiences with larger documents.

3.2 Size and Ergonomic Fit

Size is also relevant to the ergonomics of holding an e-book reader, and as such, to the placement of input sensors on the device. Figure 2 shows common hand positions for holding a book and an e-book reader with two hands. Fingers are often placed at the back of the display, with the thumbs at the front locking the device into the space between the thumb and index finger. Frequently used buttons on the front of the display should be placed in close proximity to the thumbs. There has been little experimentation in the commercial e-book market with buttons on the back of the display for paging [3]. It is unclear whether these could allow for more efficient movements than are possible with the thumb [27].

3.3 Energy Expenditure

Size also affects the efficiency of input methods. An input technique is more efficient when movement energy is minimized, given an accuracy criterion [8]. This means that, given an expert user and the task of paging forward one page, a button placed directly underneath the thumb is likely to be more efficient than a swipe gesture that requires movement of wrist and finger across the entire width of the screen. The Apple iPad e-book reader application borrows heavily from the paper book metaphor when it allows users to turn pages using a diagonal swipe gesture from the top right corner to the center of the display. While such gestures are beautiful, and facilitate learning transfer for users familiar only with paper books, they are also a good example of how too literal a use of analog book metaphors may introduce inefficiencies associated with the physics of paper books [5].

In one-handed scenarios, gestural input may be problematic when screens are larger than the range of an average thumb, approximately larger than an iPhone. If the non-dominant hand holds the device, the dominant hand is required for pointing, which increases the energy expenditure. Generally speaking, placement of frequently used buttons in locations that require a change of the natural grip produces an energy penalty to the user. Energy expenditure is particularly relevant to users reading books in a low-energy posture, such as while lying in bed. In such scenarios, books are often held up with one hand, or simply rested on a surface when they are heavy. The requirements for reading in bed are different from those in productivity-related reading scenarios.

3.4 Weight, Grasp and Force

The size of the device impacts the center of mass relative to the hand holding the device [5]. The torque this produces is particularly relevant in one-handed scenarios: the further the center of mass is from the user's hand, the more force will be required to hold the device. This impacts the availability of the thumb to press a button or execute a touch screen gesture without losing grip. It also impacts the freedom necessary for a finger or thumb to rest on a button: the more normal force the device places on the thumb or fingers, the easier it is to press a button. This may produce spurious or unwanted actions, and may require users to rest their thumbs and fingers away from buttons.

Weight impacts fatigue: the heavier the device, the shorter the maximum time of unsupported use. This again is particularly relevant in more casual mobile and bed reading scenarios. With paper documents, we often observed users supporting the entire weight in a single skin crease, where the palm meets the fingers. Decreasing the weight of the device appears to benefit most usage scenarios.

3.5 Grasp and Spurious Interactions

Spurious interactions can be a problem with capacitive buttons and touch screens that overlap with locations where the device is often held. While we, in general, advocate the minimization of screen bezel size, one side effect of the presence of a significant bezel is that it allows the hand to hold the device without spurious interactions with a touch screen. The Kindle, Nook, iPhone, and iPad incorporate lock buttons that disable interactions during transport.

3.6 Thickness and Rigidity

Weight is dependent not only on the size, but also the thickness of the display. With the advent of thin-film flexible e-ink displays (Figure 1) it has become possible to consider usage scenarios in which an e-book reader is almost as thin and flexible as paper. Thickness of display also determines, along with the electronics and casing, the rigidity of the device, and as such is an important factor in the ergonomic fit of the display to the hand. While the Kindle features rectangular shapes that are somewhat uncomfortable to users, the Nook designers curved the back to allow users a more comfortable grip. By contrast, flexible displays such as the one in Figure 1 can automatically follow the shape of the hand. When holding a sheet of paper, we observed users bending the page slightly on a diagonal, starting from the lower left or right of the side of the page, so as to fit to their palm.

When screens are flexible, thinness does not only interact with weight, but also with display size and thus portability. Rollable e-ink technology introduced by Polymer Vision [18] allows screens to be rolled up when not in use, and extended when more screen real estate is required. Thinness also opens up different interaction spaces currently restricted to paper: the kind of interleaved multitasking scenarios described by Sellen and Harper [21]. These are possible only when displays are sufficiently thin and lightweight to be tossed around on a desk, or stacked in piles, without breaking the display [7].

Thinness and flexibility, however, may negatively impact the ability to interact with surface touch screens. Touch displays larger than a smart phone cannot be operated easily with one hand. Most require users to hold the display with the non-dominant hand, typically on the left, freeing up the dominant hand for interactions with the surface of the display. However, a lack of support on one side of the display can cause it to bend when touched. This effect increases with screen width. One solution is to allow the dominant hand to continue to hold the display. However, this limits the extent of touch interactions, requiring users to rely largely on buttons placed along the bezel of the display for most frequent interactions, thus negating the need for a touch screen. An elegant solution often seen when operating a pen on paper is to seek the support of a rigid surface when needed.

One benefit of flexibility is that it opens up use of an alternate input channel that may be more robust to spurious interaction when handling the device: bend gestures. Paper document and book navigation often feature bending of a page as means of interaction. For example, when turning or bookmarking a page. Folding is a common means of scrolling content on paper maps. One potential benefit of bend sensors is that they allow interactions to remain the same when the display is folded. This is because bends can be detected relative to the current shape of the display, and can be detected relative to part of the surface using an array of bend sensors. By contrast, when a display is folded, physical buttons may be out of reach. We believe that, provided suitable solutions to some of the above problems are found, both thinness and the resulting flexibility of displays are desirable qualities.

3.7 Readability and Responsiveness

In the design of e-book readers, there are two further limiting factors that are unrelated to size, weight or thinness of the screen: contrast ratio/emissivity and refresh rate. For e-ink displays, these two relate. The use of polarized ink capsules in the display plane allows the simulation of a reflective paper display experience. The high contrast ratio caused by the lack of emissivity reduces fatigue. Reflectivity also improves battery energy consumption, an important factor in mobile technologies. However, at 150 ms, the e-ink and flexible e-ink display refresh rate is slow.

4 Flexible e-Reader Input Techniques

We summarize our design space analysis as follows:

Size Matters. While a one-size-fits-all approach is suboptimal, the sizes of the Kindle and Nook e-book readers (about 8”x5”), are a reasonable reference for evaluating the design of a flexible e-book reader.

Optimize Movement and Ergonomic Fit. Frequently used functions on the front of the display should be placed in close proximity to common thumb positions when the device is held. Experimentation with back of the display interactions might be promising.

Button Forces Matter. Most buttons on rigid devices act as force sensors. On a flexible device, the force required to activate a button may deform the device. One means of measuring this force is to measure the extent of that deformation. Capacitive touch and buttons that require minimal force to trigger may increase error, while buttons that require a large force may never be actuated due to a lack of a sufficient normal force on flexible displays.

Flexible Displays Bend on Force. Experimentation with bending the display to perform frequently used functions appears promising. In one-handed scenarios, pointing with the dominant hand may be ineffective due to a lack of a normal force.
Spurious Interactions Are a Source of Error. Capacitive, and other minimal force interactions, can increase the number of errors in actuation.

Display Weight Matters. Even a light device will cause fatigue during extended use.
Screen Response Time Matters. The refresh rate of e-ink may be a confounding variable, particularly for existing page animation techniques.

4.1 Focusing Design on Input for Most Frequent Task

To focus our design and experimentation, we restrict our evaluation to input techniques that might be most suited for paging forward and backward through a document on a flexible e-reader form factor. Of the discussed options, we chose five representative input techniques for evaluation: 1) capacitive touch sensors; 2) pressure sensors (buttons) on the front of the display with varying forces; 3) pressure sensors (buttons) on the back of the display, also with varying forces; 4) pressure sensors with one-handed interactions and 5) bending gestures on the side of the display.

We limited our exploration of intentional bending to page flicking. These gestures require less physical effort than most bending gestures. Due to the limited variance in finger movement, we consider them to be most comparable to in place tapping. We further limited our focus to the efficiency and accuracy of interactions with the device. We note that this did not allow us to study calm or mobile aspects of the presented scenarios, and consider these beyond the scope of this particular study.

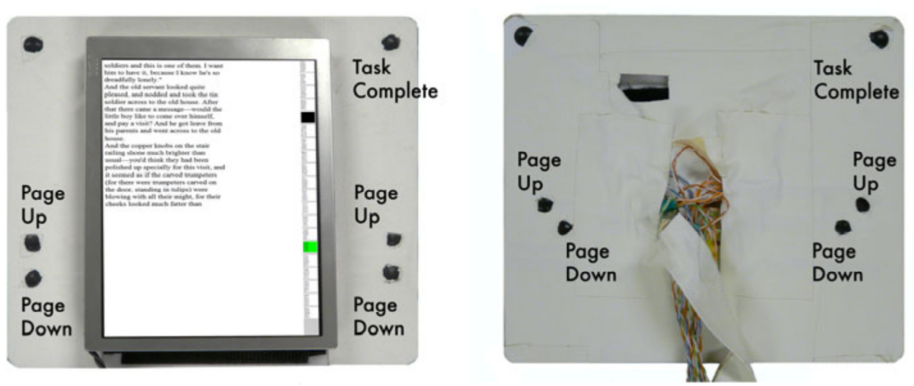


Fig. 3. Front and back of the BendFlip Experimental Apparatus

5 BendFlip: A Flexible e-Reading Apparatus

Figure 3 shows BendFlip, an e-reader that merges properties of future flexible devices with the display refresh rate of a lightweight Kindle-size LCD screen. It served as our experimental apparatus. BendFlip consists of an 8"x7"x1/32" thin, flexible surface outfitted with sensor inputs and a 200-gram rigid LCD display. We used an LCD display because the refresh rates on currently available flexible displays are too slow for our experiment (>150 ms). A 1/8" removable rigid backing, weighing 90 grams, was added to allow comparison of our set of inputs with a rigid e-book reader design. Figure 3 shows how capacitive touch, pressure, and bi-directional bend sensors were mounted on the front and back of the device. Sensor locations were selected based upon the ISO/TR 7250 standard finger measurements (70mm for index fingers). The pressure sensors created 2mm bumps on the surface, which was otherwise taped in bookbinding tape to give it a paper feel. This bump allowed participants to locate the pressure sensors on the back of the device without looking. While pressure sensors do not provide a button click, tap performance is governed by trigger force, and feedback of page turns is displayed. A conductive cloth connected to a capacitive touch sensor covered each of the front pressure sensors. This allowed for capacitive touches to be detected on each of the black pressure sensors shown in Figure 3. Figure 4 shows a user holding the device with fingers on the sensors.

5.1 Detecting Bend Gestures

We used two bi-directional bend sensors to detect bends: one in the center of both the left and right sides of the device. On the right of Figure 4 we see a user bending a side of the device downwards. After several iterations of our initial bend detection algorithm, even small bends can be detected.

Our first algorithm required a 30-degree bend. The size of the bend was selected to differentiate bend from force input. We quickly realized that sensing bend over a smaller area, specifically the inch along the edge of the device, creates a different signal pattern than force input, which distributes the bend over a larger area. This design revision still detected bend using an absolute sensor value threshold.

After several trials with the revised design, we observed that our algorithm still failed to detect certain subtle flicks. For example, many participants did not allow the

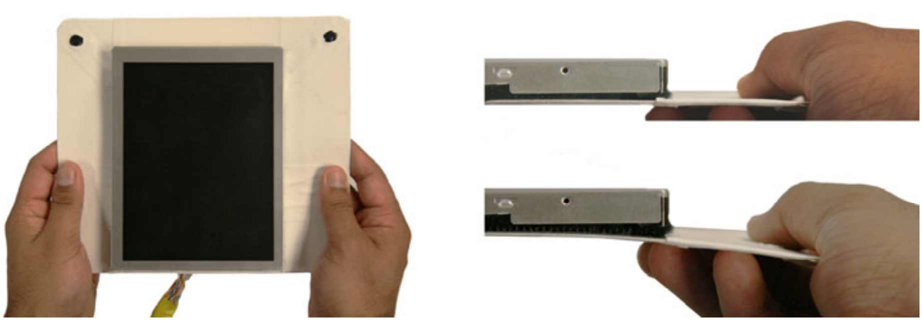


Fig. 4. Pressure and bend view

device to return to its resting state in between flicks. In fact, many pilot participants preferred to hold the device in a bent state while flicking. To address this, we adjusted our algorithm to detect flicks based upon relative changes in the sensor readings. Our final algorithm creates upwards (or downwards) flick events when the edge of the device is bent at least slightly upwards (or downwards) after a peak deformation occurring within a window of one second.

6 Evaluation

We evaluated the following set of input techniques for single page forward and backward navigation with BendFlip. These techniques are intended to represent a range of alternatives found in the current generation of e-book readers, extrapolated to future devices:

1) Front Thumb Pressure ("front press"). The user's left and right thumbs are used to press the pressure sensors on the front of the device. The use of a pressure sensor rather than a button allowed for us to set the force required for actuation, at three levels: minimum, medium, and maximum.

2) Front Finger Pressure ("point"). The user holds the device with one hand, typically their non-dominant hand. The index finger on the user's dominant hand is used to activate the same sensors deployed in the Front Thumb Pressure condition, again with three different forces.

3) Back Finger Pressure ("back press"). The user's left and right hand index and middle fingers are used to actuate the pressure sensors on the back of the device.

4) Front Capacitive ("front capacitive"). Same as Front Thumb Pressure, except the pressure sensors are replaced with capacitive touch sensors that require minimal force for actuation.

5) Bend ("bend"). The user slightly bends the left or right side of the device down with the thumb or up with the back-of-the-device fingers.

23 different variations on each of these input techniques were evaluated experimentally. 6 variations of *Front Thumb Pressure* were evaluated: three different force settings, and with or without the rigid surface mounted behind the device. The three force settings were: 0.2 N, the minimum amount that reliably activates the sensor ("min force"), 1.5 N, approximately the force required to press the Next Page button on the Kindle 2 ("med force"), and 5.9 N, the force required to bend the flexible form factor 40 degrees ("max force"). *Front Finger Pressure* and *Back Finger Pressure* were evaluated with these same 6 variations. *Front Capacitive* was evaluated with and without the rigid surface.

We also evaluated 3 different variations of *bend* gestures. In the first two, only the right side of the device is actuated:

1) Bend Right Side Only. Page up (navigating back) is triggered by a flick up, toward the user, and page down (navigating forward) by a flick down, away from the user.

2) Bend Right Side Only Inverted. The page up and page down actions are reversed from *Bend Right Side Only*.

3) Bend Split Both Sides. In the third variation, both sides of the device are used. A flick down, away from the user on the left side of the device triggers a page up (navigating back), while a flick down, away from the user on the right side of the device triggers a page down (navigating forward). A second factor in our experiment was the rigidity of the device. In these conditions, a thin but rigid surface was mounted on the back of the device.

6.1 Task, Design and Procedure

Participants were asked to navigate between the current page and a target page in the same document. The current page appeared on the left side of the display, with a row of thumbnails on the right side of the display. The thumbnail for the current page was always colored in black. The document was twenty pages in length. Participants navigated back and forth between page 5 and page 15 in the document, 8 times for each of the 23 input techniques. This allowed participants to overshoot their targets within the document. At the beginning of each trial, which constituted navigating to one target, the target page thumbnail was highlighted in green (see Figure 3). A trial was completed once the participant had navigated to the target and pressed a pressure sensor on the top right of the device (marked “Task Complete” in Figure 3), indicating the end of that trial. The time to reach the target (the Seek Time), not the time to press the “Task Complete” pressure sensor, was used to calculate our results. Participants were trained to use each of the input techniques. Training consisted of a demonstration followed by practice. Participants completed training trials until their seek time improved by less than 10%. There were 12 participants (7 male, 5 female), aged 20 to 26. 11 of the 12 participants were right-handed.

The experiment design was a 7x2x2x3 repeated measures with four factors: input technique (front capacitive, front press, back press, bend right side only, bend right side inverted, bend split both sides, and point), form factor (flexible vs. rigid), input side (front vs. back) and pressure setting (min force, med force, max force). These factors are derived from the design rationale, specifically: ergonomic fit, rigidity, grasp, and spurious interactions.

All conditions were counterbalanced using a random order of presentation. The design was not fully factorial, in that bend was only implemented in the flexible form factor, and pressure settings only applied to pressure sensor input techniques. We performed two separate repeated measures analyses of variance (ANOVA). The first compared all input techniques exhaustively using Bonferroni corrected multiple comparisons. The second examined only effects between the fastest pressure input techniques, comparing front vs. back and rigid vs. flexible via a fully factorial design. The dependent measure was seek time in all cases.

We expected rigid and capacitive touch input techniques to be faster than the flexible input techniques. We also expected the bend input techniques would be slower than the fastest flexible input techniques.

To examine error, we also measured the number of page navigations required to complete each trial. If the number of page navigations was more than 10, this

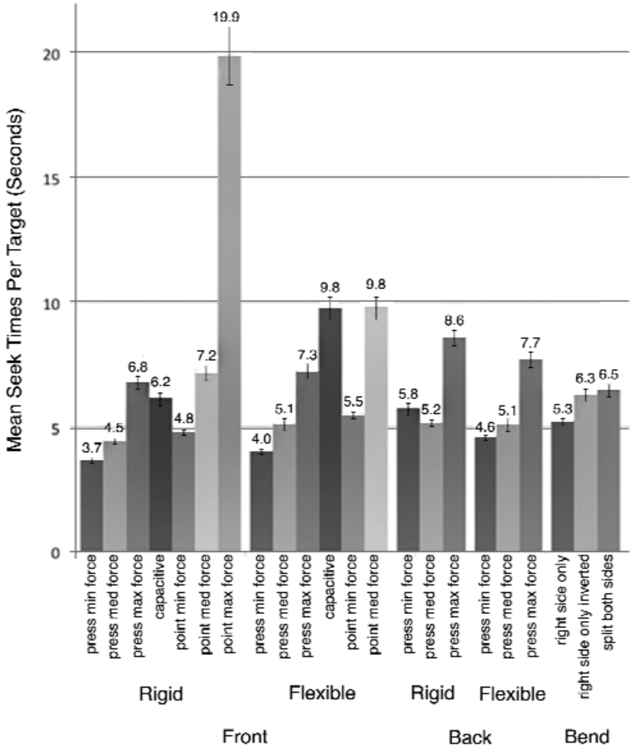


Fig. 5. Mean seek times for input techniques

constituted an error. Overshooting the target or accidentally navigating in the wrong direction would cause such errors. To evaluate performance on errors, we used an ANOVA with one factor: input technique. In this case, the dependent measure was the number of page navigations used to reach the target. To gauge their subjective opinions, participants completed a Likert-style questionnaire after each set of 8 trials.

7 Results

Two of the twelve participants were unable to complete the *point flexible med force* condition. These participants indicated that it was too difficult to activate the pressure sensors. None of the participants were able to complete the *point flexible max force* condition. We also removed the three slowest input techniques from analysis due to their unequal variances: *front capacitive flexible* (9.8 seconds, standard error 0.5), *point flexible med force* (9.8 s, s.e. 0.6), and *point rigid max force* (20.0 s, s.e. 1.3).

Figure 5 shows the results for mean seek time in seconds for each input technique. Analysis of variance showed a significant main effect for input technique: $F_{18,247} = 5.7$, $p < 0.01$. The mean seek time for *front press rigid min force* was the lowest at 3.7 seconds (s.e. 0.1), followed by *front press flexible min force* (4.0 s, s.e. 0.2). A post-hoc Bonferroni found significance between the fastest two input techniques (*front* *press rigid min force* and *front press flexible min force*) and each of *front press rigid max force*, *front press flexible max force*, *back press flexible max force*, and *point rigid med force* ($p < 0.05$).

We also performed a fully factorial ANOVA to compare differences in performance between flexible and rigid form factors, as well as between input sides: front vs. back. We chose the fastest technique that minimized error for this comparison: min press. The main effect for input side was significant, with input on the front significantly faster than input on the back ($F_{1,48} = 9.9$, $p < 0.01$).

7.1 Errors

On average, there were 11.0 page navigations for each trial (s.e. 0.25), while each trial required only 10 page navigations. *Back press rigid min force* had the largest number of page navigation errors. *Back press flexible min force* reported the second largest error (10.9, s.e. 0.28), followed by *bend right only inverted* (10.6, s.e. 0.14), *front press flexible min force* (10.5, s.e. 0.23), and *front capacitive flexible* (10.5, s.e. 0.13). All other input techniques had error rates below 3.5%. There was a significant main effect for input technique on error ($F_{21,262} = 3.7$, $p < 0.05$). Multiple comparisons showed significant differences existed between input techniques with an error rate below 5% and the two poorest techniques: *back press rigid min force* and *back press flexible min force* ($p < 0.05$ in all cases, Bonferroni corrected).

7.2 Qualitative Responses

Mean responses to the 5-point Likert scale question “I enjoyed using this device” (1-Strongly Disagree, 5-Strongly Agree) ranged from 1 (s.e. 0.0) for *front press flexible max force* to 4.75 (s.e. 0.13) for *front press rigid min force*. *Bend right side only* had a mean response of 4.08 (s.e. 0.26), comparable to *back press flexible min force* 4.08 (s.e. 0.29). We tested significance using a Friedman's Two-Way Analysis of Variance by Ranks on the Likert-scale enjoyment scores. The main effect was significant between input conditions (Friedman’s $\chi 2=157.1$, $p < 0.01$).

8 Discussion

We believe that the unexpected poor performance of capacitive input on the flexible form factor was partially due to participants exercising caution to prevent spurious interactions. Possibly for the same reason, pressure sensors on the front of the display significantly outperformed those on the back of the display. As expected, holding the device with one hand and actuating a pressure sensor by pointing with the other hand was much slower and more error prone than holding the device with two hands.

The fastest input techniques for single page navigation have comparable seek times on flexible and rigid form factors. Flexible devices also have the added advantages of being lighter, more portable, and supporting bend input.

Bend right side only performed better than expected; with seek time and Likert enjoyment ratings comparable to the other flexible input techniques. Unlike *bend right side only*, the error rate for the fastest input technique on the flexible form factor, *press med force*, was above 5%. Bend input can also easily be performed anywhere on the device, almost regardless of hand position, and without looking.

Questionnaire comments indicated that participants found *bend right side only* to be the most intuitive of the bend alternatives, and most similar to reading a paper book. Responses to the Likert scale enjoyment question for *bend right side only* were also favorable.

8.1 Design Recommendations

Our results are summarized in the following design recommendations for input on flexible e-book readers:

1. *Single page navigation performance should not be the deciding factor when choosing between flexible and rigid e-book reader designs.* Seek times for the fastest input techniques on the flexible and rigid form factors are comparable.
2. *For flexible designs, bend is an excellent alternative to buttons.* Bend input can have comparable performance and fewer errors.
3. *When using bend to navigate pages, a single bend sensor on the dominant side of the device can be effective.* Page forward should be triggered by bending the edge down, and page back by bending the edge up. Bending up to page forward and down to page up is not recommended. While one might consider placing a bend sensor on both sides of the flexible display to account for left-handed users, using one side to page back and the other to page forward is not recommended.
4. *When using bend, ensure that bends are measured relative to the current shape of the display surface so as to avoid spurious triggering.* For close to button performance, also minimize the extent of the bend.
5. *Capacitive touch and minimal force buttons may increase spurious interactions on flexible designs when placed on the back of the device.* This may be partially due to user reliance on haptic feedback. Capacitive touch is not recommended for flexible designs.
6. *Buttons on the back of the device increase error.* We believe this may be due to users not having visual feedback on their placement.
7. *When using buttons, force settings close to 1.5N appear optimal to avoid spurious interactions.* Our medium force setting allowed participants to support the device and tap in place at the same time. This is currently the case with rigid e-book readers, and transfers to flexible designs.
8. *When using buttons, they should be placed on the front edge of the device* and underneath the thumbs to support tap in place operations.

Note that the use of pressure sensors does not preclude the use of bend sensors. The two can be combined in order to increase the number of available functions.

9 Conclusions

In this paper, we examined input techniques for future handheld flexible display devices. We presented an empirical evaluation of input techniques for the most frequently performed e-book reader task: single page navigation. In our experiment, we compared the use of capacitive touch, pressure sensors (buttons with varying force), and bend sensors for navigating single pages on an e-book reader with both rigid and flexible form factors.

We found that our most efficient bend technique had comparable performance to the button input techniques for the flexible form factor, and fewer errors. For this reason and others, bend may be a compelling alternative to capacitive touch and pressure sensor input. Results also suggest that flexible devices can have comparable performance to rigid devices for e-book reader tasks.

References

1. Amazon, Inc. Kindle Wireless Reading Device, `http://www.kindle.com/`
2. Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., Singh, K.: Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. In: Proc. I3D 1999, pp. 111–118. ACM Press, New York (1999)
3. Baudisch, P., Chu, G.: Back-of-device interaction allows creating very small touch devices. In: Proc. CHI 2009, pp. 1923–1932. ACM Press, New York (2009)
4. Brewster, S., Chohan, F., Brown, L.: Tactile feed-back for mobile interactions. In: Proc. CHI 2007, pp. 159–162. ACM Press, New York (2007)
5. Chen, N., Guimbretiere, F., Dixon, M., Lewis, C., Agrawala, M.: Navigation techniques for dual-display e-book readers. In: Proc. CHI 2008, pp. 1779–1788. ACM Press, New York (2008)
6. E-Ink, `http://www.eink.com/`
7. Engadget, `http://www.engadget.com/2009/10/16/please-hammer-dont-hurt-samsungs-flexible-oled-prototype-vid/`
8. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47, 381–391 (1954)
9. Gallant, D.T., Seniuk, A.G., Vertegaal, R.: Towards more paper-like input: flexible input devices for foldable interaction styles. In: Proc. UIST 2008, pp. 283–286. ACM Press, New York (2008)
10. Harrison, B.L., Fishkin, K.P., Gujar, A., Mochon, C., Want, R.: Squeeze me, hold me, tilt me! An explora-tion of manipulative user interfaces. In: Proc. CHI 1998, pp. 17–24. ACM Press, New York (1998)
11. Herkenrath, G., Karrer, T., Borchers, J.: Twend: twisting and bending as new interaction gesture in mobile devices. In: Ext. Abstracts CHI 2008, pp. 3819–3824. ACM Press, New York (2008)
12. Holman, D., Vertegaal, R., Altosaar, M.: PaperWindows: Interaction Techniques for Digital Paper. In: Proc. CHI 2005, pp. 591–599. ACM Press, New York (2005)
13. Lee, S., Kim, S., Jin, B., Choi, E., Kim, B., Jia, X., Kim, D., Lee, K.: How Users Manipulate Deformable Dis-plays as Input Devices. In: Proc. CHI 2010, pp. 1647–1656. ACM Press, New York (2010)
14. Luk, J., Pasquero, J., Little, S., MacLean, K., Levesque, V., Hayward, V.: A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In: Proc. CHI 2006, pp. 171–180. ACM Press, New York (2006)
15. Marshall, C.C., Bly, S.: Turning the page on navigation. In: Proc. JCDL 2005, pp. 225–234. ACM Press, New York (2005)
16. Nook, `http://www.barnesandnoble.com/nook/`
17. O'Hara, K., Sellen, A.: A comparison of reading pa-per and on-line documents. In: Proc. CHI 1997, pp. 335–342. ACM Press, New York (1997)

18. Polymer Vision, http://www.polymervision.com/
19. Schwesig, C., Poupyrev, I., Mori, E.: Gummi: a bendable computer. In: Proc. CHI 2004, pp. 263–270. ACM Press, New York (2004)
20. Schwesig, C., Poupyrev, I., Mori, E.: Gummi: user interface for deformable computers. In: Ext. Abstracts CHI 2003, pp. 954–955. ACM Press, New York (2003)
21. Sellen, A.J., Harper, R.H.R.: The Myth of the Paperless Office. MIT Press, Cambridge (2003)
22. Tajika, T., Yonezawa, T., Mitsunaga, N.: Intuitive page-turning interface of e-books on flexible e-paper based on user studies. In: Proc. MM 2008, pp. 793–796. ACM Press, New York (2008)
23. Universal Display, http://www.universaldisplay.com/
24. Watanabe, J., Mochizuki, A., Horry, Y.: Bookisheet: bendable device for browsing content using the metaphor of leafing through the pages. In: Proc. UbiComp 2008, pp. 360–369. ACM Press, New York (2008)
25. Wellner, P.: Interacting with paper on the DigitalDesk. Commun. ACM 36(7), 87–96 (1993)
26. Wilson, R., Landoni, M.: Evaluating the usability of portable electronic books. In: SAC 2003, pp. 564–568. ACM Press, New York (2003)
27. Wobbrock, J.O., Myers, B.A., Aung, H.H.: The performance of hand postures in front-and back-of-device interaction for mobile computing. Int. J. Hum. Comput. Stud. 66(12), 857–875 (2008)
28. Zhai, S., Kristensson, P.: Shorthand writing on stylus keyboard. In: Proc. CHI 2003, pp. 97–104. ACM Press, New York (2003)

Who's That Girl? Handheld Augmented Reality for Printed Photo Books

Niels Henze and Susanne Boll

University of Oldenburg
Escherweg 2, 26121 Oldenburg, Germany
{niels.henze,susanne.boll}@uni-oldenburg.de

Abstract. Augmented reality on mobile phones has recently made major progress. Lightweight, markerless object recognition and tracking makes handheld Augmented Reality feasible for new application domains. As this field is technology driven the interface design has mostly been neglected. In this paper we investigate visualization techniques for augmenting printed documents using handheld Augmented Reality. We selected the augmentation of printed photo books as our application domain because photo books are enduring artefacts that often have online galleries containing further information as digital counterpart. Based on an initial study, we designed two augmentations and three techniques to select regions in photos. In an experiment, we compare an augmentation that is aligned to the phone's display with an augmentation aligned to the physical object. We conclude that an object aligned presentation is more usable. For selecting regions we show that participants are more satisfied using simple touch input compared to Augmented Reality based input techniques.

Keywords: augmented reality, mobile phone, photo sharing, mobile interaction, image analysis, photo book.

1 Introduction

The paperless office was predicted more than a quarter century ago. Despite the availability of desktop computers, notebooks and smartphones everywhere at any time this revolution has not happened yet [25]. Paper documents have qualities that current digital device fail to offer: paper has an amazing resolution, does not need a power supply, and remains accessible for centuries; furthermore, paper is cheap and one can leave it lying around without much concern about thievery or environmental influences. The digital revolution, however, happened in the last decades. While the digital world is increasingly getting dynamic and interactive - paper remains static. Once printed or written the content remains. Today, paper documents are usually produced from digital content and printing is the end of the production chain. Related digital media is often available at production time but is not used for particular prints. Digital content might also evolve, get annotated, iterated and improved but the once printed document provides no access to these changes. What is missing is a way to unify the advantages of printed documents with the flexibility and richness of online content.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 134–151, 2011.

Using mobile devices in conjunction with physical objects has been proposed to unify the advantages of paper documents with the interactivity, personalization, and real-time features provided by digital media [7]. Augmented Reality (AR) looks as a natural option to fuse digital media with paper documents [14]. As traditional AR has severe technical limitation, such as the need for AR goggles, handheld AR recently received a great share of attention in the AR domain [26]. Smartphones are used to present the phone's camera image on the screen augmented with additional content that is aligned to the recorded physical world. Recent technical development has made major progress beyond the use of artificial markers towards augmenting diverse types and objects. Handheld AR research is currently dominated by technical development. New algorithms make handheld AR feasible for more and more different types of object, incremental improvement increases the accuracy, and combining different techniques greatly improved the performance [26]. The interface design, however, has been mostly neglected so far. The HCI community only scratched the surface of designing handheld AR interfaces.

The application domain we selected to investigate the interface design of handheld AR systems is the interaction with printed photo books. Printed photo books are highly emotional enduring artefacts, handcrafted by users to preserve memories and often have interactive online galleries as natural counterparts. To create a connection between the physical and the digital, we try to enable users accessing the same information available in photo communities' galleries using physical photo books.

In this paper, we use photo books to investigate two fundamental aspects of handheld AR for printed documents: How to visualize annotations and how to select regions to create annotations. In the following Section we discuss related research. Afterwards, we motivate the augmentation of printed photo books in Section 3 and we present an initial study to collect design approaches in Section 4. The visualization techniques and selection techniques are designed in Section 5. Section 6 describes the design and the results of the experiment that compares the designed approaches. We close the paper with conclusions and an outlook to future work.

2 Related Work

Several approaches have been developed to create the link between physical objects and digital content. Fitzmaurice was one of the first who predicted the use of mobile devices for interaction with physical objects by simply pointing at them. He described for instance an application with which the user can point at locations on a physical map with a mobile device to get additional information [7]. In recent years more systems emerged that provide information related to physical objects using mobile phones. Common implementation mark objects either with visual markers [21] or with electronic tags [28]. Barcodes can be seen as the first implementation of visual markers back in 1948. Since special barcode readers are needed to read 1D barcodes 2D barcodes such as QR-codes [15] have been developed. QR-codes can be read using recent mobile phones and a few mobile phones also have an integrated RFID reader to read electronic tags. However, not all physical objects are suitable for markers. Sights and buildings are simply too large or out of range to be reasonably equipped with either type of markers. It is also questionable if objects, whose visual appeal is important, such as printed photo books, can in general be equipped with visual markers. Not only because the markers require visual space but also because visual markers affect the design of the object. Electronic marker, however, lead to ambiguity for applications with a high object density. Furthermore all markers restrict the interaction radius in a specific way. In particular, electronic markers do not to provide the relative location of the object.

Another approach to create the link between physical objects and digital content is using content-based object recognition. Davies et al. investigated user reaction to the use of a digital image capture and recognition system for mobile phone to access information about sights [5]. With their system users can take a photo of a sight with a mobile phone and receives a related description. Davies et al. found that half of the users want to use such a system even when "this is a more complex, lengthy and error-prone process than traditional solutions". Pielot et al. compared the use of object recognition and typing an URL to access information about posters [20] with a mobile phone. They showed that users are faster using object recognition and also prefer this interaction technique. In our own work, we showed that even novice users can use this interaction to get information about printed photos [10].

Handheld AR (also called Magic Lens) is the adaptation of AR for mobile phones. It extends the concept of using discrete photos by providing immediate feedback using an augmentation. Prototypes that augment different types of physical objects, such as text-heavy conference proceedings [14], real estate guides [6], and sights [1] have been developed. Different groups (e.g. [23, 18, 12]) developed prototypes to augment paper maps and conducted according user studies. Rohs et al. compared users' performance in a find-and-select task using a joystick controlled "static peephole", a dynamic peephole and a handheld AR interface (called Magic Lens) [23]. The study showed that the dynamic peephole and Magic Lens clearly outperform joystick navigation. Morrison et al. compared handheld AR for a paper map with a digital map [18] in a real world setting. One of their conclusion is that the "main potential of (handheld) AR maps lies in their use as a collaborative tool".

A number of studies showed encouraging results for handheld AR systems compared to traditional approaches and researchers investigated fundamental aspects, such as the adaptation of Fitts' law. Much less work addresses the interface design of those systems. Henze et al. proposed different interface designs for a system that augments music records based on a user study [11]. Liao et al. implemented selection and interaction techniques for printed conference proceedings [16]. Based on an initial study they identified challenges for future work (e.g. slow image recognition and inaccurate document registration). What is missing is an understanding of the alternatives of the interface design beyond qualitative results.

3 Interaction with Printed Photo Books

In our work, we investigate solutions that close the gap between printed documents and the digital realms. A particular application area that we currently focus on is the interaction with printed photo books [10]. Photo books are enduring artefacts handcrafted by users to preserve memories and often have online galleries as natural counterparts. To create a connection between the physical and the digital we try to enable users to access the information available in photo communities' galleries using their physical photo books. By combining digital content, such as comments, music, or videos, with the printed photo book, the photo book could be brought to life.

Photo books are often used by individuals or small groups possibly to tell about the last vacation, the good old times or more generally to share memories. Frohlich et al., for example, found that photos are mainly used to share memories and that "sharing photos in person was described as most common and enjoyable" [8]. People use the photos to "share the memory", and Frohlich also reports that people use printed photos for "jointly 'finding' the memory". Crabtree et al. emphasize that the sharing of printed photos relies upon the distribution of photos across group views and personal views [4]. The following scenario sketches a typical situation that could emerge around photo books:

Mary has invited her family to celebrate her birthday. After welcoming all guests and receiving birthday gifts Mary digs out a couple of photo books she produced for recent events such as her trip to Mexico, her son's baptism, and her daughter's school enrolment. A lively discussion about the events emerges between the guests and the photo books are passed from one to the other. While Mary is chatting about Mexico with her brother, her grandparents discuss the baptism in detail and her nephews and nieces debate about the school enrolment.

While the photo books are a great means to share memories as described above, they come with a few limitations. Content has been selected to create a visual appealing and affordable photo book. More information, e.g. the names of the persons on the photos or comments, as well as additional photos and videos are often available at production time. This background information remains inaccessible using the plain photo book. By augmenting the photos each user could access additional information that are of particular interest to him or her in the specific situation. Mary's brother might want to learn where the photos have been taken exactly and get more information about the photographed sights. He might also be interested in the additional video snippets Mary took during her vacation. Digital content evolves and different users can annotate the content with additional information. Mary's grandparents are, for example, interested in who attended the baptism. Mary would be happy if her parents could provide more information about distant relatives on the photos and if they would add names to the photos printed in the book. Mary's nephews and nieces are probably less interested in the family events but want to get entertained. An audio drama prepared by Mary could connect the photos to an age-specific funny storytelling of the event turning the photo book into an audible picture book for kids.

4 Requirements and Participatory Design

In order to design the handheld AR interface for printed photo books we started with an initial study to collect features and proposals for the interface design. We collected information and features participants want to access using their printed photo book. Furthermore, we asked participants to propose designs to visualize information with handheld AR using pen and paper.

4.1 Methodology and Participants

In the beginning of the study we introduced the participants to the studies purpose and collected demographic data. The remainder of the study was split into two halves. In the first part we asked participants to write down a list of information and features they consider relevant. Participants were also asked to rate each of the named features on a five point Likert scale (from not important to very important). Afterwards, we asked the participants to draw an augmentation on printed images containing a mobile phone that shows an unaugmented image of a photo book on its screen (see Figure 1). The layout of the used photo books is consistent with the sparse knowledge gained from analysing photo books [24] and is also consistent with image composition algorithms for photo books [2].

12 persons (8 male) participated in the study. The participants' age was between 8 and 54 years and the average age was 32.6 years. 4 participants had a technical background (undergraduate and graduate students) and 8 had no technical background.

4.2 Results

In the following we report the results from the user study. First the results for the desired features and information are described, followed by an outline of the sketched interfaces.

Information and Features. The participants named 6.83 (std=3.43) different information/features on average. We normalized the results by merging synonyms and very similar answers. Table 1 shows the average rating for aspects that have been mentioned at least three times.

Table 1. Frequency and rating of named information

Information	*Mentioned*	*Rating*	*Information*	*Mentioned*	*Rating*
persons' names	9	4.2	object description	5	3.4
recording time	8	4.3	comments	4	3.5
recording date	7	4.7	tags/categories	3	4.0
recording place	6	3.8	related images	3	3.5
title/description	6	3.3	links/social networks	3	3.0

The results can be further reduced considering that recording time and date is very similar information that is often presented side-by-side. Persons' names and object descriptions are also similar information that describes specific parts of a photo. Most participants proposed to not only display information but requested the possibility to also create or change additional content. In particular, participants wanted to add descriptions of persons, sights and other objects to photos similar to the way photos can be annotated in online galleries on Flickr and Facebook.

Visualization. The visualizations proposed by the participants can be differentiated by the way the overlay is aligned. Six participants aligned the information to the border of the phone. Figure 1 (left) shows an example of a sketch where the information is presented at the phone's border. The information is located at either one or two sides of the display. In contrast, three participants aligned all information to the photo. Figure 1 (right) shows one of these sketches that presents the information at the photo's border. The other participants choose a mixed design where some information is aligned to the phone's border and other information is aligned to the photo itself. Particularly, information that describes only parts of a photo is aligned to this part while general information about a photo, such as its title, is located at the border of the phone's display.

Ten participants explicitly suggested highlighting the recognized photos of the photo book in some way. Seven participants proposed to draw a rectangle around the photos. Other participants suggested to gray out the background or did not specify a particular way. Even though not requested, six participants proposed to have a way to activate additional functionalities on a separate view. Two participants proposed using the phone's menu button and two proposed to use icons (e.g. a video icon) that lead to the separated view. The other two participants did not specify a particular way to activate the additional functionalities.

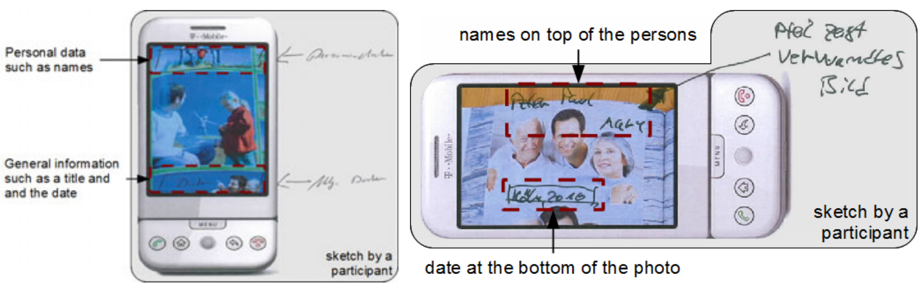


Fig. 1. Sketches of handheld AR interfaces by two participants. Information is either aligned to the display's top and bottom (left) or aligned to the photo (right). For illustrative purpose we selected a very simple sketch.

4.3 Discussion

The information that has been requested most often by the participant describes the photos content such as, persons and objects that have been photographed. When and where a photo has been taken is almost equally important. More general information textual description of a photo such as tags and a title ranks third. Participants could also envisage including social features such as comments or a pointer to social networks. In general, it can be differentiated between information that describe particular parts of a photo (e.g. the name of a person), information that describes the whole photo (e.g. recording time or a title), and content provided by other users (e.g. comments). Most participants do not only want to view information but also want to add information. They propose to be able to select regions of a photo to tag persons, sights, and other objects.

The sketches for the visualization of the augmentation produced by the participants revealed three different pattern that are consistent with a the results of a similar study Henze et al. conducted for augmenting physical music records [11]. Participants align the elements either to the augmented object or they align them to the phone's border. Furthermore, some participants propose mixed designs. Most participants propose to highlight the augmented object. While not all participants proposed the highlighting of augmented objects explicitly in both studies we assume that this a general demand.

5 Design and Implementation

Based on the results of the initial study we designed an interface to interact with printed photo book. In addition to a pure visualization of information we decided to also design interactions to select regions of photos, in order to tag persons, sights, and other objects, as this has been requested by the participants. In the following, we describe the design of the augmentations and the designs of the selection techniques. Afterwards, the implementation of the resulting prototype is outlined.

5.1 Augmentation Design

The proposed designs for the augmentation can be divided in those were the placement of information is aligned to the phone and those were the information is aligned to the object. Furthermore, some participants proposed a mixture of both approaches. We decided to not design a mixed augmentation in order to investigate the alignment aspect without the ambiguity of a mixed design. The participants proposed a number of information items they consider important to visualize. The information can be distinguished in information that describes a whole photo and information that describes a certain part of a photo. In order to support both types we decided to support a title for the photos and annotation of persons and objects in a photo. The title is a representative for information that describes the photo in general while the annotation of persons and objects is the representative for information that describes certain parts.

As proposed by the participants photos and annotated regions are highlighted. For all visualization and annotation techniques the pages of the photo book are highlighted by displaying only the page with colours and leaving the surrounding greyed out. Furthermore, individual photos and annotated regions of a photo are highlighted by drawing a rectangle around them. The centre of the display is marked with a crosshair.

Photo-aligned Augmentation. This design, shown in Figure 2 (left), attaches the information to the photos. The title of a photo is aligned to the top of the photo while the descriptions of particular parts are aligned to the top of a rectangle around this part. The augmentation follows the movement of the photo inside the camera's video. If multiple photos are visible all photos are highlighted and each has its own elements visible simultaneously.

Fig. 2. The photo-aligned interface design aligns the annotations to the photos and for all visible photos simultaneously

The advantage of this design is that all information is visible at the same time. As the text is shown as an augmentation the position and the size of the text changes according to the augmented object. Thus, the size of the presented text can naturally be increased and decreased by changing the distance of the phone to the photo book. One disadvantage of this design is that the augmentation becomes small if the phone is far from the photo. Thereby, text might become difficult to read if a user wants to get an overview about the information available for a photo book page. Furthermore, the text permanently moves and wobbles if the user moves the phone. Thereby, readability of text is affected by accidental movement of the phone.

Phone-aligned Augmentation. The second design, shown in Figure 2 (right), aligns the information to the phone. The photos' titles are at the top of the screen while the

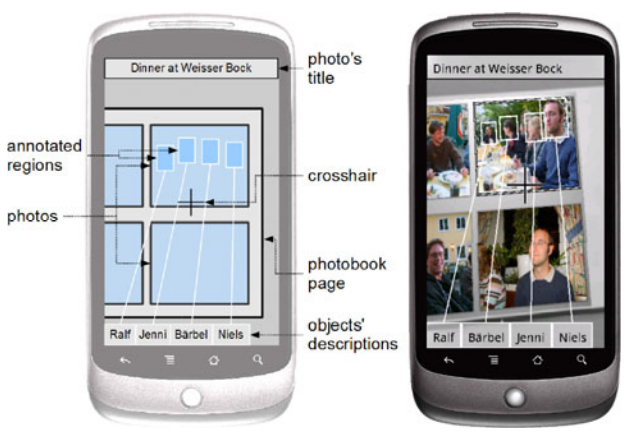


Fig.2. The design shows the annotations aligned to the display's top and bottom. Only the annotations of the photo below the crosshair are visible.

descriptions of particular regions are at the bottom. To connect the textual description of a region with the region in the photo a rectangle is draw around the region and a line connects the description with the region. Thus, the text is always located at the same position. Information is only displayed for the photo that is located below the crosshair that marks the centre of the screen. Thus, even if multiple photos are visible in the camera image only the information for one photo is displayed.

This design has the advantage that the text has always the same readable size. As the text stays at the same position as soon as the crosshair is above a photo readability is not affected by the movement of the phone. The design's disadvantage is that the information for only one photo is visible at a time. Thereby, the user must move the crosshair across all photos on a photo book page to get an overview about the available information. Since the text does not change its position selecting it would be less affected by accidental movement of the phone.

5.2 Interaction Techniques to Select Regions

As annotating regions has been requested in the initial study we designed three selection techniques to mark regions of a photo. As we did not collect recommendation for designing this interaction from participants we designed three fundamentally different approaches. With the first two selection techniques users select regions in the reference system of the augmentation. They either by have to move the phone or touch on the display. We included a third technique where regions are marked by touching a separate static image as a baseline. We did not include the selection techniques Liao et al. proposed with the PACER system [16] because we aim at true handheld AR instead of loose registration and we cannot exploit knowledge about distinct document regions (e.g. words and sentences). The two techniques touch-based techniques can, however, be seen as the basic concepts that are combined in PACER.

Crosshair-Based Region Selection. With the first technique the user aims with the crosshair that is located in the centre of the display at a corner of the region that should be selected. The technique is illustrated in Figure 3. The technique is inspired by handheld AR systems that use a crosshair in the centre of the screen to select predefined objects (e.g. [22]). By touching the display at any position the user defines the first corner (e.g. the top-left corner) of the region. The user than has to move the crosshair to the opposite corner (e.g. the bottom-right corner) by physically moving the phone while touching the display. The region is marked when the user stops touching the screen. As the region is created in the reference system of the augmentation the created rectangle is aligned to the photo.

The advantage of this technique is that the "fat-finger problem" (i.e. that users using touchscreens occlude the area they want to touch) is avoided. As the location at which the crosshair aims can be estimated more precisely than the position in which a touch results it might also be more precisely. As the user can zoom (by changing the distance to the photo) while moving the crosshair, the region could be created more precisely. A disadvantage is that the device must be physically moved. This could lead to a higher physical and mental demand and makes the technique prone to accidental movement of the phone.

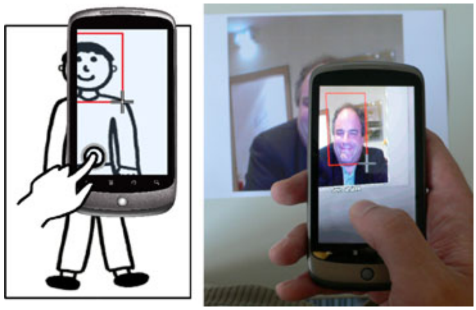


Fig. 3. The three techniques to select regions: crosshair-based on the left, touching in the augmentation in the centre and using touch and a static picture on the right

Augmented Touch-Based Region Selection. With the second technique, shown in Figure 4, the user touches at a corner of the region that should be marked to define the first corner. After moving the finger to the opposite corner the region is marked when the user lifts the finger from the screen. As the region is created in the reference system of the augmentation the created rectangle is aligned to the photo even if the user rotates the phone. This technique is inspired by handheld AR systems where users have to touch the augmentation of predefined objects to select them (e.g. [12]).

Fig. 4. The three techniques to select regions: crosshair-based on the left, touching in the augmentation in the centre and using touch and a static picture on the right

The advantage of this interaction technique is that the user does not have to physically move the phone. Therefore, it might be less physically and mentally demanding. The user can, however, imitate the crosshair-based approach by not moving the finger but only the phone. In this case the techniques have almost the same advantages and disadvantages as the crosshair-based approach. In any case the technique is affected by the "fat-finger problem".

Unaugmented Touch-Based Region Selection. The third technique, shown in Figure 5, serves as a baseline that works on a static image of the photo that should be annotated. As with the previous technique the user touches at a corner of the region that should be marked to define the first corner. The user than has to move the finger to

Fig. 5. The three techniques to select regions: crosshair-based on the left, touching in the augmentation in the centre and using touch and a static picture on the right

the opposite corner. The region is marked when the finger is lifted from the screen. We did not use more sophisticate selection techniques because the other techniques would benefit similarly from more sophisticated interaction techniques.

The clear advantage of this technique is that the user can freely move the phone as the phone is disconnected from the physical photo. Thus, unintentional jitter is avoided but using a separated view can be an imminent disadvantage for a handheld AR system. Another limitation is that the user occludes the area where she aims at with the finger. For this concrete design (but not generally) a further limitation is the lack of zoom.

5.3 Prototypical Implementation

In order to implement the designed visualization and interaction techniques we needed to implement an Augmented Reality system for mobile phones. AR systems estimate the pose of the augmenting display in relation to the scene or object that should be augmented. Using this pose a system can transform the augmenting overlay into the reference system of the physical scene and render the augmentation.

As we aim to preserve the photo book's visual appeal visual markers (e.g. QR-codes) cannot be used. Text-based document recognition [6, 14] is also not possible because photo books contain mostly images. Another widely used techniques to implemented AR systems are recognition algorithms such as SIFT [17]. SIFT and similar approaches are still too demanding for today's mobile phones. Wagner et al. simplified the SIFT algorithm to make the estimation of an object's pose feasible on mobile phones [27]. Their approach is able to process camera frames with a size of 320x240 pixels at a rate up to 20Hz. They further extended the algorithm by combining it with object tracking [27]. This extension enables to recognize and track up to 6 objects with 30Hz. However, only results from processing 6 images are reported and it was not analyzed how the algorithm performs with an increasing number of objects.

In order make handheld Augmented Reality feasible for printed photo books that can contain more than 50 pages we extended the approach by Wagner et al. Similar to [12] we integrated a Vocabulary Tree [19] in the object recognition pipeline. In the pre-processing phase, photo book pages are analyzed to extract simplified SIFT features [27]. Furthermore, the photo's metadata including titles and annotated regions is converted to a XML format that describes the content of one photo book. Installed on the phone the prototype reads the content description, the according features, and scaled versions of the photos (256x256 pixels). During runtime simplified SIFT features are extracted from the images delivered by the phone's camera and compared to the SIFT features from the photo book pages using the Vocabulary Tree and brute-force matching. If the number of matches is above a certain threshold an according homography is computed. This homography is used to draw an overlay on top of the camera image. To increase the speed, recognized pages are tracked (see [27]) in subsequent camera images. We implemented the algorithm for the Android platform using C (for the performance critical parts) and Java. The prototype recognizes objects in a 320x240 pixel camera frame with 12 FPS and tracks objects in subsequent frames with about 24 FPS on a Google Nexus One. [27] provides an extensive description of their approach and its performance, considering registration errors, and frame rate. We do not use the same implementation but we are certain that the performance is very similar.

Based on the implemented handheld AR algorithm we designed an application that provides the two visualization techniques and the three different ways to select regions. Switching between the visualizations and selection techniques is performed using a menu.

6 User Study

In order to compare the visualization and selection techniques developed in the previous section, we conducted a user study that is described in the following. In the experiment participants performed one task to compare the visualizations and one task to compare the selection techniques. A within-subject design with one independent variable (two conditions in the first task and three conditions in the second task) was used for both tasks.

6.1 Procedure

After welcoming a participant we explained the purpose and the procedure of the study. Furthermore, we asked for their age and noted down the participant's gender. Prior to each task we demonstrate how to use all conditions.

In the first task, participants had to answer five questions related to the photos in a provided photo book. To answer a question they had to read the augmentation shown on a mobile phone. Participants had to combine the information provided by the photos with information provided by the augmentation. E.g. one question was "Who watches soccer?". For this example participants must identify the photo with persons watching soccer and read the annotation that contains the persons' names. After answering a question participants were asked the next question. After completing all questions with one visualization technique they repeated the task with the other visualization and another photo book. We asked participants to answer the questions as fast as possible. The order of the conditions and the order of the used photo book were counterbalanced. We measured the time participants needed to answer the five questions. Furthermore, we asked them to fill the NASA TLX [9] to assess their subjective task load and the "overall reactions to the software" part of the Questionnaire for User Interaction Satisfaction (QUIS) [3] to estimate the perceived satisfaction.

In the second task, we asked the participants to select regions on provided photos. With each of the three selection technique the participants had to mark a region in three photos (e.g. "Mark the person's face."). They could repeat marking a region if they were not satisfied with the result. Participants were asked to mark the region as fast and precisely as possible. After completing the task with one selection technique participants repeated the task with the next technique and a new set of photos. The three conditions were counterbalanced to reduce sequence effects. We measured the time needed to mark each region, the coordinates of the region, and how many attempts participants needed. Furthermore, we asked participants to fill the NASA TLX and the "overall reaction" part of the QUIS.

6.2 Participants and Apparatus

We conducted the user study with 14 participants, 6 female and 8 male, aged 23-55 (M=31.21, SD=8.6). Five subjects had a technical background (mostly undergraduate students) none of them was familiar with handheld AR or the used application.

The prototype described in Section 5.3 running on a Google Nexus One was used for both tasks. The investigator selected the visualization and interaction technique between the tasks. For the first tasks we prepared two photo books printed on A4 and annotated each of the containing photos with a title and/or regions of the photo describing parts of it. The theme of the first photo book was a wedding and the theme of the second photo book was the visit to a fun fair. We prepared an additional photo book for the introduction with photos taken at a scientific conference. For the second task we printed 20 photos on A4.

6.3 Hypothesis

For the first task we predicted that the photo-aligned presentation is more usable than the phone-aligned presentation. With the photo-aligned presentation the user can see all information simultaneously and can quickly focus on different texts by changing the distance of the phone to the photo book. Therefore, we assumed that participants perceive this condition as less demanding and give it a lower NASA TLX score. Due to the same reasons we assumed that participants would give a higher QUIS score to the photo aligned presentation.

For the second task we assumed that the crosshair-based technique would receive a higher QUIS score and that this condition is perceived as less demanding, which would result in a lower NASA TLX score. We assumed that because, compared to the other conditions, the crosshair-based technique can be used with a single hand and the user can zoom and change the selection simultaneously just by moving the phone. For the touch-based techniques we assumed that unaugmented touch would be more usable because the movement of the hand does not move the image that should be selected.

6.4 Results

After conducting the experiment we collected and analyzed the data. We found significant differences between the two visualization techniques as well as between the three selection techniques. We did not find significant effects on the time participants needed to complete the tasks. Participants' qualitative feedback was translated to English.

Augmentation Design. Comparing the two visualization techniques we found that the augmentation design had a significant effect ($p<.05$, $r=0.81$) on the NASA TLX score (see Figure 6). The perceived task load is lower ($M=103.64$) if the augmentation is aligned to the photo compared to the augmentation that is aligned to the phone's border ($M=117.86$). The augmentation design also had a significant effect on the participants average rating of the QUIS's "overall reactions to the software" part ($p<.001$, $r=0.70$). On average the rating is higher if the augmentation is aligned to the object ($M=6.60$) compared to the score for the phone-aligned visualization ($M=5.36$). The individual scores are shown in Figure 6. The visualization technique had a significant effect on the results of all questions ($p<.001$ for the first three questions and $p<.05$ for the others). Task completion time using a photo aligned augmentation is $M=251s$ ($SD=95s$) and $M=270s$ ($SD=127s$) for the phone aligned augmentation but the difference is not significant ($p=0.07$).

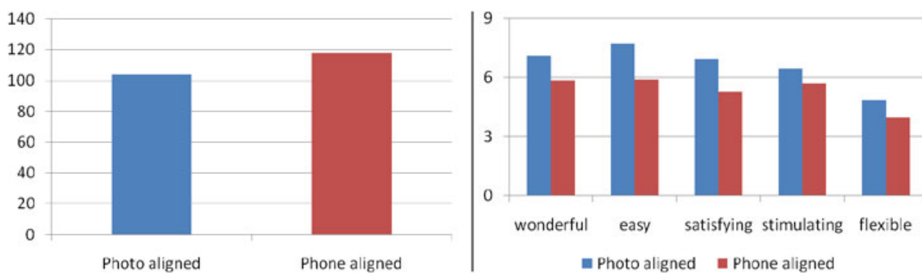


Fig. 6. NASA TLX (left) and QUIS "overall reactions to the software" part (right) for the visualizations

Most of the participants' comments addressed the performance and the accuracy of the object recognition. E.g. one participant mentioned that "it’s shaking - probably I hold the camera wrong" and another participant stated that "the recognition should be faster" and the system "should tolerate bended pages". Participants mentioned for both conditions that the recognition works better than with the other condition.

We observed for both conditions that participants prefer to hold the phone sideways. That led to negative comments about the phone-aligned presentation. E.g. "it's difficult to read because the text is skewed" or "have to turn the phone to read the

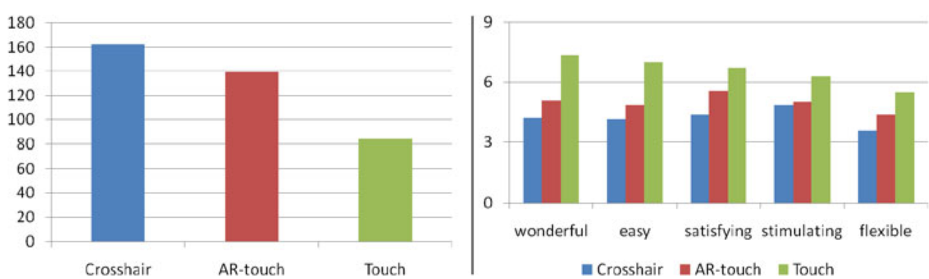


Fig. 7. NASA TLX (left) and QUIS "overall reactions to the software" part (right) for the selection techniques

text". About the photo-aligned condition participants mentioned that "it provides a good overview" and "you can see everything". However, they also mentioned that this presentation is "a bit overloaded" and "you have to go near to use the functionality".

Selection Techniques. In the second task we compared the three selection techniques. We used the Bonferroni correction to correct the significance levels. The analysis of variance (ANOVA) shows that the selection technique had a significant effect on the NASA TLX score ($p<.01$). Comparing the individual condition (see Figure 7) shows that using unaugmented touch ($M=84.07$) results in a lower score than using augmented touch ($M=139.36$, $p<.01$) or the crosshair-based technique ($M=161.79$, $p<.001$). The score for augmented touch is lower than for the crosshair-based technique but, considering the corrected significance level, the effect is not significant ($p=0.025$).

An ANOVA test shows that the selection technique also had a significant effect on the average QUIS's "overall reactions to the software" part ($p<.001$). Using unaugmented touch leads to a higher score ($M=6.57$) compared to augmented touch ($M=4.97$, $p<.01$) or the crosshair-based technique ($M=4.23$, $p<.001$). The score for the augmented touch technique is also higher than the score of the crosshair-based technique (see the individual scores in Figure 7) but without a significant effect ($p=0.027$). Average task completion time for the selection subtasks are crosshair: M=5.4s (SD=4.3), augmented touch: M=6.8s (SD=5.1), and touch M=4.1s (SD=2.2) but the differences are not significant (ANOVA: $p=0.08$).

Even though, we demonstrated the techniques prior the task and ask the participant if he/she understands the technique, some participants did not understand the crosshair-based technique. One participants, for example, noted that "it is difficult to touch the crosshair" although it is not necessary to touch it. Mentioned reasons why this condition performs worse than the others are because it is an "unusual interaction" and that it is "difficult to mark a picture by moving the phone". Another participant noted that "moving the whole body is not comfortable". Further comments are that it is "difficult to catch the crosshair where I want it to be" and the same participants stated that "I always forget paying attention to the crosshair". An advantage participants identified is that "the finger does not occlude the object" and that this technique is "usable with one finger".

For the augmented touch technique four participants appreciated that "it has zoom" (compared to the last condition). Compared to the crosshair-based technique they liked that "one can draw the window with the finger". This condition's most often mentioned limitation is that "the device moves when dragging the box" and that "touching changes the position of the phone" or more generally: "it shakes too much for me".

We got mostly positive comments about the unaugmented touch condition. However, participants identified only one advantage of this technique, even though most participants commented on this advantage. They liked that the "image does not move" and that "the image freezed". They also explicitly stated it is "easy to select because it (the image) does not move". The main limitation the participants identified is that "it has no zoom", that "zooming would be nice" and that it is "less precise than the cursor without zoom". Another problem participants mentioned is that "my finger is to fat" or with other words "there is the fat thumb again".

6.5 Discussion

The results of the first task support our hypothesis that the photo-aligned presentation is more usable than the phone-aligned presentation. Participants perceive the photo-aligned presentation as less demanding and are more satisfied. For the second task the results contradicted our hypothesis. Participants clearly prefer the unaugmented touch technique and the main reason is that image that should be selected does not move.

Based on the sparse comments we assume that the photo-aligned presentation is superior because it provides all information simultaneously and therefore helps to get an overview. The user does not have to select an object to get information about it. This is, however, also the main limitation: The photo-aligned presentation technique does not only allow the user to zoom in and out but it is required to do so. On pages with a high density of annotations the amount of text that is hardly readable can be confusing. We assume the results can be transferred to other tasks with a similar or lower object density. For those tasks the text size could be further increased, which makes it even easier to get an overview. For tasks with a considerably higher object density the text size must be adjusted accordingly to avoid overlapping texts. In this case an object aligned presentation will presumably become less usable because the user has to "zoom" often by moving the phone towards the objects.

The participants clearly preferred to select regions in a static image compared to the two techniques that use AR. This result is surprising because the design of the study favoured the two other conditions. No zoom was available even though a number of well established techniques exist to implement zooming for static images. Furthermore, we did not randomize the order of the tasks and using handheld AR in the first task certainly improved the participants' performance for the two AR based techniques. The qualitative feedback is also quite clear. Participants prefer unaugmented touch because they do not have to deal with the augmentation.

The study has two main limitations. The tasks and the setting are artificial in particular for the first task. For the intended use case it cannot be expected that users will search for particular information. Rather, users usually do not have temporal pressure or want to answer specific questions while browsing through a printed photo book. The second limitation, which applies for both tasks, is the short time participants used the conditions. From this perspective, it is even remarkable that all participants could use the conditions of the first task without any problems. Especially for selecting regions more training would certainly improve the performance with the AR-based techniques. However, it is questionable if training can invert the results. Furthermore, users might not be willing to learn using the crosshair-based technique because it is hard to use at least in the beginning.

7 Conclusions and Future Work

We investigated the design of a handheld AR interface for printed photo books. Based on an initial study we designed two visualizations for augmenting photo books with additional information. These designs and three techniques to select regions of printed photos have been implemented. The subsequent user study shows that aligning annotation to the augmented photo is more usable than aligning annotations to the phone's display. We also showed that users prefer to select regions on static images compared to selection using AR.

We assume that the results are transferable to other application domains. E.g. to interact with printed maps and text-heavy documents. Our results suggest that paper-based handheld AR systems should align text to the augmented objects even if this affects the readability. To select regions of physical objects users should not be forced to use the AR visualization. Simple selection techniques using a static image of the object is clearly preferred by the users.

If the results also apply to handheld AR for 3D objects or large objects in general needs further investigation. As next steps we propose to re-examine selection techniques for handheld AR systems. It should be investigated how more complex approaches (e.g. [16]) perform compared to traditional techniques. Furthermore, potential training effects should be studied for visualizing annotations and for select regions using handheld AR.

References

1. Alessandro, M., Dünser, A., Schmalstieg, D.: Zooming interfaces for augmented reality browsers. In: Proc. MobileHCI (2010)
2. Atkins, C.: Blocked recursive image composition. In: Proc. ACMMM (2008)
3. Chin, J., Diehl, V., Norman, K.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proc. CHI (1988)
4. Crabtree, A., Rodden, T., Mariani, J.: Collaborating around collections: informing the continued development of photoware. In: Proc. CSCW (2004)
5. Davies, N., Cheverst, K., Dix, A., Hesse, A.: Understanding the role of image recognition in mobile tour guides. In: Proc. MobileHCI (2005)
6. Erol, B., Antúnez, E., Hull, J.: HOTPAPER: multimedia interaction with paper using mobile phones. In: Proc. ACMMM (2008)
7. Fitzmaurice, G.W.: Situated information spaces and spatially aware palmtop computers. Communications of the ACM 36(7) (1993)
8. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., Ariss, S.: Requirements for photoware. In: Proc. CSCW (2002)
9. Hart, S., Staveland, L.: Development of NASA-TLX: Results of empirical and theoretical research. Human mental workload 1 (1988)
10. Henze, N., Boll, S.: Snap and share your photobooks. In: Proc. ACMMM (2008)
11. Henze, N., Boll, S.: Designing a CD augmentation for mobile phones. In: Ext. Abstracts CHI (2010)
12. Henze, N., Boll, S.: Evaluation of an Off-Screen Visualization for Magic Lens and Dynamic Peephole Interfaces. In: Proc. MobileHCI (2010)
13. Henze, N., Schinke, T., Boll, S.: What is That? Object Recognition from Natural Features on a Mobile Phone. In: Proc. MIRW (2009)
14. Hull, J., Erol, B., Graham, J., Ke, Q., Kishi, H., Moraleda, J., Van Olst, D.: Paper-Based Augmented Reality. In: Proc. ICAT (2007)
15. International Organization for Standardization: Information Technology. Automatic Identification and Data Capture Techniques - Bar Code Symbology - QR Code. In ISO/IEC 18004 (2000)
16. 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)
17. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2) (2004)

18. Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, J., Juustila, A.: Like bees around the hive: a comparative study of a mobile augmented reality map. In: Proc. CHI (2009)
19. Nister, D., Stewenius, H.: Scalable Recognition with a Vocabulary Tree. In: Proc. CVPR (2006)
20. Pielot, M., Henze, N., Nickel, C., Menke, C., Samadi, S., Boll, S.: Evaluation of Camera Phone Based Interaction to Access Information Related to Posters. In: Proc. MIRW (2008)
21. Rohs, M., Gfeller, B.: Using camera-equipped mobile phones for interacting with real-world objects. In: Proc. PERVASIVE (2004)
22. Rohs, M., Oulasvirta, A.: Target acquisition with camera phones when used as magic lenses. In: Proc. CHI (2008)
23. Rohs, M., Schöning, J., Raubal, M., Essl, G., Krüger, A.: Map navigation with mobile devices: virtual versus physical movement with and without visual context. In: Proc. ICMI (2007)
24. Sandhaus, P., Boll, S.: From usage to annotation: analysis of personal photo albums for semantic photo understanding. In: Proc. WSM (2009)
25. Sellen, A., Harper, R.: The myth of the paperless office. The MIT Press, Cambridge (2003)
26. Wagner, D., Schmalstieg, D.: History and Future of Tracking for Mobile Phone Augmented Reality. In: Proc. ISUVR (2009)
27. Wagner, D., Schmalstieg, D., Bischof, H.: Multiple target detection and tracking with guaranteed framerates on mobile phones. In: Proc. ISMAR (2009)
28. Want, R., Fishkin, K.P., Gujar, A., Harrison, B.L.: Bridging physical and virtual worlds with electronic tags. In: Proc. CHI (1999)

Looking for "Good" Recommendations: A Comparative Evaluation of Recommender Systems

Paolo Cremonesi[1], Franca Garzotto[1], Sara Negro[1],
Alessandro Vittorio Papadopoulos[1], and Roberto Turrin[2]

[1] Department of Electronics and Information, Politecnico di Milano,
Via Ponzio, 34/5 - 20133 Milano (Italy)
{cremones,garzotto,negro,papadopoulos}@elet.polimi.it
[2] Moviri srl, R&D

Abstract. A number of researches in the Recommender Systems (RSs) domain suggest that the recommendations that are "best" according to objective metrics are sometimes not the ones that are most satisfactory or useful to the users. The paper investigates the quality of RSs from a user-centric perspective. We discuss an empirical study that involved 210 users and considered seven RSs on the same dataset that use different baseline and state-of-the-art recommendation algorithms. We measured the user's perceived quality of each of them, focusing on accuracy and novelty of recommended items, and on overall users' satisfaction. We ranked the considered recommenders with respect to these attributes, and compared these results against measures of statistical quality of the considered algorithms as they have been assessed by past studies in the field using information retrieval and machine learning algorithms.

Keywords: Recommender systems, quality metrics, user study.

1 Introduction

Recommender Systems (RSs) play an increasingly important role in online applications characterized by a very large amount of data - e.g., multimedia catalogs of music, products, news, images, or movies. Their goal is to filter information and to recommend to users only the items that are likely of interest to them.

Traditionally, the quality of a RS is defined in terms of objective statistical metrics, e.g., error metrics and accuracy metrics, which do not involve users and are evaluated algorithmically, using well-known techniques developed in the fields of information retrieval and machine learning (e.g., hold-out or k-fold cross-validation).

More recently, RS research is exploring user-centric directions for measuring and improving the subjective quality of RSs. A number of researchers highlight the need of a shift of perspective, suggesting that the recommendations that are "best" according to objective metrics are sometimes not the ones that are most satisfactory or useful to the users [14]. Some works [14,15] pinpoint that the quality of the User eXperience (UX) with a RS as determined by its pragmatic factors (e.g., usability) or hedonic characteristics (e.g., aesthetics and "fun") are as important, or even more

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 152–168, 2011.

important than algorithmically assessed quality to determine the user's attitudes towards a RS, and are more influential on users' decisions to use a system and to "purchase" recommended results.

User-centric approaches to RS quality evaluation have recently received some interest in the research and industry arena of RS and HCI communities. Still, empirical research in this area is currently in its early stage and a limited amount of user-based studies exist. Empirical research in this domain is rather costly, difficult in design and implementation, partially because of the intrinsic complexity induced by the high number of variables to be controlled, the computational sophistication of RSs, and the difficulty of involving need large datasets and a wide number of users.

The paper provides a contribution to this field discussing an empirical study that involved 210 users and considered 7 recommender systems, which use the *same* dataset and user interface, but implement 7 different baseline and state-of-the-art recommender algorithms. We measured the *user's perceived quality* of each RS, focusing on three attributes - perceived *accuracy*, *novelty*, and *global satisfaction*. We prioritize the considered recommenders with respect to these attributes and compare our results against the *objective statistical* quality of the considered algorithms as it has been assessed by past studies in the field, based on accuracy metrics. We discovered some interesting mismatches that suggest that objective metrics are not always good predictors of the perceived quality of RSs.

2 Background and Related Work

Recommender Systems (RSs) are generally classified into two families, characterized by different types of *recommender algorithms* [1]: *content-based filtering* (CBF) and *collaborative filtering* (CF).

In *CBF algorithms,* items are described by means of a set of explicit features. For instance, a movie can be characterized by genre, director, and list of actors. Such RSs tend to recommend items with the same characteristics as the movies a user "liked" in the past, thus they typically propose a limited variety of unexpected recommendations [9][23].

On the other hand, *CF algorithms* are based on collective preferences of the crowd: they recommend what similar customers bought or liked. Collaborative RSs are the most used, mainly because their implementation and integration in existing domains is relatively easy and their quality, in terms of objective metrics, is generally higher than CBF algorithms. However, some criticism is addressed also to CF recommenders, pinpointing that they are biased toward popularity, constraining the degree of diversity consumers would ever prefer [8].

Two families of *objective* metrics are typically adopted to automatically evaluate RSs: *error metrics* and *accuracy metrics* [9]. *Error metrics*, such as RMSE (Root Mean Square Error), measure the capability of the system to accurately estimate the ratings real users would give to item. *Accuracy metrics*, such as precision and recall, measure the effectiveness of the "top-N recommendation task" [7], i.e., the capability of a RS to *accurately select* a small set of items that the user will surely appreciate.

Both error and accuracy metrics can be automatically evaluated by means of well-known techniques developed in the field of machine learning, such as hold-out and leave-one-out (e.g., [7] and [22]). However, such standard metrics address a single property of RS's quality, relevance, while neglecting other issues that can be perceived as important by users but are more complex and articulated to operationalize.

As an alternative or complementary approach, a number of studies investigate a *user-centric* approach to RS evaluation, carrying on user-based empirical assessment or proposing conceptual frameworks for *perceived quality*.

Celma and Herrera [3] report an experiment exploring the users' perceived quality of novel recommendations provided by a CF and a CBF algorithm in the music recommendation context. Shearer [21] describes an experiment with 29 subjects on a movie RS to determine whether recommendations based on CF are perceived as superior to recommendations based on user population averages. The recommender systems suggested movies that subjects later viewed. Participants placed slightly more confidence in the CF recommendations with respect to the recommendations based on the population averages, but the perceived quality of the two algorithms was almost the same.

Ziegler et al. in [25] and Zhang in [24] propose diversity as a quality attribute: recommender algorithms should seek to provide optimal coverage of the entire range of user's interests. This work is an example of combined use of automatic and user-centric quality assessment techniques.

Pu and Chen [18] develop a framework called ResQue, which defines a wide set of user-centric quality metrics to evaluate the perceived qualities of RSs and to predict users' behavioural intentions as a result of these qualities. The framework provides 13 quality attributes and 60 questions that can be put to users for measuring them. Quality constructs are organized in four main classes: 1) "perceived *system* qualities", which refer to the functional and informational aspects of RRs (recommended items, interaction, and interface); 2) "beliefs" (user‘s perception on ease of use, usefulness, and control on interaction); 3) "attitudes" (the user's overall feeling towards a recommender, e.g., global user satisfaction, confidence and trust); 4) "behavioral intentions" (the degree at which a RS is able to influence users' decisions to implement the suggestions by the system). The framework represents an important contribution to understand the crucial factors that influence the user adoption of RSs, and provides a useful conceptual tool to guide the design and execution of user-centric evaluation studies of RSs. Several user-centric evaluations are reported in literature employing ResQue attributes [4,10,12,15,16,17]. They focus on different RSs (employing different user interfaces [15] or implementations, or datasets in different domains [16,17] (e.g., music [10] or film [12] or investigate the different perceptions of quality in culturally heterogeneous user groups [4], thus obtaining a variety of not comparable results.

As discussed in the next section, ResQue has been partially adopted also in our work. Still, our research differs from previous works in that it is more focused – it involves seven RSs which differ in terms of algorithms *only* - and also compares the results of perceived quality evaluation with objective quality measures of the considered algorithms.

3 Empirical Study of RS Perceived Quality

The general goal of the study was to compare measures of *user's perceived quality* against measures of *objective statistical quality of RSs* in order to provide some empirical evidence about the degree at which they are aligned, and to validate, or to confute, the hypothesis, underlying most existing studies, that objective statistical quality is a good predictor for user's perceived quality. The study was designed as a between subjects controlled experiment, in which we measured *perceived quality,* decomposed into a number of measurable attributes (*dependent variables*) in seven different experimental conditions, each one using a system that support the *same user interface,* employ the *same dataset* in the movie domain, but implements a *different* recommender *algorithm* (*independent variable*).

3.1 Perceived Quality Attributes

To better scoping our research, we focused our attention on three user-centric quality metrics:

Perceived accuracy (also called *Relevance*) - how much the recommendation matches the users' interests, preferences and tastes;
Novelty - the extent to which users receive "new" recommended items;
Overall users' satisfaction - the global users' feeling of the experience with the RS.

Considering the classification of the ResQue model [18], the third metric belongs to the category "Attitudes", while the first two attributes fall in the category "Perceived System Qualities", and, in particular, the subcategory "Quality of recommended items".

Our notion of perceived accuracy is meant in the same way as in the ResQue model. Still, we operationalize their measure is a slightly different way w.r.t. ResQue, as discuss in the following section.

Our concept of novelty can be regarded as a sub-dimension of ResQue novelty, which encompasses not only the idea of "new" but also of "interesting" and "surprising", the latter being referred to as "serendipity" in Herlock [9]. In addition, we distinguish between two "levels" of novelty, called respectively "First Order Novelty (FON)" and "Second Order Novelty (SON)". FON is a weaker for of novelty: it considers a movie to be *novel* for a user only if he/she has never watched it (without discriminating whether he/she has any knowledge about it). SON is more stringent concept and subsumes FON: a recommended movie is considered novel if the user has *no* idea of it. SON is a more conservative way to measure novelty, and, as we will see in the next sections, leads to lower values than FON. Anyway, it is interesting to compute the novelty in both ways and to compare the obtained results.

3.2 Algorithms

Our study considered several state-of-the-art recommender algorithms: (i) one non-personalized algorithm used as baseline, referred to as TopPop, (ii) five *collaborative* algorithms - CorNgbr, NNCosNgbr, AsySVD, and two versions of PureSVD - and (iii) a *content-based* one – LSA. In the following we provide a short description of each algorithm. Further details can be found in [5] and in the papers quoted therein.

3.2.1 Non-personalized Algorithm

TopPop (Top Popular) implements a simple, non-personalized estimation rule, which recommends the most popular items to any user, regardless his or her ratings. Such algorithm serves as baseline for the more advanced personalized algorithms.

3.2.2 Collaborative Algorithms

There are two major approaches to collaborative filtering: (i) the neighborhood approach and (ii) the latent factor approach.

Neighborhood models

Neighborhood models represent the most common approach. Rating prediction is based on the similarity relationships among either users or items, in terms of collected ratings. Item-based similarity is usually preferred to user-based similarity for its better performance in terms of RMSE and its higher scalability [20]. Prior to computing similarities, it is advised to remove a set of biases in the collected ratings, such as: (i) *user effects*, which represent the tendency of some users to rate higher than others, and (ii) *item effects*, which represent the tendency of some items to be rated higher than others. Typically, only the most similar items – referred to as neighbors - are taken into consideration. In our experiments, the neighborhood size has been set to 200.

CorNgbr (Correlation Neighborhood) is a classical technique that computes item-item similarity by means of the Pearson linear correlation coefficient [13].

Similarly, *NNCosNgbr* (Non-normalized Cosine Neighborhood) computes item-item similarity by means of the cosine coefficient. Unlike Pearson correlation - which is computed only on ratings shared by common rater - the cosine coefficient is computed over all ratings, taking missing values as zeroes. In addition, while CorNgbr averages the ratings received by similar items, NNCosNgbr simply sums up such ratings, higher ranking items with more similar neighbors [5].

Latent factor models

Latent factor models - also informally known as SVD models after the related Singular Value Decomposition (SVD) - represent users and items as vectors in a common low-dimensional ‘latent factor’ space. In such a space, users and items are directly comparable and the rating of a user u on an item can be estimated as the proximity (e.g., inner-product) between the related latent factor vectors. This family of algorithms has been leading the Netflix contest thanks to its performance in terms on RMSE.

AsySVD (Asymmetric SVD) is a powerful matrix factorization model that reported an RMSE of 0.9000 in the Netflix context. Differently from other latent factor models, AsySVD represents users as a combination of item features. Thus, AsySVD is able to immediately compute recommendations for users not yet parameterized and to adjust recommendations as fast as the user being recommended enters new ratings, providing an immediate feedback to his or her activity [13].

PureSVD is a latent factor algorithm recently proposed [5], whose rating estimation rule is based on the conventional SVD. In order to use conventional SVD – which is not defined for matrices with missing values – unknown ratings have been treated as zeros. We have tested PureSVD with two different sizes of latent factors: 50 and 300. In fact, the larger the number of latent factors the more the algorithm is able to detect the uniqueness in users' taste. The smaller the number of latent factor the more the algorithm tends to recommend the most popular items.

3.2.3 Content-Based Algorithms

Content-based algorithms recommend items whose content is similar to the content of items the user has positively rated in the past. For instance, in the domain of movies, such content can be the movie title, the playing actors, the director, the genre, and the summary. While the basic approach to content-based recommendations is based on the analysis of term-by-item occurrences, and neglecting the semantic structure of item content, more advanced techniques try to exploit such semantic features. In our experiments we have used *LSA* (Latent Semantic Analysis), a well-known method in the field of information retrieval for automatic indexing and searching of documents. The approach takes advantage of the implicit structure (i.e., latent semantic) in the association of terms with documents. Such semantic structure comes out by representing the term-by-item relationships in a low-dimensional 'latent factor' space computed through SVD [11].

3.3 Instruments

3.3.1 Technological Framework

To run our experiments, we used a web-based commercial recommender framework - called ContentWise (www.contentwise.tv - Figure 1). ContentWise supports users with a wide range of typical RS functionalities, such as browsing a catalog of products, retrieving the detailed description of each item, rating it, getting recommendations and rating their relevance. The modularization and customization features of the system allowed us to easily create different experimental conditions by implementing different algorithms while maintaining interface and dataset invariant.

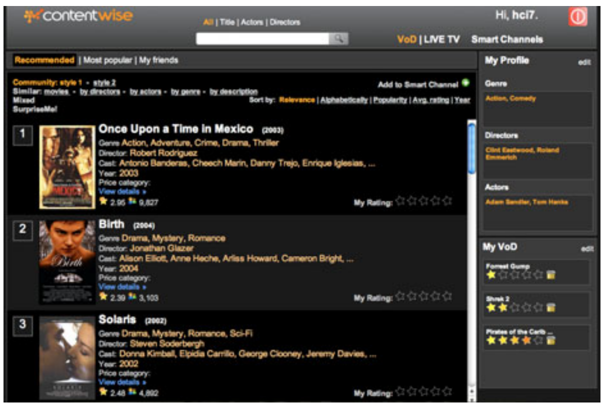
3.3.2 Dataset

The dataset included 2137 movies and about 7.7 million ratings given by 49,969 users. The data consider a subset of the well-known large-scale movie dataset Netflix, published for the purpose of the famous contest organized by the homonymous movie rental American provider. In addition, for the purpose of our study, these contents were integrated with data and metadata (e.g., movie plot, images, actors, director and genre) collected online.

3.3.3 Data Collection Technique

As discussed more precisely in the following section, the chosen user-centric metrics were measured using a *questionnaire* that evaluators completed for each user during the experiment. It collects both users' demographic attributes and their opinions about *perceived accuracy*, *novelty* and *overall satisfaction.*

(a)



(b)

(c)

Fig. 1. ContentWise interface: a) initial exploring and rating of movies (up); b) results of recommendations; c) movie details

3.4 Participants

Data collection was carried on by a team of 14 master students (two per experimental condition) selected among the best ones of those attending two courses - HCI and iTV - at our School of Information Engineering. Students were motivated in performing the evaluation to the best of their capabilities, for a number of reasons. This work represented the second assignment proposed at our courses, and accounted for 50% of the final mark. Students were pre-screened as they had to pass brilliantly the first part of our exam in order to be eligible for performing this one. In addition, they freely selected this assignment from a set of others proposed by teachers.

Students were initially trained by us to perform the study, were given written instructions on the evaluation procedure, and were regularly supervised by a teaching assistant during their activities. After a pre-screening among school maters, friends and relatives, each pair of student evaluators recruited a group of *thirty* subjects for each algorithm, almost uniformly distributed w.r.t. to gender and age. Overall, the study involved *210* users aged between 20 and 50; 54% subjects were male and 46% female. None of them had been previously exposed to the system used in our study nor had technical knowledge about RSs.

3.5 Procedure

The evaluation took place in informal environments such as university (15%), interviewer's place (32%), and interviewee's place (31%). Each interview lasted from 15 to 35 minutes. The motivation for such a temporal variability is that in case of completely novel recommendations, users were invited to explore information related to unknown items (see below) in order to express more precise and conscious opinions on the quality of the RS used.

Each participant was initially asked to provide his/her personal information (age, gender, education, nationality, and number of movies watched per month). Afterward, (s)he was invited to browse the movie catalog using the ContentWise system (pre-customized on a specific algorithm). The user was then asked to freely select five known (not necessarily watched) movies and rate his/her degree of appreciation or interest for them using a 1-5 point scale (1 = low interest for/appreciation of the movie; 5 = high). On the basis of these ratings, five recommendations were returned by the system (using the current algorithm). The user was finally invited to explore the results and reply to a set of questions related to the quality of the recommendations.

Novelty measures were collected as follows. For each recommended item we first asked the question "Have you ever watched this movie?" This answer (yes/no) was used to compute First Order Novelty (*FON*). FON for an item is 1 if the user has *never* watched the movie and 0 otherwise. If the user has *never* watched a recommended movie (FON=0), we proceeded with an in depth exploration to assess Second Order Novelty (SON). We asked the user if (s)he had ever *heard* about the movie, inviting him/her to explore the information related to the movie (director, cast, abstract, trailer,...) to refresh her memory. If a user answers "yes" to the above question (or if FON is 0), SON is set to 0, while it is set to 1 otherwise.

Perceived accuracy measures were collected as follows. For each recommended movie, if the user had already watched it, he/she was asked to rate how much he/she liked/disliked (on a 1-5 scale). Otherwise, if the user had already seen the trailer, he/she was invited to rate the degree of potential interest for the movie. If the user had never been exposed to the movie or its trailer, he/she was invited to look at the trailer and to explore additional information (e.g., director, the actors, and so forth) and then to give a rating of potential interest. For each user, each of the above attributes – FON, SON and perceived accuracy is calculated as the average on the respective values assigned to each recommended item.

Finally, the overall satisfaction was computed by asking each subject to provide a global judgment (in a 1-5 rating scale) about the list of recommended movies and was allowed to express a free comment.

4 Empirical Study Results

In this section we present the user-centric metrics of relevance and novelty computed on the basis of the questionnaire data.

4.1 Accuracy

Users' perception of the accuracy of the recommendations is measured by considering, for each user and each suggested movie, the user's opinion on the movie expressed in a 1-5 scale. Figure 2 shows the box plot of the perceived relevance for each algorithm. Upper and lower ends of boxes represent 75th and 25th percentiles. Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range. Median is depicted with a solid line, mean with a dot. Outliers are represented with empty circles.

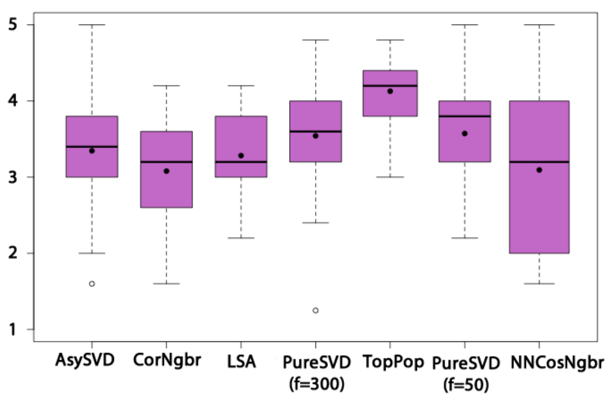


Fig. 2. Perceived relevance for each algorithm. Relevance ranges from 5 (most relevant) to 1 (not relevant).

We can see that all the algorithms have an average relevance between 3 and 4. This result shows that, on average, users are satisfied by the quality of the recommendations (the median for all the algorithms is greater than 3). Moreover, 75%

of the users have received relevant recommendations from five of the considered algorithms (AsySVD, LSA, PureSVD50 PureSVD300 and TopPop). Only NNCosNgbr produces a relatively large number of bad recommendations (25% of the recommendations are rated 2 or less).

The most surprising result is the TopPop algorithm, having the largest perceived accuracy. This result is surprising because the TopPop algorithm suggests to all users the same list of 5 movies, without taking into consideration the user profile. These movies are: "Pirates of the Caribbean: The Curse of the Black Pearl"; "Forrest Gump"; "The Lord of the Rings: The Two Towers"; "The Lord of the Rings: The Fellowship of the Ring"; "The Sixth Sense".

According to our study, any user found in this list an average of four interesting movies (more than 80% of the users rated TopPop recommendations with 3 or more stars). This result may provide evidence against the real usefulness of sophisticated recommender algorithms, a hypothesis that will be further analyzed in the following paragraphs and in the discussion section.

4.2 Novelty

A similar analysis was performed for perceived novelty. Novelty refers to the previous knowledge of the user about the suggested movies. Unlike relevance, novelty measures are based on two questions, respectively responded with either "yes" (novelty value = 1= totally novel recommendations) or "no" (novelty value = 0= no novel recommendations). Figure 3 shows the box plot of perceived first-order novelty (percentage of never-watched movies in the recommendation list). Similarly, Figure 3 shows the second-order novelty (percentage of never-heard-of movies in the recommendation list). By definition, first-order novelty (FON) is always greater than or equal to second-order novelty (SON). The two metrics provide quite similar results and are strongly correlated (the correlation factor is 0.8).

If perceived relevance was on average satisfactory across all the algorithms, the same cannot be said for novelty. On average, no algorithm was able to suggest more

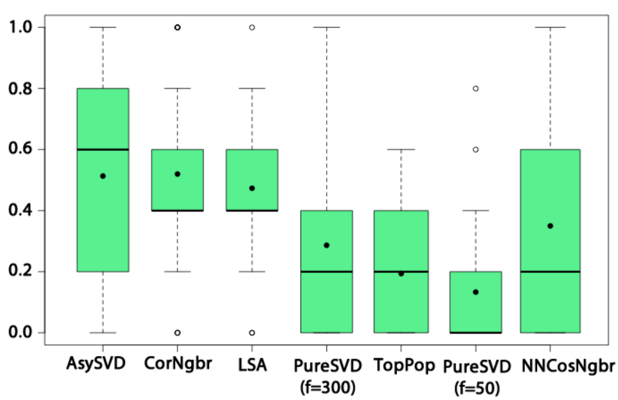


Fig. 3. Perceived First-Order Novelty (never watched) for each algorithm

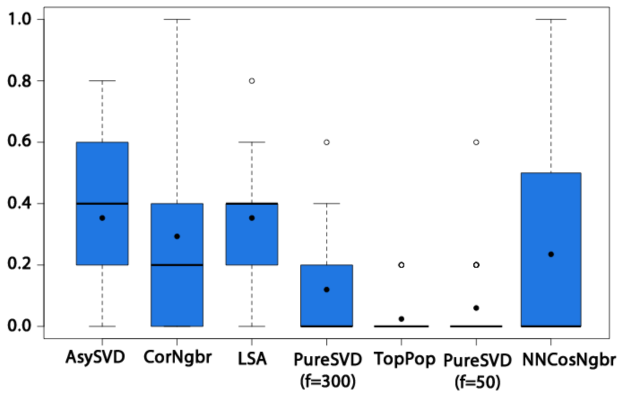


Fig. 4. Perceived Second-Order novelty (never heard of) for each algorithm

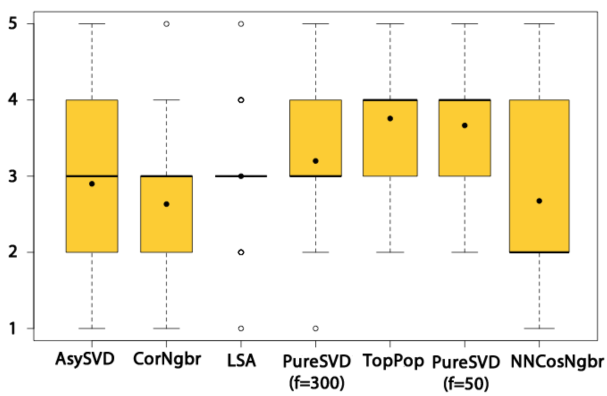


Fig. 5. Global Satisfaction for each algorithm. Satisfaction ranges from 5 (totally satisfied) down to 1 (no satisfied at all).

than 40% of totally-unknown-to-the-user movies (SON). Moreover, TopPop and PureSVD50 were not able to suggest novel movies at all (with the exceptions of few outliers, SON is always 0%).

Finally, Figure 5 shows the results for global user satisfaction. User satisfaction was measured according to a 1-5 points scale. Contrary to the relevance, the collected responses have a large variance and there seems to be no agreement in users' opinion, at least with AsySVD, CorNgbr, LSA and PureSVD300 (median and average equal or close to 3). In order to better compare the results, we first used 1-way ANOVA. The test suggests that, for each of the dependent variables, at least one of the algorithms differs significantly with respect to the others. We run multiple pair-wise comparison post-hoc tests using Tukey's method. All tests were run using a significance level $\alpha = 0.05$. Although no algorithm is significantly better (or worse) than all the other in terms of any of the quality dimensions, we can at least identify a partial order, as outlined in Table 1.

Table 1. Partial Ordering of RSs w.r.t. the various quality attributes

	Accuracy	**Novelty**	**Global satisfaction**
Maximal	TopPop	AsySVD LSA CorNgbr	PureSVD50 TopPop
Intermediate	AsySvd PureSVD300 PureSVD50	NNCosNgbr PureSVD300	PureSVD300 LSA
Minimal	NNCosNgbr LSA CorNgbr	PureSVD50 TopPop	AsySVD CorNgbr NNCosNgbr

According to this ordering, TopPop is the maximal algorithm in term of relevance (i.e., the algorithm with the best perceived relevance), while NNCosNgbr, CorNgbr and LSA are the minimal algorithms (i.e., the algorithms with the worst perceived relevance).

We have performed the same comparison for first-order and second-order novelty. Being the two novelty metrics correlated, the comparisons define the same partial ordering. According to this ordering, AsySVD, CorNgbr and LSA are the algorithms with the best perceived novelty, while TopPop and PureSVD50 are the algorithms with the worst perceived novelty.

The last column of the table shows the partial ordering according to the global satisfaction; TopPop and PureSVD300 are the algorithms which mostly satisfied the users, while AsySVD, CorNgbr and NNCosNgbr are the algorithms which less satisfied the users.

5 Objective Evaluation of Quality

5.1 Objective Metrics and Their Evaluation Method

As mentioned in section 2, RS performance is traditionally usually measured using objective metrics. In particular, there are methodologies based on accuracy metrics (e.g., precision, recall and fallout) and on error metrics (e.g., RMSE and MAE). Some of the algorithms tested in this study (TopPop, NNCosNgbr and PureSVD) cannot be evaluated with error metrics [9]. Hence, we considered only accuracy metrics in our study. In particular, we focused our attention on recall r (the conditional probability of suggesting a movie given it is relevant for the user) and on fallout f (the conditional probability of suggesting a movie given it is irrelevant for the user).

A good algorithm should have high recall (i.e., it should be able to recommend items of interest to the user) and low fall-out (i.e., it should avoid to recommend items of no interest to the user).

A measure that combines recall and fall-out is the F-measure. F-measure is defined as the harmonic mean of precision and recall. Precision can be estimated from recall and fall-out by using the definition provided in [19].

The testing methodology adopted in this study is similar to the one described in [5]. The known ratings of the dataset are split into two subsets: training set *M* and test set *T*. The test set *T* contains only 5-stars ratings. Therefore we can reasonably state that *T* contains items relevant to the respective users. The detailed procedure used to create *M* and *T* from the Netflix dataset is similar to the one used for the Netflix prize, maintaining compatibility with results published in other research papers [2,5].

In this work, the training set *M* is a subset of the original Netflix training set, while the test set *T* contains only part of the 5-stars ratings from the Netflix probe-set. The test set contains 69,039 5-star ratings.

In order to measure recall, we first trained the algorithm over the ratings in *M*. Then, for each item *i* rated 5-stars by a user *u* in *T*, we followed these steps:

1. We randomly selected 1,000 additional items unrated by user *u,* assuming that the user *u* is not interested in most of them.
2. We predicted the ratings for the test item *i* and for the additional 1,000 items.
3. We formed a top-5 recommendation list by picking the 5 items with the largest predicted ratings.

The overall recall *r* was computed as

$$r = \frac{\text{\# times the element is in the list}}{\text{\# elements in } T}$$

A similar approach was used to measure fall-out, with the only difference being in the composition of the test set *T*, that now contains only 1-stars ratings. The fall-out *f* is computed as

$$f = \frac{\text{\# times the element is in the list}}{\text{\# elements in } T}$$

5.2 Objective Metrics Evaluation Results

Table 2 presents the objective accuracy of the tested algorithm. Algorithms in the table are ordered in decreasing order of recall. Recall and F-measure suggest

Table 2. Recall, fallout and F-measure computed for Top-5 recommendation lists

	Type	Recall	Fallout	F-measure
PureSVD50	Collaborative Latent factors	0.29	0.005	0.45
PureSVD300	Collaborative Latent factors	0.25	0.005	0.40
AsySVD	Collaborative Latent factors	0.13	0.001	0.23
NNCosNgbr	Collaborative Item-based	0.12	0.010	0.21
TopPop	Collaborative Non-personalized	0.11	0.025	0.20
CorNgbr	Collaborative Item-based	0.08	0.010	0.15
LSA	Content	0.01	0.002	0.02

PureSVD as being the most accurate algorithm. Second in line are AsySVD, the two item-based neighborhood algorithms and the non-personalized TopPop algorithms, all of them with a similar recall. The content-based LSA algorithm has the worst accuracy both in terms of recall and F-measure. If we look at fallout, AsySVD and LSA obtain the best results, while NNCosNgbr and TopPop are the algorithms with the largest error rate.

6 Discussion

The analysis of the results presented in the previous sections suggests a number of interesting considerations: 1) simple, non-personalized algorithms are well perceived by the users; 2) the perceived novelty of content-based recommendations is equal or even better with respect to collaborative recommendations; 3) objective accuracy metrics (e.g., recall and fallout) are not a good approximation of user perceived quality.

Let's start from the first point. According to Figure 2, no algorithm is significantly better (or worse) than all the others in terms of *perceived relevance*. However, the partial ordering among the algorithms (Table 1) highlights that *TopPop* is the algorithm with the best perceived relevance (this is unexpected) and with the worst novelty (as expected), thus its utility is limited because oftentimes the user has already watched the suggested items. Still, *TopPop* (together with *PureSVD300*) is at the top level in terms of global user satisfaction. In summary: *simple non-personalized TopPop recommendations are better perceived by the users with respect to other more sophisticated and personalized recommender algorithms,* although users are aware of the low utility of such recommendations. Global user satisfaction seems mainly driven by the perceived accuracy than by the novelty of the recommendations. This is a somehow surprising result, especially if we consider the large academic and industrial effort in the development of new and more sophisticated recommender algorithms.

As for novelty, Table 2 highlights that AsySVD, CorNgbr and LSA are the algorithms with the best perceived novelty, while TopPop and PureSVD50 are the algorithms with the worst perceived novelty. Thus, the perceived novelty of content-based recommendations is equal or even better with respect to collaborative recommendations. *This result is in contrast with most of the existing literature in RS, which considers content-based algorithms as not able to recommend novel items* (see, e.g., [9] and [23]). To try an interpretation of this result, we should consider that collaborative algorithms, by design, are biased toward popular "Blockbuster" items, thus reducing the chances of novel recommendations. Collaborative algorithms are trained (e.g., tuned) to achieve the best performance in terms of objective accuracy. Because objective accuracy is computed on already-rated items, collaborative algorithms cannot recommend items with limited historical data. This creates the rich-get-richer effect for popular items and the opposite effect for unpopular ones, which results in lower novelty. As a consequence, collaborative algorithms tend to reinforce the popularity of already popular items and to recommend mainly common movies, which are likely not to be novel.

Finally, the comparison between Tables 1 and 2 shows the lack of correspondence between *objective accuracy metrics (e.g., recall and fallout) and users' perceived quality*. In other words, objective quality attributes are not good predictors of users' perceived quality of a recommender algorithm, at least in our case. To try an interpretation of this phenomenon, it is useful to consider that objective metrics compute accuracy of recommendations by (i) exploiting previously rated movies, i.e., user's rankings of movies that they know about, and (ii) sampling all the ratings in the dataset - the majority of which concern few popular movies. Consequently, objective metrics focus their attentions on measuring the quality of an algorithm when recommending popular items and might not be particularly effective for measuring the quality of the same algorithm when recommending novel, unrated items.

7 Conclusions

In this work we have investigated under different perspectives the quality of 7 RSs that only differ in terms of recommender algorithms. We first measured quality from a user-centric perspective and then compared these results against measures of statistical quality, in terms of recall and fallout. The considered RSs include both state-of-the-art techniques and a trivial non-personalized recommender algorithm. There are three main interesting findings:

(i) the simple, non-personalized algorithm is well perceived in terms of overall user satisfaction, although users are aware of the low utility of such recommendations;
(ii) the perceived novelty of content-based recommendations is equal or even better with respect to collaborative recommendations;
(iii) statistical accuracy metrics (e.g., recall and fallout) are not necessarily a good approximation of the quality perceived by the users.

Our research has its limitations. First, the sample size of participants used for each RS (30) is relatively small. Still, the fact that we replicated the study in seven experimental conditions using the same methodological framework, and involving overall 210 tested subjects, partially compensates for this drawback and strengthens the reliability of our results. Second, we focused our investigation of user perceived quality on a small set of attributes – perceived accuracy, novelty, and overall user satisfaction. Other approaches, e.g., the ResQue model, include many additional user-centric metrics, which we did not consider in our study. Our choice may be regarded as a weakness, but it was motivated by the need to keep data collection workload affordable. ResQue provides a 60 items questionnaire. Administrating so many questions could have been too demanding for respondents and too time-consuming for data collectors, considering our goal of collecting measures in 7 experimental conditions. In addition, we sought to focus on those user-centric attributes that are more related to standard objective quality metrics and thus are more comparable with them. Objective metrics are related to the quality of recommend items, thus we gave higher priority to measuring ResQue attributes related to these aspects. Certainly, other measures of user perceived quality are worth being investigated in relationship to objective quality, and we are planning to replicate our study in order to include them.

In spite of the above limitations, our work provides contributions both from a research and practical perspective. To our knowledge, this is the first work that systematically compares perceived quality in a significant number of different RSs isolating a precise factor – the underlying recommender algorithm – and analyzing the results against statistical, objective measures of quality. For the practice of RS design and evaluation, our results may promote further approaches that move beyond the attention to conventional accuracy metrics and shift the emphasis to more user-centric factors.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowledge and Data Engineering 17(6), 734–749 (2005)
2. Bennett, J., Lanning, S.: The Netix Prize. In: Proceedings of KDD Cup and Workshop, pp. 3–6 (2007)
3. Celma, Ò., Herrera, P.: A new approach to evaluating novel recommendations. In: RecSys 2008: Proc. of the 2008 ACM Conf. on Recommender Systems, pp. 179–186. ACM, New York (2008)
4. Chen, L., Pu, P.: A cross-cultural user evaluation of product recommender interfaces. In: Proc. of the 2008 ACM Conf. on Recommender Systems, RecSys 2008, pp. 75–82. ACM, New York (2008)
5. Cremonesi, P., Koren, Y., Turrin, R.: Performance of recommender algorithms on top-N recommendation tasks. In: RecSys 2010: Proc. of the Fourth ACM Conf. on Recommender Systems, pp. 39–46. ACM, Barcelona (2010)
6. Cremonesi, P., Turrin, R.: Analysis of cold-start recommendations in IPTV systems. In: RecSys 2009: Proc. ACM Conf. on Recommender Systems, pp. 233–236. ACM, New York (2009)
7. Deshpande, M., Karypis, G.: Item-based top-N recommendation algorithms. ACM Trans. Information Systems (TOIS) 22(1), 143–177 (2004)
8. Fleder, D., Hosanagar, K.: Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. Management Science 55(5), 697–712 (2009)
9. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. on Information Systems (TOIS) 22(1), 5–53 (2004)
10. Hu, R., Pu, P.: Acceptance issues of personality-based recommender systems. In: Proc. of the Third ACM Conf. on Recommender Systems, pp. 221–224. ACM, New York City (2009)
11. Husbands, P., Simon, H., Ding, C.H.Q.: On the use of the singular value decomposition for text retrieval. Computational Information Retrieval, 145–156 (2001)
12. Jones, N., Pu, P.: User Technology Adoption Issues in Recommender Systems. In: Proc. of the 2007 Networking and Electronic Commerce Research Conf., Riva del Garda, Italy, pp. 379–394 (2007)
13. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: KDD 2008: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426–434. ACM, New York (2008)
14. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 1097–1101. ACM, New York City (2006)

15. Pu, P., Chen, L.: Trust building with explanation interfaces. In: Proc. of the 11th int. Conf. on Intelligent User Interfaces, IUI 2006, pp. 93–100. ACM, New York (2006)
16. Pu, P., Chen, L., Kumar, P.: Evaluating product search and recommender systems for e-commerce environments. Electric Commerce Research Journal 8(1-2), 27 (2008)
17. Pu, P., Zhou, M., Castagnos, S.: Critiquing recommenders for public taste products. In: Proc. of the Third ACM Conf. on Recommender Systems, RecSys 2009, pp. 249–252. ACM, New York (2009)
18. Pu, P., Chen, L.: A User-Centric Evaluation Framework of Recommender Systems. In: Proc. of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI), Barcelona, Spain (September 2010)
19. Raghavan, V., Bollmann, P., Jung, G.S.: A critical investigation of recall and precision as measures of retrieval system performance. ACM Trans. Inf. Syst. 7, 205–229 (1989)
20. Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: 10th Int. Conf. on World Wide Web, pp. 285–295 (2001)
21. Shearer, A.W.: User response to two algorithms as a test of collaborative filtering. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, pp. 451–452. ACM, New York (2001)
22. Takács, G., Pilászy, I., Németh, B., Tikk, D.: Scalable collaborative filtering approaches for large recommender systems. The J. of Machine Learning Research 10, 623–656 (2009)
23. Weng, L., Xu, Y., Li, Y., Nayak, R.: Improving recommendation novelty based on topic taxonomy. In: Proc. of the 2007 IEEE/WIC/ACM International Conf. on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW 2007, pp. 115–118. IEEE Computer Society, Washington, DC, USA (2007)
24. Zhang, Y., Callan, J., Minka, T.: Novelty and redundancy detection in adaptive filtering. In: Proc. of the 25nd ACM SIGIR Conf. on R&D in Information Retrieval, pp. 81–88. ACM, New York City (2002)
25. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proc. of the 14th International Conf. on World Wide Web, pp. 22–32. ACM, New York (2005)

All the News That's Fit to Read: Finding and Recommending News Online

Juha Leino, Kari-Jouko Räihä, and Sanna Finnberg

Unit for Computer-Human Interaction, School of Information Sciences
FIN-33014 University of Tampere, Finland
{juha.leino,kari-jouko.raiha,sanna.finnberg}@uta.fi

Abstract. Our survey study of 147 Finns shows that online news is becoming the most important news source today: Online newspapers have bypassed paper newspapers and also TV and radio in importance, especially among young adults. Although most respondents routinely visited their preferred news sites directly, recommendations from their social network also played an important role in helping them find salient news. We analyzed the factors that affected which recommendations were read and why, and also discuss participants' expectations on the behavior of the receivers of the recommendations. The person recommending and the means of recommending affect what gets read. In contrast with previous studies, we found that the role of email as a recommendation tool is decreasing as the use of social media is becoming more common. However, personally targeted recommendations still have a better chance of being influential than recommendations made to the public at large.

Keywords: News, recommending, sociality, online.

1 Introduction

Online newspapers have become a major source of news [19] while media houses struggle to find new earning models to replace plummeting paper newspaper sales in an environment that is used to free access. Simultaneously, the huge number of possible news outlets and news items online means that we need efficient ways to find news items that we are interested to read.

News recommending is an important part of online newspapers' strategy today, as attested by the number of news sites that integrate recommending and social aspects. Huffington Post's integration of Facebook and NYTimes's TimesPeople that allows following other users and editors are popular examples, but many others, such as BBC and Reuters, are also in the game.

We conducted a survey study to see how users in Finland today keep up with news and how they find salient news items online. The most popular method of finding news was to access news sites that the user knew and liked directly. In addition, nearly half of the 147 respondents typically also used recommendations from other people. While recommender sites, aggregation services, and feeds were all used to locate news by some of the respondents, their share in the equation pales in

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 169–186, 2011.

comparison to the recommendations from other people. The popular channels for giving and receiving recommendations were social media and networks, email, IRC, and instant messaging (IM).

In fact, news recommending appears to be part of the normal social intercourse [4] and is used to maintain relations and frequently share emotions. Recommendations are often related to on-going conversations or meant to be used as material for later conversations. These conversations take place both face-to-face and online, sometimes starting in one and continuing in the other. Thus, it is easy to see why increasingly ubiquitous social media plays such an important role in news recommending.

The more personal and direct the recommendation is, the more likely the item is read. On the other hand, when the recommendation is broadcast to the world-at-large, the item itself becomes the point of decision—we ask ourselves if this is an item that interests me. While social media leads in the number of recommendations made in it, its lead in effectiveness is not as clear. Consequently, when designing ways to recommend news, especially in social media, giving people information about the item behind the link is important to help them make up their mind about reading it. Also, giving recommendations a personal feeling helps.

We undertook this study to better understand online news recommending practices as we were designing a news recommending solution in a research project. We wished to understand the underlying dynamics of news recommending to design a system that answers actual user needs rather than imposes new dynamics on users. This paper focuses on the survey results and design implications on a general level.

After reviewing briefly the related work and introducing our method, we dive into the results. We first discuss the media the respondents used to keep up with the news and then look at how they found salient news items online. Recommending practices, how and why news items are recommended, are studied in more detail. After discussing which news recommendations get read and why, we look at the implications the current practices have for supporting news recommending activity online.

2 Related Work

Today we have access to countless different news sources online and face the challenge of finding interesting news items from the seemingly limitless number of items [2, 15, 19]. This problem of information overload has given birth to such services as Google News and Digg.com, and the research community is constantly looking for new ways to improve news recommendations [18], e.g. by combining collaborative and content-based filtering with each other [e.g. 15] or with other approaches, today increasingly with social network information [e.g. 7, 8, 16, 18].

The phenomenal growth of social networking [7] has resulted in them "fast overtaking traditional web as the preferred source of information" [6] and "transforming the way information is created and distributed" [13]. Social scientists have long known that social networks are central to the spreading of information [13], and today online networks are central to information dissemination, search, marketing, and discovery [14]. In addition, social networks are used for spreading and recommending news [10]. In fact, today social media sites are often "the first to break the important news" [14]. In contrast, Bernstein et al. [1] found that email is the preferred medium for content sharing and that social networking sites are used for that to a much lesser degree.

In real life, we rely on recommendations from our friends and family in decision making [16]. The person recommending an item also matters online, as we are interested in items on which our friends have shown interest [7, 8]. Sharing online is after all a social act meant to develop social capital by strengthening weak ties (bridging social capital) and maintaining strong ties (bonding social capital) with friends and family, although especially social networking is more associated with bridging than bonding social capital [1, 3]. In fact, familiarity network is a clearly better indicator of user interest than similarity network [8]. In contrast, collaborative filtering does not distinguish friends from other neighbors who share similarities with the user [16]. The explicit social information and social processes offered by social networking sites lend themselves well to recommendation generation and can be a reliable source of recommendations [2, 7].

Consequently, many social networking sites are adding recommending features [8]. Facebook, for example, is encouraging users to "set up their Facebook accounts for news reading" [12], and pundits are discussing the possibilities of Facebook becoming the "world's leading news reader" [12]. Furthermore, Hitwise data shows that Facebook was in fact the 3rd biggest source of visits to news and media sites while Google News was 11th and Twitter 39th (March, 2010—one week's data) [9].

In the beginning of social networks, users mainly posted personal information but today these networks "have metamorphosed into a forum where people post information such as news that they deem to be of common interest" [6]. In effect, users have become strategic thinkers who weigh in various factors when deciding if and what to post and propagate [6]. Thus, we need to understand better the characteristics and driving motivations behind the user activity and the effects of the social networks [14].

3 Method and Participants

The study was carried out as an online survey designed to provide foundation for designing a news sharing and recommending service. The survey topics concerned reading news (online and offline), recommending news, receiving news recommendations, recommender systems, and the impact of recommendations. The survey form was built so that the respondents needed to answer only the relevant questions. For instance, if the "Yes" radio button was ticked on a question about using news aggregators, the form opened to display more detailed questions (2–13 questions per topic), but if "No" was ticked, no further questions about the topic were asked.

The form had 29 questions where different properties (such as the importance of various news media or the importance of a recommender's identity) were evaluated with a 7-point interval scale (1–7; 7 signifying strong agreement) while activity frequencies were asked with a 5-point scale. All these non-open-ended questions had an openable text field for commenting. In addition, there were several open-ended questions with text fields for answering. An English translation of the survey form, originally in Finnish, is available at: http://tinyurl.com/6e9n5af.

The participants were solicited at three Finnish universities through mailing lists distributed to staff and students in several departments. Four movie tickets were raffled among the respondents. We received 147 responses (from 83 males and 63 females). Fifty-eight respondents were students and further 36 worked at a university.

Our respondents were on the younger side with the mode being the 20–29 age group (Table 1). The slant towards younger respondents means that we have to be careful when contrasting age groups, given the smallness of Below 20, 50–59, and Over 60 age groups. Consequently, our results as a whole represent more the attitudes of adults and young adults than those of high-schoolers (or below) or seniors. Therefore, the results offer a good foundation for designing, as they provide a glimpse of up-and-coming use practices.

Table 1. Participants by age (one did not provide) and gender (one did not provide)

	< 20	20-29	30-39	40-49	50-59	60 >	Total
Males		40	32	7	2	1	**83**
Females	4	37	15	5		2	**63**
Total	**4**	**77**	**48**	**12**	**2**	**3**	**146**
Percentage	2.72%	52.38%	32.65%	8.16%	1.36%	2.04%	

Survey-studies face well-known problems: We cannot be sure how seriously the respondents took their task, and the results are based on self-report, not observation. On the other hand, there is no other practical way of collecting such information in large scale. To improve the reliability of the data, the responses were screened for bogus entries; most respondents included verbal comments indicating that they had given considered responses.

When we study representatives of one culture, as most studies in fact do, we have to be careful when generalizing the results to other cultures. Culture is in many ways a significant factor in communication, as collectivist cultures (e.g. South Korea) emphasize relationship-building aspects in communication while individualistic cultures (e.g. USA) focus on information [5, pp. 196–205]. Scandinavian countries tend to be amalgamations of both approaches [11].

4 Reading News

Our results show that online news constitutes today the most important source for news, at least for young adults. While TV and paper newspapers are still clearly in the picture, their grasp on the audiences is weakening, especially when it comes to younger people. Table 2 summarizes how important the respondents found different media for themselves in following news on average and by age and gender groups.

Overall, online newspapers were clearly the most important medium for following news. It had both the highest rating (5.99) and the lowest standard deviation (STDEV) (1.28). Paper newspapers (4.56/1.96) and TV (4.13/1.96) followed online newspapers, but with lower ratings and higher STDEVs, underlining that they were important only to some respondents. The same applies to online tabloids (4.12/2.10) that had the

Table 2. Importance of different media by age and gender (differences between all distributions are statistically significant except for paper newspaper–Internet tabloid and TV–Internet tabloid pairs)

	TV	Radio	Newspaper (paper)	Newspaper (Internet)	Tabloids (paper)	Tabloids (Internet)
AGE (146 respondents as one did not provide age)						
<20	5.75	3.75	6.50	6.75	3.00	4.00
20-29	3.94	2.91	4.36	6.14	2.30	4.08
30-39	3.75	2.96	4.30	6.08	2.21	4.25
40-49	5.42	3.33	5.75	4.75	1.58	3.33
50-59	5.50	4.00	6.00	4.00	2.50	4.00
60>	7.00	4.33	6.00	6.33	6.00	6.33
Below 40	3.92	2.95	4.41	6.14	2.29	4.14
40 and over	5.71	3.59	5.82	4.94	2.47	3.94
GENDER (146 respondents as one did not provide gender)						
Female	4.78	3.46	4.87	6.02	2.57	4.54
male	3.65	2.70	4.30	5.95	2.13	3.83
TOTAL SAMPLE (147 respondents)						
AVG	**4.13**	**3.03**	**4.56**	**5.99**	**2.32**	**4.12**
STDEV	*1.96*	*1.79*	*1.96*	*1.28*	*1.49*	*2.10*

highest STDEV in the sample. (Tabloids in Finland are between traditional newspapers and yellow journalism—credible sources of news but tend to go for a more scandalizing angle.) Online tabloids were clearly more important than paper tabloids (2.32/1.49).

Together, online tabloids and newspapers are a more important source of news than their paper cousins and, in fact, dominate as a source of news for the respondents on average. While online newspapers are a more important source of news for the younger, practically all read them (96%).

TV is still an important source of news for many as are paper newspapers. Paper newspapers were read by 119 (81%) respondents and subscribed to by 59 (40%). Still, 19% claimed never to read them while only 4% claimed not to read online newspapers.

4.1 Computers vs. Mobiles for Reading Online News

On average, the 96% of the respondents who read online papers spent 1 h 10 min a day to read them on PC. The respondents who read online news on a mobile phone (27%) spent on average 15 minutes a day reading news on it.

Altogether 102 respondents (69%) had an Internet connection on the mobile, but only 38% of them read news with the mobile. The four respondents who described their use mentioned reading news when on the move, e.g. sitting in a bus or train.

4.2 How the Participants Found Online News

We asked the respondents how they *typically* found the news they read (Table 3). On average, each respondent used 2.5 ways to find online news. Although the respondents predominantly found news by going to news sites they knew and visited regularly, recommendations also had an important role, as 70 respondents found news through recommendations from others, with 22% of them receiving recommendations *Several times a day* and 53% at least *Several times a week*. Moreover, 15 used recommendation systems, such as Slashdot.org or digg.com, and 15 found news by newsletters from news services.

Table 3. How participants *typically* found online news (later in the paper, the numbers of users for a medium may be higher than here, as here we show the number of respondents who mention using the method *typically*)

Method	Users
Accessing online papers directly	123 (84%)
Accessing TV news online services directly	68 (46%)
RSS etc. feeds	32 (22%)
News aggregation services	30 (20%)
Recommendations from people	70 (48%)
Recommendation systems	15 (10%)
Newsletters from the news services	15 (10%)
Other	8 (5%)

Of the *News aggregation services*, Ampparit (a Finnish news aggregator) was the most popular (25 mentions) and Google News the second most popular (5 mentions). Respondents overwhelmingly used only one news aggregator.

Among the *RSS feeds* mentioned by name (one user mentioned using over 400 feeds and another about 80, and predictably neither named them), the most popular ones were HS (the largest-circulation newspaper in Finland) (8 mentions), YLE (Finland's national broadcasting company) (6 mentions), and BBC (6 mentions).

5 Recommending News

Roughly one third of the respondents (31%) recommended news at least *Several times a week* and over half (59%) recommended them at least *Several times a month*. There is a clear symmetry and reciprocity in making and receiving recommendations: The respondents who made more recommendations also received more of them, and the respondents who made fewer recommendations also received fewer. The correlation is highly significant ($r = .5943$, $p < .0001$).

In addition, the respondents reported receiving more recommendations than making them (which might be at least partially explained by one person often recommending one item to more than one person, considering that Social media and IRC were the most popular ways of recommending).

Reciprocity was not as evident in the rating of how important the respondents considered recommending news back to people from whom they had received recommendations (3.75). On the other hand, respondents did find being able to comment or otherwise acknowledge the recommendation rather important (4.54). In addition, when asked why they recommended news, two respondents stated they did it to get recommendations back. Moreover, another two respondents mentioned having stopped giving recommendations to somebody because of lack of response. Thus, reciprocity appears not limited to recommending news back but also involves other means of responding to the recommendation, as responding showed appreciation.

5.1 How Respondents Made News Recommendations

Figure 1 shows how the respondents recommended news. *Social media* and *IRC* are the winners, closely followed by *IM*. *Tell a friend* (emailing through the news site) type of features, on the other hand, were not popular, as further attested by the fact that in *Other*, giving the recommendation face-to-face was mentioned 21 times, consequently making it more typical than using *Tell a friend*.

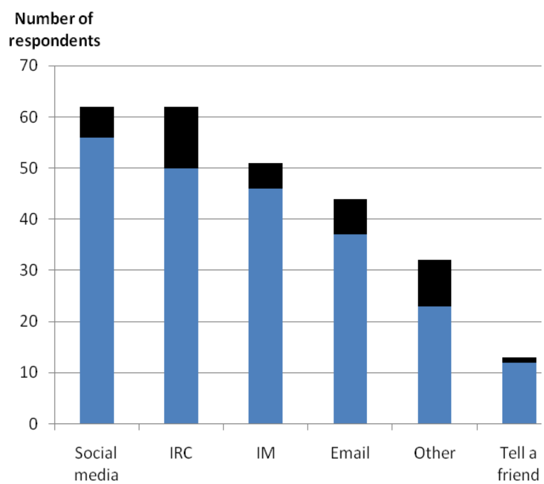


Fig. 1. Recommending means used by respondents (darker top area indicates respondents who did not use any other means to recommend news)

While the difference to the results of Bernstein et al. [1] is striking— they found email to be the most common tool for online sharing—the importance of social media in sharing evident in our results concurs with some other studies (e.g. [6]). The reason behind the difference can be the different respondent profiles (inc. cultural factors) or that they studied sharing any web content while we focused solely on sharing news.

On average, the respondents who made recommendations used two means for it. Some means appeared to go better together than others. *IRC* and *Social media* use were clearly connected, as 65% of *IRC* users also used *Social media* and 58% of *Social media* users also used *IRC*. Similar reciprocities were found for *IRC* - *IM* and

IM - *Social media*. *Email*, on the other hand, was a natural companion to *Tell a friend* (69% of its users also used email) and to some degree, those who used *Other* (34% also used *Email*, the most popular second method for *Other* users). It appears that email is the core means for all those not in the previous cluster.

In effect, we see two clusters (circled in Figure 2): One for *ICR*, *IM*, and *Social media*, and one for *Email*, *Other*, and *Tell a friend*. To be sure, the borders are porous and the clusters are not clearly defined, but some formation is evident, nonetheless.

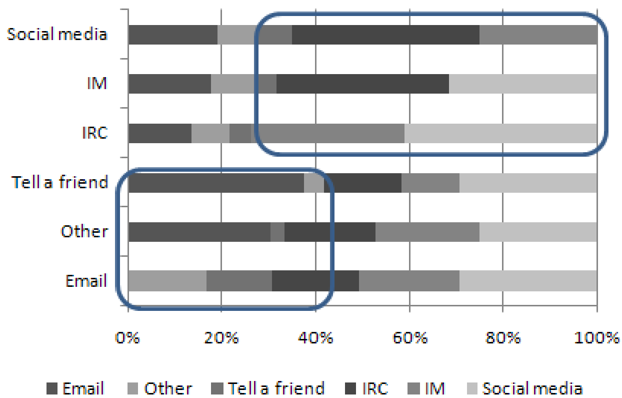


Fig. 2. What other means were used with each means

The same clusters can be seen in some other characteristics as well. For instance, TV as a medium for news did not appeal to those who used *Social media*, *IM*, and *IRC* to recommend news. These groups rated its importance below 4 (3.65–3.95), while those who used *Email*, *Tell a friend*, and *Other* means to recommend news rated it somewhat higher (4.41–4.66).

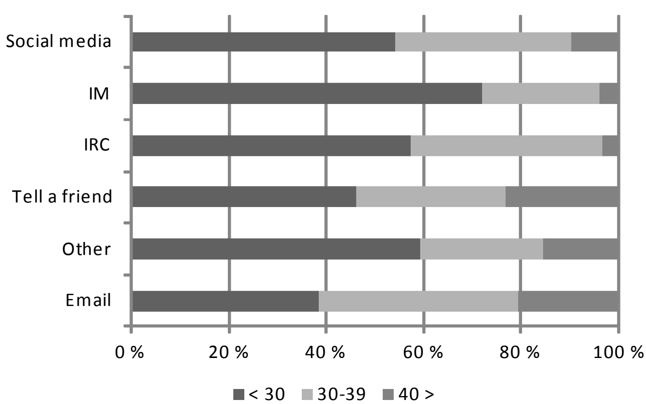


Fig. 3. Recommendation means used by age groups

We also analyzed the age distribution of users of each recommendation medium (Figure 3). Those using *Email* to recommend news users were on average slightly older: 61% were 30 or over while in the whole sample only 45% were 30 or over. In contrast, those using *IM* were chiefly in the younger age groups.

In *Other*, in addition to the 21 mentions of recommending news face-to-face, 6 recommended them over the phone and 1 mentioned using SMS. Bernstein et al. [1] also found face-to-face link sharing to be very common.

5.2 What Kinds of News Are Recommended

Figure 4 shows the most common types of news categories (those mentioned in 10 responses or more). The similarity in the sent and received news types is evident.

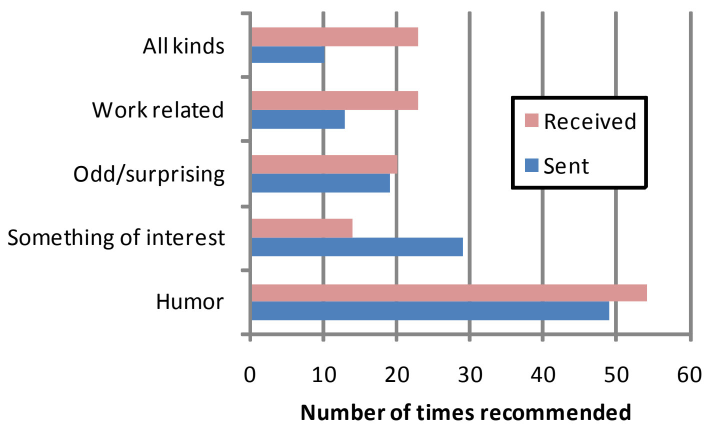


Fig. 4. Most common news types recommended

5.3 Why News Are Recommended

When the respondents were asked why they recommended news, 6 reasons were mentioned by more than 10 respondents: *Sharing funny things* (77 times), *Informing* (56 times), *Conversation* (24 times), *Participation* (15 times), *Feeling that the item would interest the receiver* (13 times), and *Staying in touch* (12 times).

In addition to *Conversation* and *Participation* being common reasons for recommending news, many reasons mentioned by fewer than 10 respondents were also related to maintaining social bonds: *Topic of common interest* (3 times), *Sharing an emotion* (3 times), *Giving a good feeling* (2 times), and so on.

With 24 respondents referring to starting or maintaining an ongoing conversation as a reason for recommending news, news recommendations are clearly part of overall social behavior. The conversations can take place face-to-face, virtually (computer-mediated), or partially virtually and partially face-to-face.

Furthermore, out of the 9 respondents who reported having stopped recommending news to somebody, 2 explained that it was because they were having less and less to do with the person. All this goes to underline how recommending news is part of overall maintaining of human relationships within our social network.

5.4 Framing the Recommendation: What to Include

The respondents saw being able to include one's comment (4.81/1.90) and being able to include a piece of the article (4.20/1.90) as quite important. Commenting helps avoid misunderstandings as to sender's opinion—"*That's where I say what I think about the article, I don't want somebody to think that I actually consider the item important, usually the opposite*"—and allows the receiver estimate if he or she should read the item: "*Comment could be used to emphasize what made the news article important in the first place or why it's sent to this person*".

5.5 Using Social Media to Recommend

Social media was the single biggest means of sending and receiving news recommendations. *Social media* here is largely synonymous with Facebook, as only it and Twitter were mentioned, and Twitter use was very low in comparison: 56 respondents recommended news in Facebook while only 4 did so in Twitter.

Many *Social media* users gave a description of how they recommended news in social media. These descriptions show that like with *Tell a friend*, tools provided by news sites to add news items to Facebook are not that popular—only 2 mentioned using them. One reason might be that other methods offer better ways to comment the recommendation: "*... Facebook—I post the interesting URL there and write as short a comment as possible*".

In fact, no method offered by Facebook for recommending was a clear winner. Respondents mentioned using Wall, Inbox (private messages), Feeds, Status, Share feature, Chat, and Groups for sharing news items. In effect, Facebook offers both means to recommend directly to persons and means to recommend to the world-at-large. As a result, *Social media* recommendations can be direct or indirect, i.e. personal or aimed at a larger audience. In addition, Facebook is also an IM environment: "*Facebook. That's where I talk with my homies the most as I am surfing and I drop them a hint if I read something funny...*" Thus, categorizing Facebook simply as a *Social media* is somewhat misleading—as a social media site, it actually covers many bases in online communication. Although Twitter has caught on slowly in Finland, it appears to be coming along, as more people receive (and would like to receive) recommendations from it than currently use it for making recommendations.

6 Which Recommendations Get Viewed and Why

6.1 How Respondents Acted on Recommendations

Overall, a significant number of recommendations get read (Table 4): 65% read *All* or *Most* news items recommended to them, 83% read at least *About half*, and only 4% claimed never to read them. Interestingly, the decision on whether or not to read an item appears to hinge significantly on the relation between the sender and the receiver of the recommendation: "*I don't get that many news recommendations but when I do,* *I read them (if they're from somebody I know well)*". In fact, the respondents rated the importance of the person making the recommendation to reading the recommended news item as 5.45, thus underlining the importance of the sender to reading.

Table 4. How many recommended items respondents read and expected others to read

	Read		**Expected others to read**	
Frequency	**No.**	%	**No.**	%
All	37	26 %	34	26 %
Most	57	39 %	62	48 %
About half	26	18 %	23	18 %
Less than half	19	13 %	7	5 %
None	6	4 %	3	2 %
No reply	*2*		*18*	
Altogether	**145**		**129**	

In addition, the medium through which the recommendation is received also plays a role. Figure 5 summarizes the effect of the medium through which the recommendation was received. If *Social media* has shone in other aspects, here it shines for all the wrong reason: Only *Other* had a lower reading rate. While 61% of the users who received recommendations through *Social media* read *All* or *Most* recommended items, the respondents who received recommendations through *Email*, *IM*, and *IRC* read over 70% of them, and the respondents who received recommendations through *Tell a friend* read 90% of them.

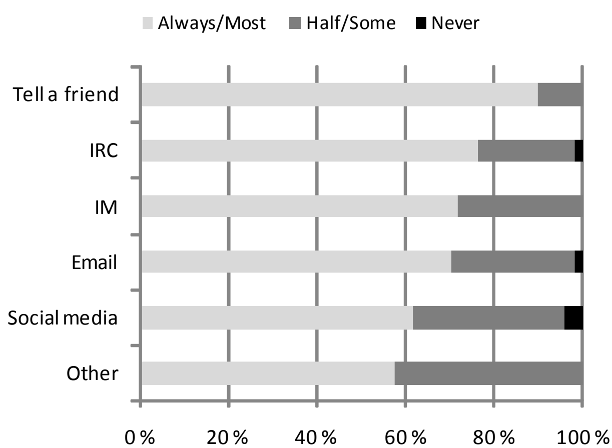


Fig. 5. What share of the recommended news gets read by the means of recommending

We also found that the perceived importance of *Social media* was lower than that of others for its users—only *Tell a friend* was perceived as less important by its users than *Social media*. Thus, although *Social media* was used a lot, other (less used) means were considered more important means of recommending, perhaps due to their directness. The importance of the medium shows even more pronouncedly in the respondent comments: "*The recommendations I get in Facebook are not as 'important' as the ones I receive in email. I check out the links in Facebook only if I got extra time. The links I get in email I check almost always, because they are directed to me and are thus more personal.*" Repeatedly, the message is the same—if it is directed personally to me, it is important, but if it is not directed at me personally, it is not as important. Facebook is mentioned repeatedly because of its popularity, but at the end of the day Facebook simply appears to represent a medium where the recommendations are not direct or personal but rather broadcast at the world-at-large. Email is considered more important because it is both direct and personal. Thus, how direct and personal a recommendation is determines to a large extent if it gets read, somewhat reducing the importance of *Social media* in news recommendations.

We conjecture from our data that when it comes to indirect, non-personal recommendations, different factors determine if the item gets read or not. First, the significance of the headline to an item getting read was rated very significant at 5.63 (1.33). The significance of the title is likely to increase when the recommendation is neither direct nor personal. Therefore, whether or not the receiver reads an article recommended in Facebook depends much more on the information available about the item itself, such as title, since its importance cannot be determined from the relationship (e.g. "Because of my relationship to this person I should check it out" and "This person often makes good/bad recommendations" type of factors).

The title is a good indication of the topic and theme of a news item, and resultantly a very important factor in deciding whether or not to check the item out. If the topic and theme are shown in the recommendation, the receiver can easily decide whether it is of interest. When it comes to news in particular, the source is also likely to be important as it gives hints of the treatment of the topic (e.g. Fox News vs. NYTimes). Therefore, when the person making the recommendation is not the deciding factor, the available information on the topic and source becomes a more significant factor.

When there are too many recommendations that are not salient to the receiver, they all tend to get treated as spam. For example, while some respondents received news recommendations from online services, at least some ignored them: "*I get some but I consider them spam and I don't read them…*" Consequently, in giving news recommendations, a large enough number of the recommendations need to interest the receiver or the recommendations are perceived as spam.

This applies not only to services but also to people who make personal recommendations, as seen in: "*Well, if it's from some jerk-off, then I won't bother*". As O'Donovan and Smyth [17] state, we also need to trust the quality of the recommendations a person makes, and the past quality of the recommendations the person has made appears to weigh in the equation. Also, the trustworthiness of the person making the recommendation may be context-specific.

6.2 How Respondents Expected Others to Act

The respondents expected the receivers to read their recommendations roughly as often as they reported reading the recommendations they received. The correlation is highly significant ($r = .4016$, $p < .0001$). Interestingly, the respondents rated the importance of reading recommendations sent to them (4.79) much higher than the importance of others reading their recommendations (3.58). In addition, the respondents rated the possibility of being able to comment received recommendations at 4.54 but getting a response from the receiver of their own recommendations only at 3.51: "*This is a difficult one, because [getting a response] is not that significant but … since we tend to talk and joke [about it], it is important in that sense.*" In effect, the respondents placed more responsibilities on themselves than others.

In summary, making recommendations between people who know each other and engage in many forms of social intercourse is different from recommending news to unknown people. When the sender and the receiver know each other, the sender considers the interests of the receiver, and news recommending takes place within the larger frame of social intercourse. On the other hand, when making recommendations to a faceless mass, for instance "digging" something, we are more involved in spreading the word. The content of the piece becomes important instead of the recommendation being aimed at building and maintaining relationships. It seems likely that the recommended items and topic are also different.

7 Recommendation Sites, Aggregation services, Feeds, and Recommendation Emails

While recommendation sites (e.g. Slashdot, Digg, and Reddit), aggregation services (e.g. Ampparit and Google News), feeds, and email recommendations from online services were used by some respondents, their share of the news recommending cake pales next to the human-to-human(s) recommending activity taking place. However, to their own users, they were important to finding news.

The 36 respondents who used aggregation sites and services (*typically* used by 30) rated their importance at 5.03. A few mentioned using filtering to get rid of Big Brother (a reality TV-show) and similar types of news or to get news only from some categories. Their comments underline that aggregation site users are happy with the services: "*An easy way to get a general view of especially the foreign news without any personal trouble*" and "*They are easy and pleasant to follow during the day…*"

Feeds were used by 34 respondents (*typically* used by 32) of whom 85% were males. The feed users rated their importance to news following at 5.28, therefore more important than aggregation sites or recommendation sites were to their users.

Recommendation sites, mainly Slashdot (15 mentions), Digg (5 mentions) and Reddit (3 mentions), were used by 19 respondents (89% males). Recommender site users rated their importance to following news at 4.47, therefore also fairly high.

The 19 users who received email recommendations from online services (*typically* used by 15) all used different services—no service was mentioned even twice. Some were connected to the user's work, some to hobbies, and some to more general interests, such as economic newspapers. What set the respondents who received such email recommendations apart from the overall sample was their age: 37% were 40 years old or older, while in the whole sample only 11% were 40 or older. Interestingly, the respondents who received email recommendations from online services did typically not pay that much attention to them. No respondent said anything positive about them but many mentioned not reading them: "*I don't remember. I ignore most of them.*", "*I get some but I consider them spam and I don't read them…*" and "*Much of the time I don't read them…*"

8 Who to "Follow" for News Recommendations

The respondents were also asked to rate the importance of following friends, editors/journalists, specialists (in a given field), unknown people who shared the same interests, and of getting automatically generated recommendations that took their interests into consideration. We asked them to rate them independent of whether they used such systems, as we wanted to probe their attitudes towards such features.

It turned out that the respondents were not that enthusiastic of any of these options. Getting automatic recommendations based on their interests was the most highly rated (3.06/1.85). The comments, while often mildly positive—"*Could be interesting…*", "*If such feature existed…I'd sure use it*"—also focused on potential problems and showed a decided lack of faith in automatic systems, e.g. "*I don't believe there could be an efficient automatic news recommending system for my interests*" and "*Would be an interesting feature if it worked, otherwise it would simply be irritating if it produced extraneous, uninteresting stuff.*" The lack of trust on automatic recommendations appears symptomatic, and this can be one significant reason behind their relative lack of popularity among our respondents.

Following friends or editors were not greeted with enthusiasm, either, at 2.29 and 2.20, respectively. Neither did following unknown people with shared interests at 2.48 fare much better. In fact, following specialists in a particular field was seen as more important than the other options at 2.77. We postulate that the reason for this is partially the control it offers over the type of news. In addition, the effect of the authority position such people may enjoy should also be studied, as our results do not shed light on this aspect of the equation.

Consequently, recommendations coming from friends were important and likely to be read (if direct and personal) but following friends was not considered important. We conjecture that the reason is that simply following friends means that the friends have not pre-selected news items that are likely to be of interest to us (quality). Also, following friends does not strengthen our social ties the way making and commenting recommendations does (part of larger social intercourse). Thus, if a news recommending system enables following, it should also provide features that enable sociality or improve the recommendation quality from a dull stream of what others are reading.

9 Implications of Current Practices

9.1 Quality vs. Sociality

There appear to be two dimensions working in news recommending practices: 1) getting good recommendations, and 2) recommending being part of the larger social intercourse between people. The two are, of course, not contradictory to each other, as the people who know us are in position to give us good recommendations. Social recommending serves both needs but emphasizes social aspects while automatic news recommendations, such as those based on collaborative filtering, tend to emphasize quality aspects and ignore social aspects.

Significantly, many recommendation providers have begun to add social aspects to their service, e.g. following friends in Digg. On the other hand, sociality-emphasizing recommending systems cannot ignore the quality issue, or they might end up being regarded as spam and therefore ignored, as has happened to recommendation emails. The reason why *Email* and *Tell a friend* recommendations get read is not only because they are from our social contacts (personal) and likely of high quality. They are also typically not too many. On the other hand, Facebook recommendations not made personally to us are judged on other merits than maintaining social connections, and so the title and other aspects that tell us about the news item itself become more important. With this type of recommendations we need to help the user receive only good recommendations and not be flooded with semi-targeted ones.

The challenge is to apply the two ingredients in right proportions to different services. Is the purpose of Digg to provide a social experience or help people find salient reading? If it is to provide good recommendations, then the focus should be on quality of recommendations and social aspects should be used to spice the experience but not allowed to come in the way of recommendations. In contrast, in Facebook news recommendations are a spice and a consequence of the larger social experience, something that nicely integrates into the social intercourse. Consequently, Facebook needs to give tools that make news recommending within its ecosystem an easy and enjoyable addition to the overall social experience and make sure that these tools do not distract users from the social experience.

Of course, these examples describe exactly what Digg and Facebook are doing. There are also many other approaches being developed and tested today, such as NYTimes's TimesPeople (based on following other users and NYTimes editors and journalists) and Huffington Post's integrations with Facebook (the integration means that the readers do not need to build their social networks once again in yet another service). The key is to know what one's focus is and not to try to be a be-all-end-all.

9.2 Design Implications

How to approach news recommending and the inherent recommendation quality vs. supporting social intercourse equation hinges very much on the service and its focus and its business model. News providers need to decide how and at what level they want to take part in existing social networks and recommending services. Judging by the number of Facebook and Digg icons popping up next to news items, the news providers do see some level of integration with these services as desirable.

Provide Item Title and Other Salient Information in the URL: Whether a news item is recommended directly to the reader or to the world-at-large, the news provider should strive to provide as much information about the news item as possible. A trusted person making a recommendation combined with an interesting item is a winning combination.

Although in the context of social network integration the news provider can to some degree affect the amount of information provided to the receiver, many news items are still recommended as URLs and whatever else accompanies the recommendation is beyond the provider's control. News providers should therefore strive to provide as much information about the item in the URL as feasible. In an URL like *http://www.somepaper.com/article/iU6040?type=mN?src=mv* all we see is the source (*somepaper*). In comparison, *http://www.somepaper.com/news/asia/2010-April-12/Is-the-West-Engaging-China.html* gives us a lot of information to whet our appetite.

Make it Personal and Direct: Personal and direct recommendations are likely to get read: If a friend who knows us recommends us personally something, it means that he or she has selected it for us and has a reason to believe that we should read the item and by reading it we can strengthen the relationship (we might discuss the article over lunch) and by not reading we might hurt the relationship (the sender asks us about the article or refers to it and realizes that we did not read it). Therefore, the recommendation is likely to be good and we stand to gain social capital (or lose it, perhaps even a stronger motivator).

Consequently, a social networking site—or any other site, for that matter—should make sure that there are tools for direct recommending, from one user to one or more users (that may provide features to facilitate commenting or discussing the item). In fact, this is what social networking is about in part. By providing such tools, we can make it more likely that users do news recommending within our service rather than switching to another service or email.

Recommending to the World-at-Large: On the other hand, social networking sites also need to allow recommending news to the world-at-large. While enabling such recommendations and providing social networking tools, such as a commenting feature, is a start, there are also other possibilities to enhance such recommendations.

Although following friends or other users was not rated that important in our study, the success of TimesPeople has shown that it does have its place even if it is not to everyone's liking. Giving the users who agree to be followed a possibility of selecting what they recommend (instead of a stream of everything they view) has two advantages. First, it allows the users being followed to show explicit liking—not everything they view is to their liking—and consequently improves the quality of recommendations and keeps their number down, and second, it provides privacy protection that encourages users to allow the following to begin with.

To make it easier for users to follow recommendation streams, the number of items has to be sensible and they have to have a high enough occurrence of salient items not to be seen as spam. One approach is to support filtering the streams (also discussed by Bernstein et al. [1]). We found that some feed and aggregation site users had a tendency to filter recommendations explicitly to get certain types of news and to avoid other types. While these means were not used by the majority, their users were quite happy with them, suggesting that services should support such means. Another approach is to automatically filter the recommendation stream to reduce the number of items and improve the overall quality of recommendations. Social networks offer an especially good foundation for filtering with the ready networks of friends in place.

10 Conclusions

While our participants found most news by going directly to the news sites they knew and liked, recommendations from others also played a significant role in how they located salient news items. In effect, news recommending is part of the normal social intercourse where we keep in touch with people and share emotions and information.

The more directly the news recommendation is made, the more likely we are to read the item. For less direct recommendations, we make the decision more based on what we know about the item, i.e., we try to judge if the item interests us. Consequently, the sites involved in news recommending have to decide whether their focus is on sociality (with news recommendations being only a part of the larger social intercourse taking place between people) or quality recommendations (with social aspects providing spice to the experience). The question is to understand one's focus and to blend the right mixture for it. Understanding the larger social intercourse taking place between people helps us understand better how to integrate recommending news and other online information into that intercourse.

If we do not provide means for recommending with sociality taken into consideration, humans will find a way to add sociality to the experience, for example by emailing the URLs to each other. Email was not built with news recommending in mind but it continues to serve it well, in part because it opens way for all kinds of related social discourse. If we want to take things beyond this, we need to design for supporting human behavior. One of the secrets behind Facebook is that it supports well human sociality and consequently offers support for what we do all the time in any case—deal with other humans for fun and profit.

Acknowledgments. This work was supported by the ICT SHOK project on Flexible Services. The second author was supported by the Academy of Finland grant 1130044.

References

1. Bernstein, M.S., Marcus, A., Karger, D.R., Miller, R.C.: Enhancing Directed Content Sharing on the Web. In: Proc. CHI 2010, pp. 971–980. ACM, New York (2010)
2. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and Tweet: Experiments on Recommending Content from Information Streams. In: Proc. CHI 2010, pp. 1185–1194. ACM, New York (2010)
3. Ellison, N., Steinfield, C., Lampe, C.: The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. JCMC 12(4), 1143–1168 (2007)
4. Erdelez, S., Rioux, K.: Sharing Tools on Newspaper Web Sites: An Exploratory Study. Online Information Review 24(3), 218–228 (2000)
5. Goldstein, N., Martin, S., Cialdini, R.: Yes! Profile Books, London (2007)
6. Gupte, M., Hajiaghayi, M.T., Han, L., Iftode, L., Shankar, P., Ursu, R.M.: News Posting by Strategic Users in a Social Network. In: Leonardi, S. (ed.) WINE 2009. LNCS, vol. 5929, pp. 632–639. Springer, Heidelberg (2009)
7. Gürsel, A., Sen, S.: Producing Timely Recommendations from Social Networks through Targeted Search. In: Proc. AAMAS 2009, pp. 805–812 (2009)
8. Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S., Ofek-Koifman, S.: Personalized Recommendation of Social Software Items Based on Social Relations. In: Proc. RecSys 2009, pp. 53–60. ACM, New York (2009)
9. Hopkins, H.: Twitter and News and Media Websites. Hitwise Intelligence (2010)
10. Java, A., Song, X., Finin, T., Tseng, B.: Why We Twitter: Understanding Microblogging Usage and Communities. In: Proc. WebKDD/SNA-KDD 2007, pp. 56–65. ACM, New York (2007)

11. Kalogeraki, S.: The Divergence Hypothesis in Modernization Theory Across Three European Countries: the UK, Sweden and Greece. Culture Unbound 1, 161–178 (2009)
12. Kirkpatrick, M.: Facebook Could Become World's Leading News Reader (Sorry Google). ReadWriteWeb (2010)
13. Lerman, K., Galstyan, A.: Analysis of Social Voting Patterns on Digg. In: Proc. WOSP 2008, pp. 7–12. ACM, New York (2008)
14. Lerman, K., Ghosh, R.: Information Contagion: An Empirical Study of The Spread of News on Digg and Twitter Social Networks. In: Proc. ICWSM 2010 (2010)
15. Liu, J., Dolan, P., Pedersen, E.R.: Personalized News Recommendation Based on Click Behavior. In: Proc. IUI 2010, pp. 31–40. ACM, New York (2010)
16. Liu, F.K., Lee, H.J.: Use of Social Network Information to Enhance Collaborative Filtering Performance. Expert Systems with Applications 37, 4772–4778 (2010)
17. O'Donovan, J., Smyth, B.: Trust in Recommender Systems. In: Proc. IUI 2005, pp. 167–174. ACM, New York (2005)
18. Phelan, O., McCarthy, K., Smyth, B.: Using Twitter to Recommend Real-time Topical News. In: Proc. RecSys 2009, pp. 385–388. ACM, New York (2009)
19. Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., Olmstead, K.: Understanding the Participatory News Consumer. Pew Internet & American Life Project (2010)

Helping Users Sort Faster with Adaptive Machine Learning Recommendations

Steven M. Drucker, Danyel Fisher, and Sumit Basu

Microsoft Research, 1 Microsoft Way, Redmond, WA. 98052
{sdrucker,danyelf,sumitb}@microsoft.com

Abstract. Sorting and clustering large numbers of documents can be an overwhelming task: manual solutions tend to be slow, while machine learning systems often present results that don't align well with users' intents. We created and evaluated a system for helping users sort large numbers of documents into clusters. iCluster has the capability to recommend new items for existing clusters and appropriate clusters for items. The recommendations are based on a learning model that adapts over time – as the user adds more items to a cluster, the system's model improves and the recommendations become more relevant. Thirty-two subjects used iCluster to sort hundreds of data items both with and without recommendations; we found that recommendations allow users to sort items more rapidly. A pool of 161 raters then assessed the quality of the resulting clusters, finding that clusters generated with recommendations were of statistically indistinguishable quality. Both the manual and assisted methods were substantially better than a fully automatic method.

Keywords: Mixed initiative interactions, adaptive user interfaces, information interfaces, interactive clustering, machine learning.

1 Introduction

A project manager at a software corporation is faced with tens of thousands of pieces of feedback from beta users. She wishes to divide the feedback into groups to ensure that each item will be addressed. Going in, she does not know what the appropriate categories for the data will be. The categorization she creates is organic and personal – she splits up the work with her team in mind, creating appropriate categories according to their abilities. A different project manager with a different team would surely create a different clustering for the same task. Existing tools for this type of personalized sorting are limited in their ability to help a user with the task – today, the program manager would likely use a tool such as Excel, but would have to do the categorization manually.

This need exists across many domains. A graduate student organizing related work for his thesis, or a human resources manager organizing resumes, faces a similar challenge: they have very large amounts of data to sort, and their particular choices are unique to their own needs and interests. Indeed, a different graduate student, or different manager, would come up with a distinct clustering.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 187–203, 2011.

1.1 Using Machine Learning to Speed Clustering

Numerous algorithms are available to help automatically cluster documents. These algorithms make implicit or explicit assumptions about what items should be grouped together, which may not be well aligned with a user's own notion of appropriate groupings. Furthermore, users may have different notions of what items should go together, and may not be able to explain why, at a high level, certain items should or should not go together. We wish to take advantage of the assistive power of machine learning, while allowing users to express their needs naturally.

Methods of combining machine learning with manual clustering have met with mixed results. Both the Implicit Queries for Data Mountain [8] and Scatter-Gather [1] research found that machine learning made sorting and retrieval tasks, respectively, *slower* for users, though they did help users to recall the locations of items and better understand the target corpus.

We differ from these past approaches. Starting with machine-generated clusters can have two undesirable outcomes. Primarily, it can diminish trust in the system if automatic clusters do not correspond to the users' schematization. Additionally, we are concerned that automatically-generated clusters might bias users, guiding them away from creating clusters that are meaningful to their interests. While fully automated clustering can be very effective for datasets in which cluster centers are well-separated (such as disambiguating search results as in [18]), we show that automatic clusters are less successful for ambiguous datasets. We want to allow users to grow clusters organically, yet still gain the benefits of machine assistance.

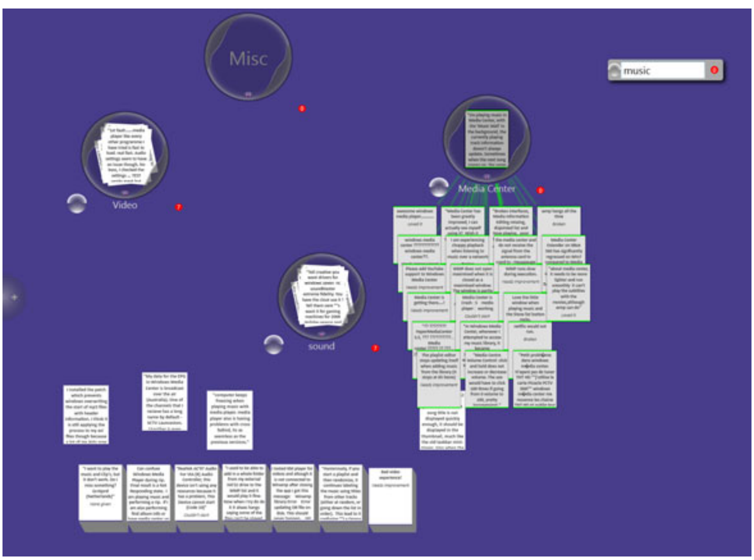


Fig. 1. The iCluster interface. The user has asked for suggestions of documents that might fit the "Media Center" cluster.

1.2 The iCluster System

Our system, called iCluster, assists users in organizing a set of items after observing their actions; it learns a model of each individual cluster as well as a global distance metric. The system makes recommendations to the user; these recommendations improve as the user organizes more data. As the system is training on the work users have done so far, users' individual clustering habits will lead to personalized recommendations.

The primary contributions of our work are an adaptive, interactive system to assist people in sorting items, and an evaluation of how recommendations affect the sorting task. We explored time and accuracy tradeoffs using iCluster in comparison with both a manual system and fully automated clustering.

The body of this paper is organized as follows. We first examine our work in the context of systems for organizing documents spatially, past work on clustering in the machine learning literature, and systems that use interactive learning to search and organize information. We then describe the design of the system, including an observational study and subsequent iterative development as well as the backend algorithmic support. Next, we discuss the evaluation of the system; we find that assisted clustering can be substantially faster than unassisted clustering while maintaining the quality of the clusters. Last, we discuss our findings and future work.

2 Related Work

2.1 Manual Clustering

A number of research projects have examined the ways that people sort their data, both looking at personal data and shared information. Dourish *et al* [10] examine shared files in a government office. While the government workers nominally sort items with a defined, global taxonomy, the sorting and clustering process is based on local needs and information. Malone [12] examines office workers' organization of their desks; Whittaker and Hirschberg [17] look at how people manage personal paper archives. Both studies find that users use local, idiosyncratic organizations, moving things into categories that make sense for their own tasks. Similarly, they find the now classic breakdown between 'filers' and 'pilers', and that people very naturally arranged items in piles. While these piles reflect the users' interests, Whittaker and Hirschberg also show that it can be difficult to find papers once they have been piled, or to get an overview of what a pile represents.

2.2 Techniques and Interfaces for Adaptive Clustering

The machine learning community has a history of work on unsupervised (i.e., fully automatic) clustering of data by various means, from the well-known k-means algorithm to the more modern approach of Latent Dirichlet Allocation [5]. More recently, there has also been some work on "interactive clustering," in which an algorithm is presented with pairs of items that *must* or *cannot* belong together. The algorithm then optimizes a distance metric which best satisfies these constraints, and items are clustered using k-means (e.g. [7]). However, the work in the machine learning community has been theoretical in nature; there has been very limited work on incorporating interactive clustering into an interactive *system*, and none measuring its effect on human behavior. The closest such work is that of desJardin et al. [9], which proposes an interface in which users could specify these constraints. This work contained only a simulation of human behavior and did not study how or whether the system helped humans with the task.

Basu *et al.* [3] propose an algorithm to support online interactive clustering that works somewhat differently: instead of reclustering the entire dataset, their method provides online recommendations for new items that could be appropriate for a given cluster. That work was not incorporated into a final system nor evaluated with real users. In this work, we adapt Basu *et al.*'s technique with a spatial interface.

2.3 Spatial Layout for Document Collections

There has been quite a bit of research into how people use spatial organization in real life to help manage their collections of documents. Malone [12] explores how people organize their office desks, and Whittaker and Hirschberg [17] look at how people manage personal paper archives. They find the now classic breakdown between 'filers' and 'pilers', and that people very naturally arranged items in piles. However, they show that it can be difficult to find papers once they have been piled, or to get an overview of what a pile represents. Our interface uses a spatial layout to build on these observations while trying to alleviate the difficulties.

User interfaces have been exploring spatial organization since the early 1980's. Work at Apple expanded on simple 2D spatial layouts with piles [13] that allowed users to group related documents into stacks. More recently, DataMountain [15], Bumptop [1] and DynaPad [4] all expand from the simple 2D metaphor by adding 3D, physics, or zooming. Our contributions are orthogonal to this work – we could augment any of these systems with the techniques described in this paper. Other interfaces allow users to cluster items spatially: for instance, Bubble Clusters [16] make clustering an implicit operation based on spatial distance between items. While our approach has a more explicit containment model, our interface could be adapted to work with the bubble-clustering approach.

Czerwinski et al. [8] adds "implicit queries" to the DataMountain system. DataMountain [14] allows users to spatially layout web thumbnails in a pseudo-3D plane; their project extends the work by highlighting related documents. Related documents are computed with a fixed similarity metric (unlike our dynamic learning system), based on either a previous subject's categorizations, or a cosine distance between document word vectors. The authors evaluate categorization time, number of categories and subsequent re-retrieval; they found that users using implicit queries took longer to place items but were more able to recall where they were.

Scatter/Gather [15] uses a search interface to build clusters of related objects; users can then dynamically recluster based on a selection. Scatter/Gather uses this for interactive refinement; it was designed for question-answering tasks, rather than information sorting tasks. As such, it was used and evaluated in a very different manner than our system.

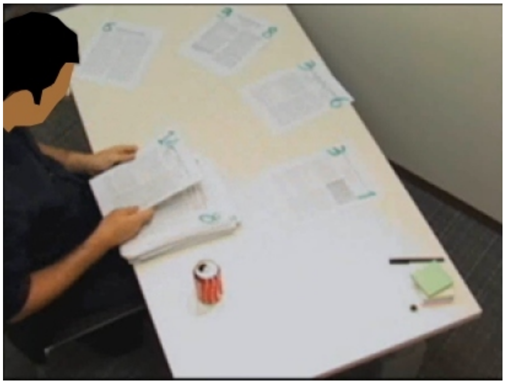


Fig. 2. A user clusters 60 CSCW papers by hand during the observational study

3 The iCluster System

We developed iCluster to address the challenges of clustering large sets of data. Our design was based first on an observational study of clustering behavior. In this section, we discuss first the observational study, then the prototype we constructed.

3.1 Clustering Paper Documents Manually

In order to better understand how people currently go about the task of clustering a large number of papers, we observed six people who were given the task of organizing 60 papers from the last few years of the ACM CSCW conference. The subjects were given one hour to get through as many documents as possible. Subjects were asked to group the papers according to their topics and addressed problems. They were given print-outs of the papers, sticky notes, paper-clips, pens, highlighters, and folders; users were permitted to organize the papers in any way they wished (Figure 2).

Consistent with Malone *et al.*'s results [12], all of the participants organized their documents into piles. Subjects spread their piles on the table, and labeled them with sticky notes. Some people put papers that they were not yet certain of into a temporary "holding" pile, while others left those papers in the initial pile. Several users refactored their piles when a new paper made them realize that they had mis-categorized a past paper.

In general, the lessons learned from this preliminary study were the following:

- The desktop was used as a holding area; papers were then piled in distinct stacks. People used spatial memory to re-find piles, although this could be slow and error-prone.
- Labels on piles helped users return to them more easily; unlabeled piles were confusing.
- People often did not put items in a cluster until several related items were found.
- People seemed to use different organizational principles to cluster the information (agreeing with Dourish *et al.*'s findings [10]). For instance, some grouped methodological techniques, while others grouped based on a particular technology.

3.2 iCluster Prototype

We used the feedback from the observational study in order to build a prototype for users to cluster papers online. The iCluster prototype uses spatial design concepts: documents and clusters are arranged on the desktop, and can be dragged into new locations. Documents can be dragged into clusters; users can request recommendations of what documents might fit into a given cluster. Figure 1 shows the interface. The prototype can be seen in action in the supplemental video figure.

iCluster represents each document as a small “trading card” (Figure 3). For instance, in a corpus of textual content, we include the first few lines of the text on the front, and the entire statement on the back. For conference papers, we put title and authors on the front and the abstract on the back. Initially, the cards are stacked in piles (seen at bottom, Figure 1); users place the cards into clusters.

Fig. 3. The front (left) and back (right) of a single document in the iCluster interface

Fig. 4. Full-text search. The user has searched for the word “collaboration,” matching documents are brought to the foreground

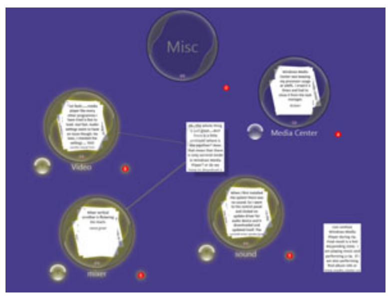


Fig. 5. Item-to-group recommendation. The recommendation system believes that this document would fit better in the “mixer” group (lower left, thick line) than the “video” or “sound” group (thinner lines to left and bottom right), and not at all to the media center group

We experimented with various forms of loose spatial layout (such as DataMountain and bubble clusters), but found that our users liked the ability to hide information in a container. Users create clusters in iCluster with a button on the left side of the screen which creates an empty cluster that can be moved around the screen. Users can caption clusters with a click and by entering a name on the keyboard.

A text entry box on the screen enables keyword search. When a keyword is entered, up to 70 matching items are spread out underneath the search query box in columns which allows for effective use of space for the revealed documents. The items are highlighted, and lines are drawn to the search box, to make the items easier to find. Figure 4 illustrates a text search in progress. Based on user feedback, we also highlight any groups that contain items that match the search.

The novel contributions of iCluster are the ways in which the system can help users to make clustering decisions. Assistance is based on items within clusters, and so does not begin until users have placed items into clusters. Recommendations can be triggered in two ways. When an item is selected, lines of different weights are drawn to clusters, showing which ones might be most closely related to the selected item. A thicker line indicates a more likely match. We refer to this as an item-to-group recommendation. Figure 5 shows one such recommendation.

The second form of recommendation is the group-to-item recommendation. When a user clicks on a cluster, the system computes which twenty documents would be the best match for the cluster, based on entries already placed in clusters. It then lays out those documents using the same column strategy as before, underneath the selected group. Group-to-item recommendation is illustrated in Figure 1. As with search results, there is no “accept” button for group-to-item recommendations; rather, the user simply drags the object they are interested in into the relevant group. Indeed, a user may even choose to drag a recommended item into a different group.

We now describe the learning algorithm that drives both the item-to-group and group-to-item recommendations.

3.3 Backend Learning Algorithm

The recommendation algorithm is based on Basu *et al.* [3]; we provide an overview here for completeness. The algorithm is a hybrid approach that incorporates aspects of both distance metric learning and per-cluster classification. The algorithm dynamically learns both types of models and reweights their relative contributions. In so doing, it can leverage both approaches in modeling the users’ intent. The algorithm was shown to make good recommendations after fairly few items were rated, and can be applied to arbitrary types of data, as long as each item can be described by a set of candidate features (continuous or categorical) and the distances between pairs of items can be described by a set of candidate distance measures.

Given a set of features and distances, and the list of items which belong to each cluster, the algorithm optimizes a distance metric between items that best satisfies these constraints. Unlike previous work (such as the fomulations in Cohn and Caruana [7] or desJardins et al.[9]) that only optimized the weights on the candidate distance measures, this method also trains a classifier for each cluster and converts it into a new distance measure; details of the formulation are in [3]. In this way, this method is able to take advantage of the benefits of both metric learners and classifiers. To recommend the closest items to a cluster, the algorithm computes the mean distance using this learned measure to all the elements in the cluster. The items are then ranked according to this quantity.

We extended Basu *et al.*’s algorithm to get recommendations of the closest cluster to an item. To do so, we compute the rank of the item for each cluster, then sort the clusters based on this rank. We use rank rather than distance because distances from different clusters are not calibrated with respect to each other.

For our data, the per-item features were produced by combining all the text from the text fields and representing this via standard TFIDF features, which were further compressed via principal components analysis (PCA) to yield the top 50 features per item. The 25 distance measures between items came from computing the absolute distances between the top 25 features. Because we had some concerns about the model overfitting when only a small number of items had been labeled, we made the small extension of scaling the number of features according to the data available. Specifically, we limited the number of features available to the classifier to five times the total number of clustered items so far; we found this led to better recommendation behavior for the first few filed items.

Basu *et al.* [3] validated their algorithm's recommendations through simulation: five expert humans manually clustered 300 papers from the SIGIR, CSCW, or SIGGRAPH conferences, depending on their area of expertise. The authors then simulated the recommendations that the system would have generated based on the data input, playing back the clusters that the users formed in random order. With only three examples as labels, more than 25% of the recommended items were correct suggestions on average; with eight examples, nearly 40% were correct. Furthermore, the algorithm suggested items at a significantly higher rate of precision than either a classifier or a metric learning approach on its own.

This relatively high level of precision even with small numbers of examples led us to believe that the algorithm would be an appropriate choice for the engine of the iCluster system during the sorting task.

4 System Validation

In order to explore both the need for and the effects of using iCluster, we performed several different evaluations. First, we wanted to establish that people do indeed cluster items in different ways from each other. Second, we wanted to investigate whether using iCluster would allow users to generate clusters more rapidly. This speed should not be at the cost of quality: the clusters they generate should be of similar quality to ones generated without machine assistance; and that the clusters should be of higher quality than purely-automatic clusters.

In order to compare the effects of recommendations, we created a version of iCluster that does not contain either item-to-group recommendations or group-to-item recommendations. We refer to this limited version as BASELINE, and to the full version as ASSIST.

We also wanted to make sure that the iCluster system could be generalized to work for several different problem domains.

We explore the following four hypotheses:

- **H1:** Individual users will develop different and distinct clusters.
- **H2:** Users will cluster more items in the ASSIST condition than in BASELINE.
- **H3:** Clusters created in the ASSIST condition be of comparable quality to those generated in the BASELINE condition.

- **H4:** Both ASSIST and BASELINE clusters will be of higher quality than automatically-generated clusters.

We tested these hypotheses through several studies. The first study did not use any assistance whatsoever, but rather had people use the base system to cluster several corpora and we subsequently examined the overlap of items to establish that people do indeed use different strategies to cluster documents. This was used to evaluate H1.

The second study was a within subjects study to explore how people perform with and without assistance from the iCluster system. It allows us to address H2.

Finally, a third study was performed to evaluate the quality of the clusters generated in the previous studies (and to an automated clustering process) to address hypotheses H3 and H4.

Study 1: Generating Clusters Baseline

For our first study, we recruited four sorters who were familiar with the SIGCHI literature, and another four sorters who were generally familiar with computing. CHI sorters were given a dataset based on a selection of 400 recent CHI and UIST papers. The other sorters were assigned to the FBK condition, which consisted of a dataset of 300 items of free-text feedback about a piece of software. In both cases, papers were stacked in 5 initial piles in random order. SIGCHI papers had the title and authors on the front, and the abstract on the back, and so averaged 138 words of content. The FBK dataset, in contrast, had no abstract; messages averaged just 31 words of content. Users started with no clusters.

We asked CHI sorters to imagine that they were on the program committee of a conference and needed to sort the different documents into separate tracks. FBK sorters were to sort the feedback items into functional groups that could be addressed by hypothetical product teams. They were asked to get through as many items as possible in 45 minutes while still creating useful groups.

These sorters were all assigned to the BASELINE condition of the program where no assistance was given. Sorters were able to use text-search to find documents by keyword.

After being introduced to the system and allowed to try it out on briefly on a test data set, sorters were given 45 minutes to cluster as many documents as they could. We logged all interactions with the system as they clustered in that time period, including the number of items sorted.

Study 2: Within Subjects Generating Clusters with and without Assist

Because of large variations in individual style and familiarity with the corpus, it was clear that a within subjects study was the only way to allow us to meaningfully compare sorting time or quality (H2 and H3). In addition, our system was vulnerable to unusually strong learning effects: once a user has clustered any sample from a corpus, they have a good intuition for the clusters that they will use for another sample.

We recruited 16 new sorters; we used as our corpora software feedback in two comparable but different areas (Network and Multimedia). We balanced the order of presentation (with and without assistance) as well as the order of the corpus. Sorters were introduced to the features of the interface between trials, and given a chance to try the interface. They were then given an unspecified amount of time to cluster as many items as they could into 'high quality' groups. After 20 minutes, they were interrupted and the task was restarted on a different corpus using the other condition of the interface.

Study 3: Evaluating Cluster Quality

Sorters could easily bias the cluster speed test, of course, merely by dropping items into clusters randomly. To address H3 and H4, we wanted to establish whether the clusters generated groups of high quality.

Judging cluster quality in an objective way has traditionally been quite difficult. A rater, presented with a grouping and asked for an assessment of quality, will search for some rational explanation of how the items fit together. Recent work by Chang *et al.* [6] presented a clever new way to judge cluster quality that alleviates this problem. The basic idea was to present some items from a cluster and one outlier; the accuracy with which raters could correctly detect the outlier was their measure of quality.

For our work, we adapt and expand this model by measuring the quality of a clustering with two properties:

Coherence: A cluster is coherent if a user can recognize that its members go together (and so identify an outlier).

Distinctness: A set of clusters is distinct if a user can recognize that elements from different clusters do *not* go together (and so pick out a matching element for a set).

In an Outlier test, adapted from [6], raters are presented with four items from one cluster, and one from a second, both chosen from a single sorter. The rater must identify the outlier that was chosen from the second cluster. If raters are unable to identify the outlier, then the cluster is likely to be a weak one; if raters are successful across many trials, the cluster is more likely to be coherent.

We augmented Chang *et al.*'s work with Match questions, which are meant to measure distinctness. We prompt the rater with two items chosen from one cluster. The rater selects one matching from the same cluster from among four distractor items chosen from other clusters (Figure 7). If raters can tell which item matches, then the clusters are more likely to be distinct from each other.

We created and deployed a web-based study within our organization to test the quality of the clusters generated in the first user study. Questions were selected randomly from among the ASSIST, BASELINE, or AUTO conditions (described below). After a training round, raters were given a series of 21 outlier and 21 match questions in randomized order.

Automated Condition. A third condition was added to explore how well a fully-automated clustering algorithm could perform in comparison to human-sorted clusters. A brief note on our automatic clustering approach. We use Latent Dirichlet Allocation (LDA) instead of k-means since the task presented to users was to determine the underlying topics of the collection, assigning appropriate documents to each cluster according to how well they belong: this is precisely what LDA was designed to do [4]. LDA computes the underlying topics and estimates for each document the "responsibility" of each topic in explaining the information in that

Choose the item that does not belong.

- HeartBeat: an outdoor pervasive game for children; *Remco Magielse, Panos Markopoulos*
- A comparative study of speech and dialed input voice interfaces in rural India; *Neil Patel, Sheetal Agarwal, Nitendra Rajput, Amit Nanavati, Paresh Dave, Tapan Parikh.*
- Design influence on social play in distributed exertion games; *Florian Mueller, Martin Gibbs, Frank Vetere*
- O' game, can you feel my frustration?: Improving user's gaming experience via stresscam. *Chang Yun, Dvijesh Shastri, Ioannic Pavlidis, Zhigang Deng*
- Designing digital games for rural children: a study of traditional village games in India. *Matthew Kam, Akhil Mathur, Anuj Kumar, John Canny*

Fig. 6. A sample "Outlier" question. The second item in the list is different from the others.

Choose the item that goes with both of these

- Collaborative interaction with volumetric display;
 Tovi Grossman, Ravin Balakrishnan
- MightyTrace: multiuser tracking technology on lc-displays; Ramon Hofer, Patrick Kaplan, Andreas Kunz

From among these

- Bonfire: a nomadic system for hybrid laptop-tabletop interaction; Shaun Kane, Daniel Avrahami, Jacob Wobbrock, Beverly Harrison, Adam Rea, Matthai Philipose, Anthony Lamarca
- Codex: a dual screen tablet computer; Ken Hinckley, Morgan Dixon, Raman Sarin, Francois Guimbretiere, Ravin Balakrishnan.
- Stirring up experience through movement in game play: effects on engagement and social behavior; *Sin Lindley, James Le Couteur, Nadia Berthouze*
- More than face-to-face: empathy effects of video framing; *David T. Nguyen, John Canny*

Fig. 7. A sample "Match" question. The second item matches the first two.

document; we interpret the topics as clusters and use the most responsible topic for a given document as its assigned cluster. The top documents belonging to each topic are those that had the highest value for this entry; those with the lowest values were left uncategorized. To best align with the clusters created by our subjects, we set the number of topics to be the average number of clusters from the subjects. We left uncategorized as many documents as the average number of uncategorized items from our subjects. We refer to this as the AUTO clustering condition. To produce four clusterings from this condition, we randomized the order of the items four times and ran LDA separately on each set.

5 Results

In this section we first evaluate the hypotheses and then delve deeper into a single user's results to better understand the interaction of search and recommendations.

H1: Individual Users will Develop Different and Distinct Clusters

We base our analysis for this hypothesis on the first user study. We wished to develop a measure of cluster similarity: that is, given two users, how similar are their clusters?

We pairwise compared every user's clustering to that of every other user. To measure the degree of agreement between two users, we took all pairs of items that were in the same cluster for the first user (where both those items were also filed by the second user); we then computed the fraction of those pairs which were in the same cluster for the second user. The agreement measure thus is between 0 and 1.

If the natural clusters of a dataset are the primary driver of users' clusters, then we would expect different users to create similar clusters to each other (with an agreement of near 1.0); conversely, if users' individual styles or biases have a strong effect, we would expect substantial disagreement. Users had a mean overlap of 0.49 (standard deviation 0.15) of their items. This low overlap shows that clusters were very different from each other. We consider H1 to be verified: different users created distinct clusters from the same items in a dataset.

H2: Users will Cluster More Items in the ASSIST Condition than in BASELINE

For each user, we counted the number of items that they successfully placed into a cluster. Figure 8 shows the number of items filed under each condition; Figure 9 shows the relative speedup between BASELINE and ASSIST. Both illustrate that ASSIST made a significant difference to filing speed, offering as much as a doubling in speed.

We used Wilcoxon signed-rank test to compare the number of items filed in the ASSIST and BASELINE conditions. There was a significant difference between ASSIST and BASELINE, z=-1.966, p=.049.

Figure 9 breaks down the *relative* speedup per user—that is, the number of items sorted in ASSIST divided by BASELINE. 12 of the 16 users were faster with assistance; 5 were 30% or more faster than in BASELINE. (Later, we look at the use of suggestions for one individual's run, which suggests that greater gains are possible.)

We consider H2 to be verified: users clustered more items in ASSIST than in BASELINE.

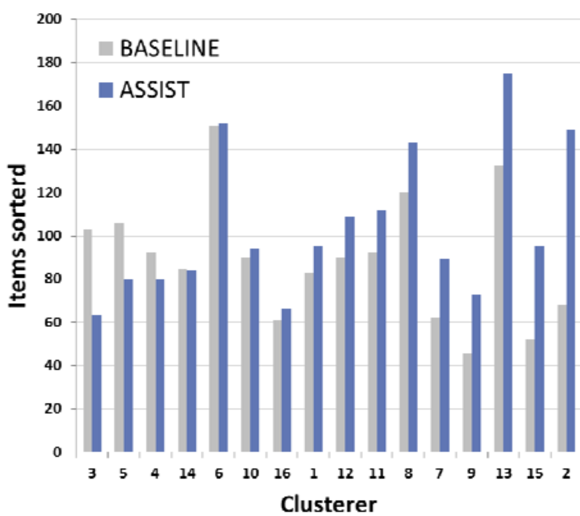


Fig. 8. Number of items filed under ASSIST and BASELINE for the second (within subjects) study. Subjects are sorted by absolute difference between their ASSIST and BASELINE runs.

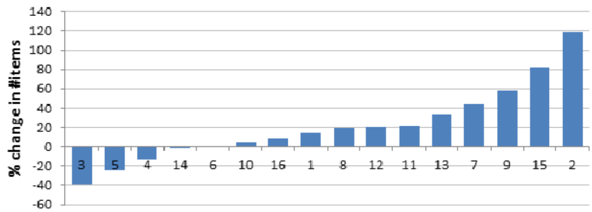


Fig. 9. Relative (percentage) speedup from the BASELINE to the ASSIST condition. One user was more than twice as fast with ASSIST.

H3: ASSIST Clusters will be of Comparable Quality to BASELINE Clusters

Our definition of quality is based on multiple raters' evaluations of clusters. Each rater answered a sample of Match and Outlier questions about the different sorters' clusters; the quality of a given sorting of a dataset is then the percentage of questions for that sorter that were judged correctly. As in H2, we base our result on the within-subjects study.

Some sorters produced clusters that the raters found easy to interpret; others produced clusters that were more difficult to understand. A set of clusters from a single user is scored as the percentage of questions (as described earlier) correctly answered by raters, and so ranges from 0.0 to 1.0. As raters selected from five possible options for Outlier questions and four options for Match questions, we place chance at 0.225.

A Wilcoxon Ranked Samples test was conducted to compare average quality of clustering in ASSIST and BASELINE conditions. There was no significant difference between ASSIST and BASELINE; z=-1.221, p =0.222.

H3 was verified: users with assistance created clusters not distinguishably worse from users without assistance.

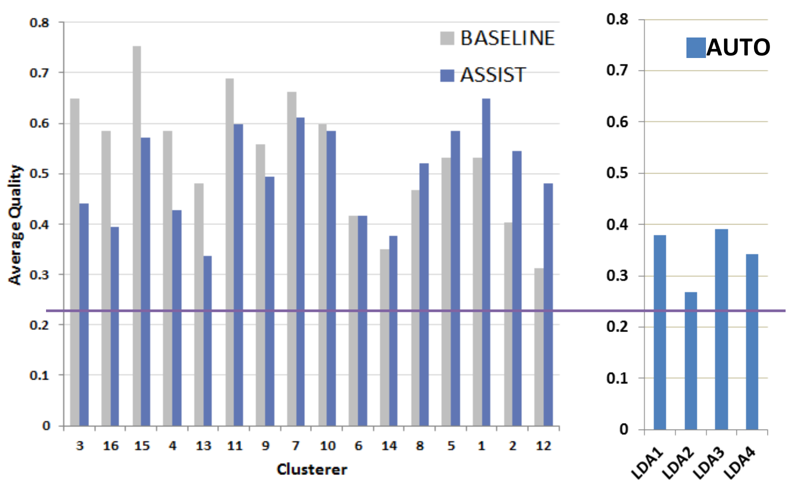


Fig. 10. Support for H3 and H4. (left): Relative quality of ASSIST, and BASELINE. Bar height is the average quality for a given user. The bars are ordered by absolute difference in quality. Chance is 0.225 shown by the purple line. Figure 10 (right): Quality measurements of AUTO (4 different runs of an LDA clustering algorithm).

H4: Both ASSIST and BASELINE Clusters will be of Higher Quality than Automatically-Generated clusters

As an additional condition for the quality evaluation study, we included clusters that were automatically generated to compare with both the passive and baseline conditions. We conducted a 3×1 ANOVA on mean rater scores and found $F(2,21) = 15.56$, $p < 7.17e-5$. Unlike the results in H3, this test is **not** paired samples, so it is not surprising that we have weaker p-values for the ASSIST-BASELINE condition. Figure 11 illustrates the three conditions; Table 4 shows the p values for the tests.

We consider H4 to be verified: Both BASELINE and ASSIST were significantly different than AUTO, and not from each other.

Table 4. (left) Mean and standard deviation for quality of within-subjects groups. (right) Significance values for ANOVA tests between conditions. Starred entries signify significant differences (i.e., the two conditions' scores are different from each other, thus ASSIST and BASE both differ significantly from AUTO, but do not different significantly from each other.)

	Mean (sd)
ASSIST	0.58 (0.117)
BASE	0.652 (0.197)
AUTO	0.348 (0.081)

Test (difference between conditions)	p	
ASSIST-AUTO	0.001	*
BASE-AUTO	8.53e-5	*
ASSIST-BASE	0.73	

5.1 How Suggestions Helped

In order to better understand how sorters used recommendations, we can look at a sample user's track over time (Figure 12). This user, from the between-subjects study, was in the FBK ASSIST condition.

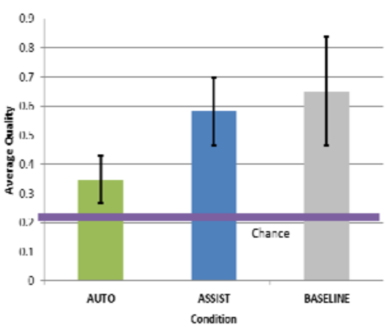


Fig. 11. The relative quality of all groups across AUTO, ASSIST, and BASELINE. Bar height is the average quality; error bars represent the std. dev. ASSIST and BASELINE are each significantly different from AUTO, and not from each other.

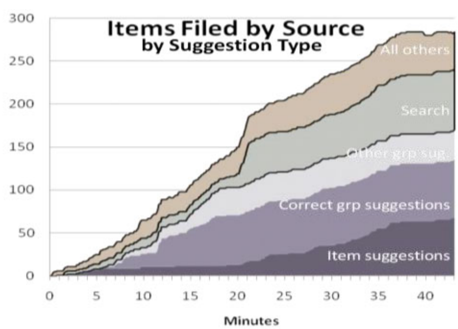


Fig. 12. Stack chart showing numbers of items clustered over time for one sorter, separated by item source. Note the large proportion of items from suggestions.

A user can place documents into a cluster from three sources: they could take documents from the piles or from the desktop, carry out a full-text search, or get a suggestion from either the item suggestion or the group suggestion. We track the time at which the user dropped each item into a cluster, and the source of the item. Thus, if a user did a full-text search that revealed a set of documents, but only sorted two of them into clusters, then the "search" line would increment twice, for each of those two documents.

We can see from Figure 12 that this sorter sorted far more documents from recommendations than from search or from the pile. Text search never provided more than half of the sorted documents at any point: indeed, text search provided most of its value in one particularly fruitful search (around minute 22).

The suggestions are further subdivided into three parts. When the user clicks on a group and asks for suggestions, we distinguish between when a user places the item in the recommended group ("correct") and when a user places the item in another group ("other"). In addition, item-to-group suggestions are always available; we count these only if the user follows the suggestion.

Note that item suggestions grew rapidly over the course of the study, especially after the half-way point: at this point, the suggestions are increasingly useful and the user is able to very rapidly sort items into groups.

6 Discussion

6.1 Comparing iCluster to DataMountain Implicit Queries

It is interesting that we found a significant improvement in the speed of sorters in comparison to the implicit query work on DataMountain (DM-IQ) [8]. The previous work had found that sorters were slowed by recommendations; we found the opposite.

We believe that the differing structure of the learning algorithm may help explain the distinction. DM-IQ was based on a static recommendation, and did not adapt as the user placed their items. If a user's mental model matched the system's model, the suggestions would be helpful; however, if the user had a different sort in mind, then the highlighting would be distracting. There is some evidence for this distraction in the original DM-IQ work: the paper reports that users "often considered whether or not to follow the system recommendation" [8].

As our Hypothesis 1 suggests, different users' clustering were very different from each other (which [8] also finds), which would reduce the utility of static suggestions.

6.2 Starting with a Blank State

The iCluster interface begins with a blank slate: users do not get an overview of the dataset. This differs from systems like Scatter/Gather, which provide an initial structure. We received mixed feedback on this: several users preferred to start from scratch; others wanted more of an overview about the dataset as a whole.

The current form closely models the real-world task of dealing with a pile of data. However, it suffers from a slow start, with weak recommendations until clusters contain several items. The drawback to an overview is that it requires a naïve distance metric to generate the initial view; this might introduce a level of bias and possibly distrust. We expect the tradeoff between these approaches is problem- and user-dependent, and hope to further explore this in our future work.

6.3 Current Use

The system has been in use in several different groups at our organization. Each of these groups collects user feedback on software products; their interest is in organizing the feedback into groups that can be assigned to particular debugging and development teams. During periods of intensive feedback collection, iCluster has been used for triaging larger collections of data.

6.4 Scaling Up

In Basu *et al.*'s original algorithm [17], suggestions improve as users sort more items (from 25% accuracy after three examples to 40% accuracy after eight). As clusters get larger, and the user creates more of them, suggestions should become more relevant, allowing users to evaluate them more rapidly and to accept a larger percentage of suggestions. We hope to add automatic features to allow a user to rapidly accept all recommendations for a cluster, and to cluster thousands of items in the background. This should help speed the system more as users work longer.

6.5 Qualitative Feedback

After the study, several sorters asked for the system to sort their email. One former program committee member said that the system would be "awesome for a Paper Committee sorting task." Several sorters noted that even when the recommendations for a cluster were not quite right, the suggestions were often appropriate for another cluster and so were useful. We attribute this to the learner building a model of items the user was likely to cluster.

7 Conclusion

iCluster implements an interactive clustering algorithm which learns as users add more data. While it becomes increasingly accurate as clusters get larger, it begins to show effects almost immediately, allowing users to rapidly train the model and get useful recommendations. With the help of our adaptive tool, users were faster than without it, and produced clusters of equal quality. Clusters produced with our tool were significantly better than clusters created automatically on the same datasets. We've been able to apply this tool to a wide variety of domains, ranging from conference paper sorting, to user feedback, and are currently exploring using it to cluster medical patient progress graphs.

References

1. Agarawala, A., Balakrishnan, R.: Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In: Proc. CHI 2006 (2006)
2. Jain, A.K.: Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31(8), 651–666 (2010)

3. Basu, S., Fisher, D., Drucker, S., Lu, H.: Assisting Users with Clustering Tasks by Combining Metric Learning and Classification. In: AAAI 2010 (2010)
4. Bauer, D., Fastrez, P., Hollan, J.: Spatial Tools for Managing Personal Information Collections. In: Proc. HICSS 2005 (2005)
5. Blei, D.M., Ng, A., Jordan, M.I., Lafferty, J.: Latent Dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (2003)
6. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.: Reading Tea Leaves: How Humans Interpret Topic Models. In: Proceedings of NIPS (2009)
7. Cohn, D., Caruana, R.: Semi-Supervised Clustering: Incorporating User Feedback to Improve Cluster Utility. In: Proceedings of the Conf. on Artificial Intelligence. AAAI Press, Menlo Park (2000)
8. Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S., van Dantzich, M.: Visualizing implicit queries for information management and retrieval. In: Proc CHI 1999 (1999)
9. DesJardins, M., MacGlashan, J., Ferraioli, J.: Interactive Visual Clustering. In: Proc. of the Int'l Conf. on Intelligent User Interfaces. ACM Press, New York (2007)
10. Dourish, P., Lamping, J., Rodden, T.: Building Bridges: Customisation and Mutual Intelligibility in Shared Category Management. In: Proc. GROUP 1999 (1999)
11. Jones, W., Dumais, S.: The spatial metaphor for user interfaces: experimental tests of reference by location versus name. ACM Trans. Inf. Syst. 4(1) (January 1986)
12. Malone, T.: How do people organize their desks?: Implications for the design of office information systems. ACM Trans. Office Info. Sys. 1(1), 99–112 (1983)
13. Mander, R., Salomon, G., Wong, Y.Y.: A "pile" metaphor for supporting casual organization of information. In: Proc. CHI 1992 (1992)
14. Pirolli, P., Schank, P., Hearst, M., Diehl, C.: Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection. In: Proc. CHI (1996)
15. Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., van Dantzich, M.: Data mountain: using spatial memory for document management. In: Proc. UIST 1998 (1998)
16. Watanabe, N., Washida, M., Igarashi, T.: Bubble clusters: an interface for manipulating spatial aggregation of graphical objects. In: Proc. UIST 2007 (2007)
17. Whittaker, S., Hirschberg, J.: The character, value, and management of personal paper archives. ACM Trans. on Comp.-Human Int. 8(2), 150–170 (2001)
18. Zeng, H.-J., He, Q.-C., Chen, Z., Ma, W.-Y., Ma, J.: Leaning to Cluster Web Search Results. In: Proceedings of SIGIR (2004)

Sharing Ephemeral Information in Online Social Networks: Privacy Perceptions and Behaviours

Bernardo Reynolds, Jayant Venkatanathan, Jorge Gonçalves, and Vassilis Kostakos

Madeira Interactive Technologies Institute
University of Madeira
{bernardo.reynolds,vjayant,jorge.goncalves,vk}@m-iti.org

Abstract. This paper presents a study where the online Facebook practices of a sample of users (n=103) was analysed over a period of two years, via the scraping of data in Facebook and the collection of questionnaire data. The data allows for a contrast between implicit and explicit attitudes regarding Facebook and online sharing. Our analysis reveals that while overall privacy concerns are not reflected in posting behaviour, awareness and familiarity with privacy controls is. This is supported by contrasting users' attitudes regarding day-to-day sharing against actual behaviour on Facebook. We theorise that there exists a failure in translating users' privacy needs into a social-technical environment such as social networking sites. This work demonstrates how aspects such as demographics and usage influence and shape users' behaviour and practices towards privacy. We therefore argue that the factorization of these aspects may augment the translation of users' privacy needs and improve the design of privacy sensitive mechanisms for day-to-day information sharing.

Keywords: Social media, sharing, privacy.

1 Introduction

Information sharing in the form of hyperlinks, videos, status updates or photographs, has gained increasing popularity with the advent of online social networking sites. With a wide variety of content and objectives, publishing comes in the form of small posts answering to questions such as "What's happening?" or "What's on your mind?" These mechanisms enable users to satisfy a natural tendency to explicitly disclose or publicise information about themselves, their opinions and activities [21].

The study presented here contrasts users' attitudes regarding day-to-day sharing against their actual behaviour on Facebook. While previous studies have examined this dichotomy in terms of "static" information such as email address, telephone number, or spouse names, relatively little work has considered this dichotomy in the context of everyday information sharing, i.e. sharing newly generated information, thoughts and ideas. This work is important in understanding how to design privacy mechanisms more appropriate for day-to-day sharing of information.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 204–215, 2011.

2 Related Work

Emerging communication technologies are fundamentally changing the way we behave, interact, and socialise [16]. As Boyd describes, information shared online becomes persistent, searchable, replicable and subject to invisible audiences [5], giving rise to significant privacy concerns. Because online social networks are denser and more diverse than those offline [17,18], thousands of users may be classified as friends of friends of an individual and be able to access shared personal information [9]. Another detail crucial to privacy is the fact that posts are typically displayed in forward or reverse chronological order. This allows others to infer contextual information about the author, especially when other contextual information like time of creation, platform used (web or mobile) or eventual responses to said posts are aggregated and grouped together, which is the convention in social networking sites. The diversity and richness of online social networks makes it challenging for users to group or categorize the individuals composing their social network, and adds complexity to the process of rule development. At the same time, emergent network effects may further affect group behaviour in terms of attitudes and usage of social systems [15].

Substantial work has considered rule development and privacy regulation, suggesting that these can be understood as a presentation of self [9][10], a formulation of dialectic and dynamic behavioural mechanisms depending on circumstantial context [2] or the conjugation of disclosure, identity and temporal boundaries [21]. What was once achieved with walls, doors and other physical or architectural constraints is still to be adapted to today's communication means [14,24]. Privacy management is an intricate process and is further augmented in a computer-mediated environment. On social networking sites, privacy regulation is a socio-technical activity involving interaction with the technological system and the group context. Individuals' privacy behaviour in such systems involves a mixture of technical and mental strategies. For instance, a technical strategy may involve the use of privacy settings to regulate content distribution to select audiences [23], while research has also shown that considering tie strength is another strategy for developing rules for disclosure [25].

Despite the evidence suggesting that users adopt objective strategies for controlling their privacy online, previous work has identified a discrepancy between people's privacy attitudes towards sharing information and their actual sharing patters, albeit in the context of static information such as contact details and preferences [1,20]. This behaviour has been termed the "privacy paradox". For instance, studies on the sharing of information by students on Facebook suggests that the number of users who actually change their privacy settings is remarkably scarce [8,11], while a follow-up study further revealed a high discrepancy between stated concerns and actual behaviour towards sharing static profile information on Facebook [1]. Other studies have further established the privacy paradox on social networking sites [24]. A tentative explanation for this dichotomous behaviour is that the perceived benefits of Facebook usage outweigh the risks of disclosing personal information [6]. In addition technical issues may also contribute: Facebook has received considerable criticism due to their technical shortcomings in the subject of privacy, possibly due to the fact that the norms developing around this technology are subject to continuous evolution [4].

Despite these changing norms, however, previous studies have consistently identified a dichotomy between users' proclaimed privacy concerns and their actual behaviour [1,20].

While previous studies have highlighted the privacy paradox in the context of online social networking, most studies have focused on the sharing of static profile information. Arguably, this information is qualitatively different from the information that users share on a day-to-day basis. Research shows posting about current activity and implying location is a common practice on social networking site users [22]. Such information is more likely to be of an ephemeral nature, and users can always decide on the spot about how and with whom to share this information. The study described next aims to address this issue and to uncover users' posting practices in dealing with ephemeral information.

3 Study

3.1 Theoretical Approach

A study was designed to test whether users' perceptions of privacy towards the sharing of ephemeral information matches their actual behaviour. Given that significant evidence suggests the existence of the privacy paradox, it was hypothesised the same applies to disclosure of ephemeral information:

- H1: Users' privacy attitudes towards sharing ephemeral information do not match their actual behaviour.

Previous research also shows an influence of gender on privacy attitudes [13,23]. Furthermore, while not all teens are members of social network sites, these sites developed significant cultural resonance amongst American teens in a short period of time, as well as Europe. Given that age has a significant influence on social networking site adoption, it is hypothesised that this also shapes privacy concerns in relation to sharing.

- H2: Different demographics have contrasting privacy concerns towards sharing of ephemeral information

3.2 Method

This study adopted a multi-pronged approach to collect data. Two complementary datasets were collected: a quantitative data set gathered using a Facebook scraping application, and a qualitative dataset gathered through an extensive online questionnaire.

Participants were recruited by advertisements in Facebook, via email, and via notice boards. Participants were promised a chance to enter a draw that would award 40 Euros to 10 participants. Interested participants were invited to add to their profile a Facebook application and notified about the crawler application upon installation. Contacts to address any concerns regarding the data collected were provided at various stages of the studies but never solicited.

The application collected information about participants' past behaviour, including wall posts, status updates, friends, privacy settings for posts and basic user information. Status updates were explicit posts written by participants and wall posts were any other items in participants' wall. The software then has to complete a follow-up questionnaire. Apart from gathering data the application also provided the link to the online questionnaire.

The questionnaire study consisted of 40 questions capturing data for demographics, Facebook usage, privacy attitudes, posting practices, and community. In addition the questionnaire had 10 scenario exercises that asked participants to imagine themselves disclosing various types of information and deciding how to share that information with different people in their network. In example, "last night was so much fun but I have a huge hangover" or "I'm late for a class/meeting".

The questionnaire, partially inspired by a previous study on the subject of privacy on Facebook [23], was designed as a between subjects study. The manipulated variable was the optimism of the questions collecting disclosure attitude data: in the optimistic condition participants were asked about what information they wished to share, in the pessimistic condition they were asked about which information they wished to hide. In both conditions the scenarios themselves were identical and the information to be shared was related to participants' activity, location or even personal pictures.

Although Facebook started in a college-oriented context and grew mostly through student networks it has now expanded and diversified to reach all sorts of demographics. The participants mainly came from the university environment. Faculty staff and students, under the premise of being able to win gift certificates, participated in the study. These factors may have introduced some response bias.

4 Results

The study was conducted between April and May 2010 with 103 participants (61 male) completing both parts of the study. The age ranges used in the questionnaire were: 1. less than 18 years old (no participants), 2. 18-24 years old, 3. 25-34 years old, 4. 35-44 years old 5. more than 44 years old (no participants). The logging application collected data from August 2007 onwards; earlier data was not accessible due to technical restrictions. The data included 4942 wall posts and 2769 status updates. As a baseline level of relationship between participant's questionnaire and log data, the correlation between stated and actual number of friends was 0.783, with participants on average underestimating their true number of friends by about 16.

H1: Users' privacy attitudes towards sharing ephemeral information do not match their actual behaviour.

Hypothesis 1 was partially confirmed. Data from the questionnaire showed that most respondents claimed to be Facebook users for 1-2 years, to visit their account daily (38.8%) for 15 minutes (30.1%) and use it mostly to read the newsfeed (76.7%) and posting content (66.9%). With respect to posting practices, the participants asserted that they did not post frequently on Facebook (avg. 2.4 on a 5-point Likert scale), but the majority claimed to post more often compared to 6 months ago (42.7%). The log data showed an overall average of 8.95 posts per person per month. This average was 26.4 for the three months leading up to the study, while it was 2.04 over the 3-month period 6 months earlier. Similarly, the average number of explicit status updates was 5.8 per person per month for the whole log, while this average was 5.25 for the three months leading up to the study, and 5.28 for the 3-month period six months earlier.

Regarding privacy attitudes, the questionnaire data shows that participants showed little intent to hide the information they shared from a friend or group of friends (avg. 1.9 on a 5-point Likert scale) and most (54.4%) admitted never having used Facebook's feature that would allow them to do so. The logged data confirms this claim as only 4 different participants chose to hide a total of 68 posts and only one did so for more than 1 friend. Figure 1 shows a both the stated and actual privacy settings that participants attributed to their own posts.

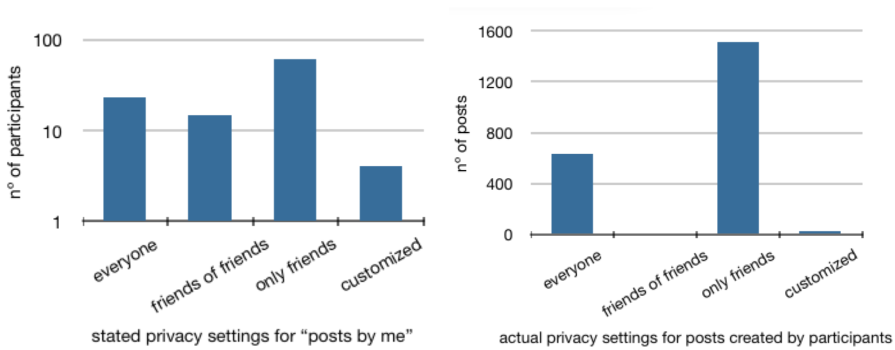


Fig. 1. Stated vs. actual distribution of privacy settings for individual posts

In their last privacy settings review, participants claimed to have opted for a more private setting (avg. 2.1 on a Likert scale ranging from private to public). Regarding privacy concerns, respondents acknowledged to be worried if all their Facebook friends knew where they had been last night (avg. 3.3 on a 5-point Likert scale), what they were doing at the moment (avg. 3.4). Participants were more worried about non-friends (i.e. strangers) knowing their location at the moment (avg. 4.1), and what they were doing at the moment (avg. 4.1).

Table 1 shows the relationship between participants' subjective concern about privacy with the number of posts that are shared with big audiences (everyone or all friends) and the frequency of posting. Similarly, Table 2 shows the relationship between participants' self-perceived frequency of changing their privacy settings on the number of posts that are shared with big audiences (everyone or all friends) and the frequency of posting.

Finally the results revealed no significant impact of the phrasing of the questionnaire (optimistic vs. pessimistic) on participants' responses, and hence all questionnaires were considered in further analysis.

Table 1. Mean ratio of user's posts visible to "everyone" or "all friends", and mean posts per month (ppm) broken down by users' concerns towards privacy

	(low)				(high)
How concerned are you with privacy?	1	2	3	4	5
Ratio of respondents' posts with privacy set to "everyone" or "all friends"	27%	32%	26%	27%	31%
ppm:	43	85	51	56	62

Table 2. Mean ratio of user's posts visible to "everyone" or "all friends", and mean posts per month (ppm) broken down by users' attitudes towards privacy

	Rarely				Very Often
How often do you change your privacy settings?	1	2	3	4	5
Ratio of respondents' posts with privacy set to "everyone" or "all friends"	50%	28%	23%	16%	20%
ppm:	65	59	57	72	23

H2: Different demographics have contrasting privacy concerns towards sharing of ephemeral information

Hypothesis 2 was also confirmed. In assessing the effect of gender, a one-way ANOVA shows a significant effect of gender on posts per month ($F(1,101)=46.404$, $p<0.0001$), showing that females post more than males. Furthermore, a chi-square test shows a significant relationship between gender and current privacy setting for "posts by me" ($x^2=9.721$, dF=3, $p<0.05$), showing that females have more "open" privacy settings than males.

In terms of the effect of age, a chi-square test shows a significant relationship between users' age group and the frequency with which they change their privacy settings ($x^2=15.58$, dF=8, $p<0.05$), showing that older users change their settings more frequently. Additionally, a chi-square test shows a significant relationship between users' age group and their current privacy setting for "posts by me" ($x^2=34.48$, dF=6, $p<0.0001$), showing that older users have stricter settings.

On average males had 8.04 posts per month on their wall, while females had 9.58. Additionally, 35% of posts by females were visible to "everyone" or "all friends", while 25% of posts by males had these privacy settings. Privacy attitudes based on gender are shown in Table 3.

Table 3. Gender effects on privacy attitudes

	Gender	
	Female Mean (SD)	Male Mean (SD)
How concerned are you with privacy?	3.88 (1.06)	3.49 (1.09)
How often do you change your privacy settings?	2.55 (0.94)	2.43 (0.97)

Furthermore, a chi-square showed a significant relationship between participants' age and their reported frequency of changing privacy settings ($x^2 = 240.1$, df = 15, $p<0.0001$), with older participants claiming to change their settings more frequently. In addition, the youngest age group posted 7.73 posts per month, gradually increasing to 11.35 posts per month for the oldest age group. The average ratio of posts visible to "everyone" and "all friends" gradually decreased from 33% for the youngest age group to 26% for the oldest group, while their response to "how concerned are you with privacy" increased from 3.5 to 4.0 on a 5-point likert scale respectively.

5 Discussion

5.1 Subjective Perceptions vs. Actual Behaviour

The baseline level of correlation between perceived and true number of Facebook friends shows that participants don't accurately account for their entire network, which may result in poor boundary regulation as users occasionally, and by accident, may give out too much information and experience unwanted approaches [3].

Users were accurate in their assessment that they now post more than in the past. While this is true on average, it does not apply to explicit status updates, but rather to wall-sharing posts. The relative frequency of each activity provides further evidence that on-the-fly ephemeral sharing is more popular than sharing explicit ideas and thoughts, and certainly more popular than the sharing of static profile information.

In terms of participants' perceptions on privacy, the findings contradict previous results pointing in the direction of an opt-out dynamic for sharing, suggesting that users should have to choose what they wish to hide [5]. Participants responding to the questionnaire had limited intention of explicitly hiding the content they create, and in fact very few actually use of the available means to do so.

However, participants did claim concern about potential privacy threats on Facebook, even though most ephemeral information shared by participants did not reflect these concerns by means of explicit privacy settings. In fact, there was little correlation between participants' broader concern about privacy on Facebook and their actual posting practices: both the number of postings and the portion of those posts visible to a large audience appear to be independent of general privacy attitudes (Table 1).

On the other hand, participants' self-assessment of their frequency of change of privacy settings are in line with their actual posting behaviour: the number of posts and the ratio of posts visible to large audiences decrease as user's self-perceived frequency of changing privacy settings goes up (Table 2).

5.2 Gender

In terms of gender differences, females were more concerned about privacy, as also shown in previous studies [13,23]. However, in their actual behaviour females post more often and with fewer privacy restrictions than males. We further analysed the effect of gender in the questionnaire data. In the disclosure scenarios, females were more open across all scenarios except for one. Aside from the first scenario, the optimistic approach was more popular among females than among males. The same was verified on the section where optimistic vs. pessimistic approach was the manipulated variable on a between subjects design. Females were less prone to want to hide their posts when compared to males. Moreover, females made less use of the on-the-fly privacy settings present in the publisher interface.

The results do not imply that females are not worried about privacy, but rather they suggest that females are less pro-active on this issue. These findings demonstrate an incongruity. Although females demonstrate higher concerns in privacy matters, that is not reflected in the strategy they prefer to manage the information they share, nor in the use of available means to better manage the privacy of that information.

5.3 Age

Older users claimed to be more concerned with privacy, and their attitude was actually reflected in their posting practices: even though they posted more frequently, their privacy settings were more restrictive than those of younger users. When considering their choice for a pessimistic approach on the information disclosure scenarios, there was no significant difference between younger and older adults. On the other hand, the pessimistic approach resonated more among older participants than among younger ones in the between subjects section of the questionnaire.

Furthermore, older users were less satisfied with the control Facebook provided over their privacy and stated to change their privacy settings more often. Despite this fact, the number of older participants actually using on-the-fly privacy settings to manage the information they share was significantly lower than of younger ones. However, although less elders use this features, the ones who do, do it more frequently than youngsters. These findings contradict recent research demonstrating that young-adult Americans have an aspiration for increased privacy [12] and that the majority of young adult Facebook users are far from being nonchalant and unconcerned about privacy matters [7]. Our findings contradict these claims to the extent that when compared with elders, young-adults exhibit less concern on the subject of privacy. We find that these findings hold for satisfaction with Facebook's privacy controls, frequency of adjusting privacy settings and last privacy controls review.

In addition, we observed that older users claimed to be more concerned with privacy and their attitude was actually reflected in their posting practices: even though they posted more frequently, their privacy settings were more restrictive than those of younger users. Stated preferences in privacy settings for sharing ephemeral information are consistent with this finding. Older participants stated to use more restrictive settings and also to use less permissive settings when compared to younger participants. Previous research [6] also found that teenagers' rhetoric about online safety with regard to social media mirrored the narratives presented by the news media. This is in agreement with our results in reference to the importance of privacy in the public debate, which was not affected by age. Both young-adults and older ones equally valued the issue.

5.4 Behaviour over Time

Boyd and Hargittai [7] give prominence to an escalation of engagement with privacy settings from 2009 to 2010. This is congruous with participants' declared adoption of a more private profile in their latest privacy settings review. The proposed justifications for this behaviour are: the increase in public attention to privacy matters, the increased changes in Facebook's default settings and the site prompts to review privacy settings. While we concur with these assumptions we also propose the upsurge in sharing information as a cause for the rise in engagement with privacy settings. Even though our findings reveal an inverse relationship between posting frequency and adjustment of privacy settings, users often rely on the privacy settings to regulate content distribution to select audiences [23].

5.5 Privacy by Design

Overall, our findings on the influence that gender and age have in privacy attitudes are inline with previous research [7,13,19,23]. Women, in general, are more worried about their privacy when compared to men. Taking into account different criteria such as stated privacy concerns, choice of approach for managing disclosures, preferences in privacy settings to control posts, and other, we can infer that males state to be more open when it comes to sharing information. Lewis et al [19] propose safety concerns as the central cause for this phenomenon. Such safety issues revolve around fear of abuse of social networking sites (SNSs) by sexual predators and other information disclosures risks associated with SNSs. However, such concern is not reflected on their actual behaviour nor on the strategy females chose to manage disclosure.

As for age, younger participants were more permissive and adopted looser controls when sharing information. There is a number of factors that can serve as justification for this circumstance. One is youngsters being born amidst the digital world of today, therefore having more familiarity and trust in services like Facebook. Another is ingenuity as conscience and accountability of actions is something humans grow through life. And as with many other matters, youngsters are typically careless of the consequences that rise from their actions. Assuming this applies to the act of sharing ephemeral information, it is also safe to assume youngsters are less concerned with the consequences of not managing their privacy.

Significant privacy implications can arise from over-sharing, and the results presented here provide some guidance on how to limit the extent of this phenomenon. First, the results show that users underestimate the size of their audience; therefore it is important to highlight and remind users of its extent.

In addition, users' general privacy concerns are not necessarily manifested in their posting behaviour, with both young and female user groups sharing with larger audiences, thus making them good target groups for redesigned privacy controls. The potential benefit of such a design strategy is underlined by the fact that users are aware that in fact they now share more than in the past, and those who claim awareness and frequent use of privacy controls do in fact share less often and to less people.

5.6 Design Guidelines

Summarising the results and findings from the study, we present here the following set of principles and respective guidelines.

Principle: *Demographics shape behaviour towards privacy*
We found divergent privacy concerns and practices across gender and age. These two demographic factors had an influence on privacy concerns, choice of approach to disclosure and others.

Guideline: Consider the demographic factors we presented and accommodate those in the design.

Example: Allowing both experienced and inexperienced users to translate their privacy intentions into privacy settings.

Principle: *Usage patterns shape behaviour towards privacy*
We demonstrated usage patterns such as time since joining Facebook, frequency of logins to Facebook, time spent on each login and others have a massive influence on privacy practices and concerns like privacy concerns, latest privacy controls review choice and frequency of changing privacy settings.

Guideline: Consider all the usage factors we presented and accommodate those in the design.

Example: Build usage based privacy settings, e.g. frequently suggest new users of the system to protect their content.

6 Conclusion

This paper presents a study that contrasts users' privacy perceptions about sharing ephemeral information against their actual behaviour. Previous work suggests a dichotomy between these. The findings reveal that while overall privacy concerns are not reflected in posting behaviour, awareness and familiarity with privacy controls is. We theorise that there exists a failure in translating users' privacy needs into a socio-technical environment such as social networking sites. This study suggests that such a translation relies on successfully factorizing aspects such as, but not limited to, gender and age.

There are a number of ways that the work presented here can be extended. First, one could implement a narrowcasting solution grounded in this work to perform usability tests. In addition, quantitative analysis on user adoption for the Facebook Groups feature could be conducted because it supports narrowcasting shared information. More specifically, user impressions and attitudes towards Groups could be captured. Finally, this kind of work could be extended to a network like Twitter, which is 100% based on status updates, thus rendering further insights.

Acknowledgments. We thank Tasos Spiliotopoulos for his valuable comments and everyone who took part in the study. This work is funded by the Portuguese Foundation for Science and Technology (FCT) grant CMU-PT/SE/0028/2008 (Web Security and Privacy).

References

1. Acquisti, A., Gross, R.: Imagined communities: awareness, information sharing, and privacy on Facebook. In: Danezis, G., Golle, P. (eds.) PET 2006. LNCS, vol. 4258, pp. 36–58. Springer, Heidelberg (2006)
2. Altman, I.: Privacy Regulation: Culturally Universal or Culturally Specific? Journal of Social Issues 33(3), 66–84 (1977)
3. Baarkhus, L., Tashiro, J.: Student Socialization in the Age of Facebook. In: Proc. of the Conference on Human Factors and Computing Systems: CHI 2010, pp. 133–142. ACM Press, New York (2010)
4. Bellotti, V., Sellen, A.: Design for Privacy in Ubiquitous Computing Environments. In: Proc. Third European Conf. Computer-Supported Cooperative Work ECSCW 1993, Milano, Italy, pp. 77–92 (1993)
5. Boyd, D.: Facebook's Privacy Trainwreck: Exposure, Invasion, and Social Convergence. Convergence 14(1) (2008)
6. Boyd, D.: Taken out of context: American teen sociality in networked publics, Ph.D. dissertation, University of California at Berkeley (2008)
7. Boyd, D., Hargittai, E.: Facebook Privacy Settings: Who Cares? First Monday 15(8) (2010)
8. Debatin, B., Lovejoy, J., Horn, A., Hughes, B.: Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. Journal of Computer-Mediated Communication 15, 83–108 (2009)
9. DiMicco, J.M., Millen, D.R.: Identity management: multiple presentations of self in Facebook. In: Conference on Supporting Group Work, pp. 383–386. ACM Press, New York (2009)
10. Goffman, E.: The presentation of self in everyday life. Doubleday Anchor, Garden City (1959)
11. Gross, R., Acquisti, A.: Information Revelation and Privacy in Online Social Networks. In: Workshop on Privacy in the Electronic Society. ACM Press, New York (2005)
12. Hoofnagle, C., King, J., Li, S., Turow, J.: How Different are Young Adults from Older Adults When it Comes to Information Privacy Attitudes and Policies? (2010)

13. Joinson, A.N.: Looking at, looking up or keeping up with people?: Motives and use of Facebook. In: Conference on Human Factors in Computing Systems (CHI), pp. 1027–1036. ACM Press, Florence (2008)
14. Kostakos, V.: A design framework for pervasive computing systems. PhD Thesis, University of Bath, UK. Technical Report CSBU2005-02, ISSN 40-9497 (2005)
15. Kostakos, V.: Is the crowd's wisdom biased? A quantitative analysis of three online communities. In: Proceedings of Computational Science and Engineering, Vancouver, Canada, August 29-31, pp. 251–255 (2009)
16. Kostakos, V., O'Neill, E., Little, L., Sillence, E.: The Social Implications of Emerging Technologies. Interacting with Computers 17, 475–483 (2005)
17. Kostakos, V., Venkatanathan, J.: Making friends in life and online: Equivalence, micro-correlation and value in spatial and transpatial social networks. In: Proceedings IEEE SocialCom, Minneapolis, USA, pp. 587–594 (2010)
18. Lenhart, A.: Adults and Social Network Websites. Pew Internet and American Life Project (2009), http://www.pewinternet.org/PPF/r/272/report_display.asp (retrieved January 14, 2009)
19. Lewis, K., Kaufman, J., Christakis, N.: The taste for privacy: An analysis of college student privacy settings in an online social network. Journal of Computer-Mediated Communication 14(1), 79–100 (2008)
20. Norberg, P., Horne, D., Horne, A.D.: The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. Journal of Consumer Affairs 41(1), 100–126 (2007)
21. Palen, L., Dourish, P.: Unpacking Privacy for a Networked World. In: Proc. of the Conference on Human Factors and Computing Systems: CHI 2003, pp. 129–136. ACM Press, New York (2003)
22. Patterson, D.J., Baker, C., Ding, X., Kaufman, S.J., Liu, K., Zaldivar, A.: Online everywhere: evolving mobile instant messaging practices. In: Proceedings of UbiComp 2008, vol. 344, pp. 64–73. ACM, New York (2008)
23. Stuntzman, F., Kramer-Duffield, J.: Friends only: examining a privacy-enhancing behaviour in Facebook. In: Proc. of the Conference on Human Factors and Computing Systems: CHI 2010, pp. 1553–1562. ACM Press, New York (2010)
24. Tufekci, Z.: Can You See Me Now? Audience and Disclosure Management in Online Social Networking Sites. Bulletin of Science and Technology Studies 11(4), 544–564(21) (2008)
25. Wellman, B., Wortley, S.: Different Strokes from Different Folks: Community Ties and Social Support. American Journal of Sociology 96(3), 558–588 (1990)

An Investigation into Facebook Friend Grouping

Patrick Gage Kelley[1], Robin Brewer[2], Yael Mayer[3],
Lorrie Faith Cranor[1], and Norman Sadeh[1]

[1] Carnegie Mellon University
[2] University of Maryland
[3] Harvey Mudd College
{pkelley,lorrie,sadeh}@cs.cmu.edu
rbrewer@umd.edu, yael_mayer@hmc.edu

Abstract. With increasingly large friend networks, Facebook users may be losing sight of exactly with whom they are sharing content they post to Facebook. When Facebook released a new privacy interface in sum- mer 2010 they simplified privacy controls; however, group-based permis- sions remain at the core of fine-grained privacy control. In order to use these fine-grained controls, users must be able to accurately and usefully specify friend groups. In a series of 46 semi-structured interviews, we investigated how participants group their online friends using four differ- ent grouping methods. Our results show that these different mechanisms alter the strategies and groups that users create, that groups created a priori need further refinement before they can adequately address pri- vacy decisions, and that users are adapting their online behavior to avoid the need to specify groups in the current Facebook interface. We con- clude with several recommendations that would allow users improved group-based access control.

Keywords: grouping, online social networks, privacy, access control.

1 Introduction

Media organizations present Facebook privacy as a pervasive problem, with Internet shopping behavior disclosed, employees fired for behavior in tagged photographs, locations being shared, and the frequently changing settings described as a "bewildering tangle of options," the situation seems bleak [3, 16, 13, 7].

A May 2010 New York Times infographic detailed more than 170 privacy options that Facebook users could then manipulate. To restrict access to any group more selective than all of one's Facebook friends, the "Customize" option must be chosen, at which point pre-specified friend lists or individual friends can be explicitly granted or blocked from that type of content. When Facebook released a new privacy interface in summer 2010 they simplified privacy controls; however, the new controls still rely on the Customize option for selective access control.

While the friend list interface and privacy settings continue to change, as of April 2011 Facebook users continue to have the option to control their privacy using friend lists. In order to use this option, users must first create one or more groups of their friends using the Facebook friend list interface. They can then control access to

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 216–233, 2011.

specific posts or to general categories of content by designating specific friend lists to which access should be granted or denied. We seek hereto examine how users categorize their Facebook friends into lists and whether they tend to do so in a way that supports using friend lists for access control. In addition, we seek to understand the impact of grouping mechanism on the types of lists users create. We conducted 46 semi-structured interviews with Facebook users, investigating how participants group their friends using four different friend-grouping mechanisms.

We found that when presented with the task of grouping their friends, nearly all users relied on similar attributes such as place, time, and context to form initial groups. However, we also found that the different grouping mechanisms we tested affected the strategies users employed. More importantly we found that these mechanisms impacted qualities that one would expect to remain consistent, such as size of groups, number of groups, and overlap. This implies that while it remains possible there could be an internal “ground truth,” defined as a user’s set of ideal groups, the grouping mechanism can be used to—and necessarily does—influence the resultant groups. We also found that one-time grouping is not sufficient for privacy control, even in the small number of possible scenarios we focused on. And finally, we found that Facebook users tend to refrain from posting certain sensitive information on the site rather than posting it and using Facebook privacy controls to restrict access. These findings lead to a number of recommendations for designing group-based privacy controls for social networks.

2 Related Work

Social psychologists and cognitive anthropologists have studied how people categorize objects, concepts, and people for decades. Mervis and Rosch provide a summary of studies regarding the categorization of objects, specifically looking at what categories people define, the fuzziness of categorical boundaries, and the human process of classification [15]. Stangor et al. provide an explanation of how people categorize others, specifically people they do not know, based on simple physical attributes (sex, race) [20]. Categorization has also been looked at within the HCI community, where Malone’s early work explored how people categorize the (paper) files at their desks [12]. Malone recognized, and then studied, the pilers, whose apparently-messy desks do have a physical organizational structure. Kwasnik, commenting on her own investigation of categorization of files adds, “The research indicates that we choose categories, at least for concrete objects, that have the most usefulness for the least cognitive effort” [9]. While these foundational studies provide a holistic picture of how humans group, we seek to understand the grouping task specifically in the context of social networks.

Friend networks are composed of people that users meet in distinct contexts and at different life stages. Facebook and most other social network sites combine all of these people into a single group of “friends,” even though in real life users may have different sharing expectations for different friends and groups [2]. Additionally, sharing across the entire network may not be expected, and one study demonstrated that people having misconceptions about the size and extent of the network they are sharing their information with [1].

In a 232-participant survey, Binder et al. found that as networks continue to grow, broadcasting information across different social groups increases "tension" [4]. In a 20-participant online observation study, Lampinen et al. investigated how people currently managed "group co-presence" on Facebook. Their results showed that participants managed the differences in their groups by dividing the platform into separate spaces such as closed Facebook groups and by "using suitable channels of communication," i.e. directed messages, private pages, and off-Facebook space. They also relied on self-censorship and trusting their friends to be responsible and not share inappropriate data [10].

The Facebook privacy controls allow users to use their friend lists to customize their sharing preferences according to their groups of friends. However, creating friend groups is a secondary task completed by few Facebook users. In a small thesis study, Smart found that none of his 10 participants used friend lists to control their privacy settings [19]. Facebook founder Mark Zuckerberg stated in October 2010 that approximately 5% of Facebook's users have created even a single friend list [5].

While, most work in interfaces for privacy has not looked directly at online contact management, many of the same general principles, such as those catalogued by Lederer et al. apply [11], for example they specifically call out excessive configuration over action, a common issue in contact management. In Davis and Gutwin's research on awareness servers, they found that some types and fidelities of information will be more public than others, yet for more sensitive information the specific recipient will greatly impact the disclosure decision [6].

In a recent experiment, Jones and O'Neill created an automatic method to group Facebook users' friends. They downloaded 15 participants' entire friend networks—that is, their friends and their friends' connections—and used a clustering algorithm to group them. Participants also grouped their friends separately using the card sorting software xSort1[1] "as if they were grouping them for controlling their privacy on Facebook." The xSort software displays the names, but not the photos, of each friend. The participant-created groups were used as ground truth to test against the automatically-generated groups. Jones' results showed that the automatically-generated groups were at best around 70% accurate in comparison to the groups made by the participants [8].

In the final phase of this study, Jones and O'Neill asked participants to select a privacy-sensitive item from their profile and indicate their "willingness to share" this item with each of 100 contacts. These contacts were a stratified sample of the groups they had made earlier. The researchers found that using the human-generated groups 77.8%-90.8% of contacts were correctly granted or denied access to a privacy-sensitive item. This shows even groups created by participants with privacy in mind may not be adequate for access control [8]. Our research goes beyond the study by Jones and O'Neill to explore how grouping mechanisms impact the groups users create and their suitability for access control.

3 Methodology

To understand how Facebook users group their online contacts, we performed semi-structured interviews with 46 participants in our lab. We view this as an exploratory

[1] http://www.xsortapp.com/

study, consisting of a pre-survey, an open-ended grouping task, an interview exploring a series of common social network sharing behaviors, and a short post-survey. While we gave each participant one of four different mechanisms to group his or her contacts, we did not set out to test whether one was superior to the others; our goal was to illuminate the effect each interface had on grouping behavior, exploring participants' overall opinions and attitudes toward group-based access controls in social networks.

We recruited participants through a university research pool website, flyers around the city, and local Craigslist postings. Participants completed pre-surveys online before coming to our lab. Upon arrival at our lab, they signed consent forms allowing us to audio and video record (through video cameras or screen capture). For each interview, one member of our research team recorded the time and observations throughout the grouping task, including: size and labels of groups, grouping strategies, participant questions, hesitation, group rearranging, and additional notes. Quotes used in the following sections, from both the grouping task and subsequent privacy scenarios were transcribed from recordings after the interview. The post-survey was completed on paper, immediately after the main interview. Participants were paid $15 for successful completion of the entire task.

3.1 Grouping Mechanisms

Our participants were each assigned to one of four grouping mechanisms:

– **Card sorting (CS).** Participants were provided with a set of physical paper cards, each with the name and profile picture of one of their Facebook friends. The participants sorted the cards into piles during the grouping process as shown on the left in Figure 1.

– **Grid tagging (GT).** Participants were provided the names and profile pictures of each of their friends arranged across several sheets of paper in a grid. Participants "grouped" their friends by tagging the pictures with different colored markers as shown on the right in Figure 1. The participants indicated what group they were creating with each new color used. Participants tagged pictures with multiple colors to place people in multiple groups.

– **File hierarchy (FH).** Participants were provided a folder on a computer containing profile picture files of each of their friends. The picture files were named with the corresponding friend's name as shown in Figure 2. To group their friends, participants were instructed to create labeled folders and put the appropriate pictures in the folders. By default, friends were displayed as large icons so pictures were visible. Participants could choose whether to use a Mac or Windows computer for this task.

– **Facebook friend lists interface (FB).** Participants were asked to log in to their Facebook account and categorize their friends by creating standard Facebook friend lists. We directed participants to the "Account – Edit Friends – Friends" screen (see Figure 2, right) that, while unfamiliar to most users, allows users to see if each of their friends had been placed into a friend list. (By default the edit friends view also includes Facebook "Pages" that a user might also have "liked" and can also be placed into friend lists, which is why we directed users to click the Friends tab on the left).

Fig. 1. Left: A participant using the card sorting mechanism points to a group of one friend during the privacy scenarios section of the interview. Right: A participant using the grid tagging mechanisms reviews her friends, after having tagged them with colored markers during the privacy scenarios of the interview.

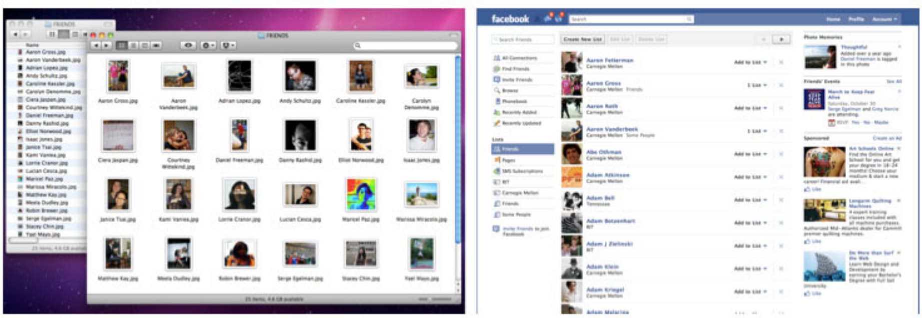


Fig. 2. Left: Screen capture of possible view options and friend photos in the file hierarchy mechanism. Right: Screen capture of Facebook's edit friends page, with the friends tab selected.

We did not explicitly tell participants that placing participants in multiple groups was something they could do, so as to not bias them towards doing so. However, we provided a way for this to occur in each mechanism, and did affirmatively answer questions regarding grouping when asked.

Through several early pilot tests we explored other interfaces and variations of those above. The xSort tool mentioned above, used by Jones [8], was eliminated because it cannot display pictures and it has no provision for allowing a friend to be placed in multiple groups. While we could have implemented our own digital card sorting application that met these requirements, we would then have had a condition fundamentally similar to physical card sorting.

We also tested several variations of our two physical methods, including using post-it notes, allowing for a drawable landscape underneath the card piles (for labels, subgroups, and group relationships), and starting with the cards spread on a table instead of beginning in a stack. Ultimately, these variations introduced more problems than they solved and were thus discarded. For example, beginning with the cards spread across a large surface caused a cognitively troubling and unfamiliar search problem for one pilot participant.

For the final study we chose the four mechanisms described above, representing a diverse set of alternatives, each with their own benefits and challenges. Card sorting was the initial motivation for the study, and remained the most intuitive to sort a set of objects. Grid tagging allowed for a folksonomy tagging structure of friends, where the groups were not central, but the contacts were. The file hierarchy was added to most easily allow for subgroups, encouraging nested folders and simple duplication. Finally, the Facebook friend list interface was used as the status quo, a common baseline and the interface participants have most likely previously used.

3.2 Survey and Interview

Potential participants were directed to an online survey as well as a custom Facebook app, that accessed a list of their Facebook friends. We used the survey and friend data to prepare the grouping task before the participant arrived at our lab. Our pre-survey included basic demographics questions, two questions about general Facebook usage, five questions about trust and privacy concerns on Facebook, four questions about current privacy controls, and six questions about Facebook's privacy settings changes. Participants completed this survey on their own computer when scheduling an appointment to come to our lab.

At their appointment, participants began with the open-ended grouping task. Next we walked participants through a number of specific scenarios. First, we discussed sharing certain pieces of information, including cell phone number, email address, political and religious views, relationship status, and sexual orientation. We also walked participants through a number of thought experiments to consider who they would share content with if they were posting pictures, sharing their location,[2] inviting friends to a party, or posting a status update. We also explored some privacy-sensitive scenarios, including asking about vacation photos, gossiping about a superior, untagging a photograph, and removing a post from a participant's wall.

At the conclusion of the interview task, we had participants fill out a short paper survey that asked about their use and frequency of updating friend lists, their use of Facebook's "limited profile" feature, if they accept friend requests to "just be nice," and if they enjoyed the grouping task.

4 Demographics and Survey Responses

64 participants took our online pre-survey. Of those, 46 participants came to our lab and completed the interview and grouping tasks. For the remainder of the paper we will focus on these 46 participants. For the purposes of the paper we will refer to each participant who completed the study by condition/number (for example, Participant GT2 is the second participant in the grid tagging condition). For a summary of basic demographics see Figure 3. Our study included 27 male participants (59%) and 19

[2] This study was conducted before the release of Facebook Places, so in the study it was posed as a theoretical feature that offered real-time location sharing.

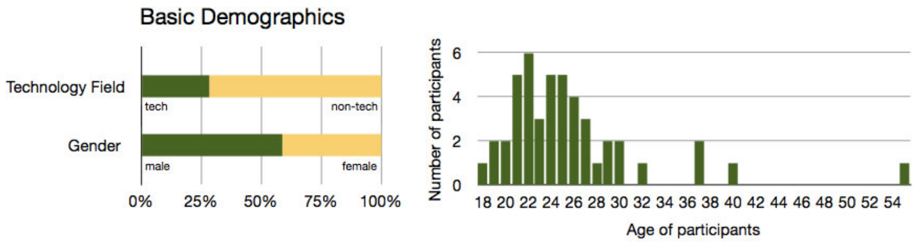


Fig. 3. Basic participant information including: field of study/employment, gender, and age

female participants. The average age was 25.6 years, younger than the national population, common among Facebook's population. 13 participants (28%) self reported that they studied or were employed in a technology-related field.

Our participants also closely align with usage patterns of active Facebook users, with all but 2% of participants reporting they log in at least several times per week, and 69% reporting they use Facebook multiple times per day. Facebook reports that 50% of its active users log in on any given day. When asked how frequently users "update your status or post content on Facebook," 63% of users reported that they contribute information (not just consume) at least once per week. Facebook reports an overall average of 130 friends across all users. Our users have an average of 280 friends (median 222.5 friends).

35% of participants agreed with the statement "I trust Facebook with my personal information" and 43% disagreed (22% neutral). As a possible compensation for users' distrust in Facebook, 41% of our participants agreed that "I don't worry about Facebook privacy because I have a strong set of privacy rules" (39% neutral, 19% disagree). A striking 87% of users reported agreement with "Once I put information on Facebook, it's not truly 'private' anymore." 32% of participants agreed that "Facebook changes privacy controls too frequently." However, 71% of participants believed "Facebook should ask for user input before making changes," and 96% agreed that "Facebook should announce any planned changes in advance." Nonetheless, 83% of participants agreed that they did not read Facebook's new privacy policy in detail.

From our post-survey at the end of the study we found that, while many might assume the task of sorting all of one's Facebook friends to be a daunting one, 76% of participants agreed or strongly agreed that they enjoyed the study, with the rest expressing neutrality. Many participants with several hundred friends stated they enjoyed the study, suggesting people may be willing to group their friends, if the correct incentives are in place.

Of the 46 participants, 17 participants (37%) reported using friend lists in their Facebook accounts. However, only 8 of these people stated that they used these lists to control their privacy settings. In addition, only 9 participants (20%) stated that they try to do monthly maintenance of their friend lists. While these numbers seem to suggest our participants are biased to be more concerned about privacy, these numbers are self reported (we did not check for list usage nor maintenance) we also

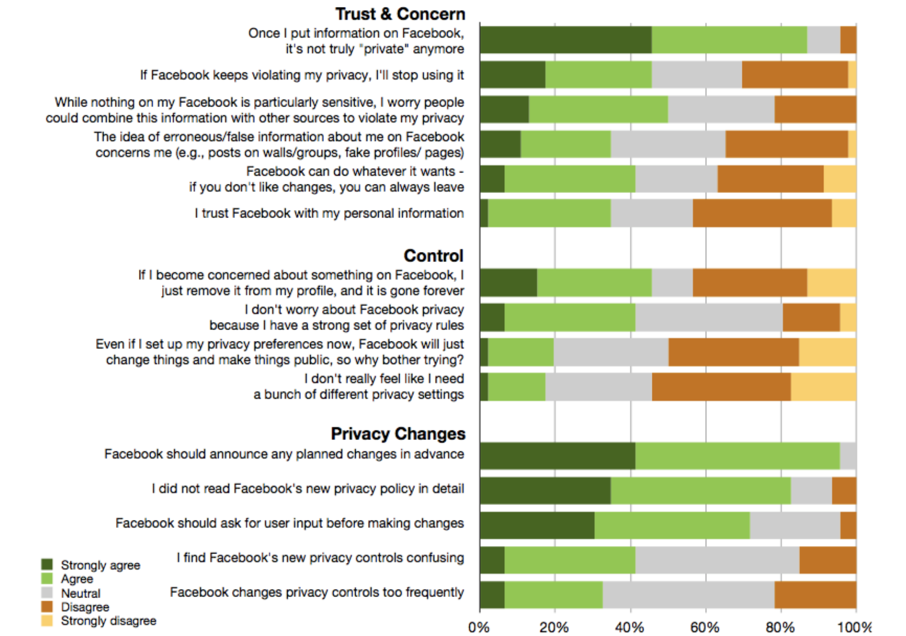


Fig. 4. Pre-survey responses from our 46 interview participants. Each of the above statements was asked with a 5-Likert scale from 1 (Strongly agree) to 5(Strongly disagree).

note that while Facebook normally gives figures distinguishing between active and all users, the 5% figure for friend list usage was not clearly associated with either group, and may have been taken from total usage to make friend lists seem less popular (as this quote is taken from an event with a product intended to replace friend lists as antiquated and useless).

5 Common Grouping Strategies

Many of our participants when tasked with the intentionally open-ended prompt "using (the cards/the markers/interface), categorize your friends into different groups and tell us what you're thinking as you do so," began instinctively and immediately sorting through and categorizing their friends, while others spent time looking at their friends before deciding how to proceed. FB10 began the task stating: "There are all sorts of different ways to categorize the people you know." GT3 said "Oh gosh, like I kind of have to look at all of them before I start a group." Some wanted further details. GT10 asked, "Is this open-ended?" FH2 asked, "By different characteristics, or by female, male?" FB1, at a loss for what to do, asked if we had any recommendations for groups he could make.

We observed two common strategies for performing the original grouping task. The first, which we refer to as "by friend," consists of a participant examining a single

Table 1. Overview of our 46 survey participants, including the mechanism they were assigned, the number of friends they had, the number of groups they made, the strategy they used to first categorize their friends, and whether or not their original groups included overlap. Timing information is given in seconds per friend, and was not available for all participants.

Participant overview

#	Gender	Age	Num. friends	Num. groups	Style	Overlap	secs. per friend	#	Gender	Age	Num. friends	Num. groups	Style	Overlap	secs. per friend
Card sorting								**Facebook interface**							
CS1	Female	21	593	8	friend	no	2.10	FB1	Male	22	227	2	group	no	2.29
CS2	Male	24	126	7	friend	no	4.20	FB2	Male	27	169	3	group	overlap	4.37
CS3	Female	21	436	9	friend	no	2.07	FB3	Female	20	229	6	group	overlap	n/a
CS4	Male	32	97	5	friend	no	2.62	FB4	Male	20	428	6	group	overlap	n/a
CS5	Male	22	215	8	friend	no	n/a	FB5	Male	29	152	4	group	overlap	4.13
CS6	Male	23	658	5	friend	no	1.85	FB6	Female	21	407	7	group	no	n/a
CS7	Male	26	99	3	friend	no	n/a	FB7	Male	25	214	5	group	overlap	4.54
CS8	Male	30	212	9	friend	no	2.59	FB8	Female	21	375	2	group	no	0.50
CS9	Male	26	156	6	friend	no	3.67	FB9	Male	21	259	4	group	no	2.08
CS10	Male	28	54	5	friend	no	1.96	FB10	Male	22	395	2	group	no	0.95
CS11	Female	27	6	3	friend	no	1.50	FB11	Male	22	366	5	group	overlap	n/a
CS12	Female	37	274	14	friend	no	n/a	FB12	Male	19	889	11	friend	overlap	6.05
Grid tagging								**File hierarchy**							
GT1	Male	24	108	5	group	no	6.12	FH1	Male	23	145	5	group	no	5.59
GT2	Female	25	218	7	group	no	5.31	FH2	Female	27	149	5	friend	no	5.01
GT3	Male	22	671	30	friend	overlap	5.27	FH3	Female	24	379	6	friend	no	n/a
GT4	Female	19	705	7	friend	no	2.90	FH4	Male	26	269	5	group	no	4.08
GT5	Male	25	391	9	group	overlap	6.47	FH5	Male	55	12	3	friend	no	12.42
GT6	Female	26	62	7	group	no	9.71	FH6	Female	24	206	3	group	no	4.08
GT7	Male	30	164	4	friend	no	3.16	FH7	Female	22	155	13	friend	overlap	5.30
GT8	Female	37	255	8	friend	overlap	n/a	FH8	Male	25	422	8	group	no	4.76
GT9	Male	29	21	4	group	no	9.90	FH9	Female	40	69	4	friend	no	n/a
GT10	Female	24	174	5	group	overlap	4.94	FH10	Female	18	599	2	group	no	1.30
GT11	Female	25	245	11	friend	no	4.16	FH11	Male	23	442	7	friend	no	n/a

friend and placing him/her into a group, creating new groups as needed, and continuing until all friends had been grouped. The second, which we refer to as "by group," consisted of participants creating a group, and then searching for friends or systematically looking through their list/stack of friends and placing them in that group if applicable. They continued this for each group they could come up with, until they decided they had finished. The strategy each participant used is displayed in Table 1. These strategies are similar to those used by Malone's file organizing participants [12].

Participants grouped their friends largely based on when or where they met them, or some specific context. Each of the groups was given a label during the grouping task (the interviewer prompted the participant for a group label or description if it was not provided). Each of these labels was coded into general groups, summarized in Table 2. The most used type of label was a description of friends from school/college, which 41 participants (89%) had. Of these, 35 participants (76%) specifically referred to a group of university friends. Family-related labels were also widely used with 29 participants (63%) creating one or more family groups. 14 (30%) participants created a group that specifically referred to a location other than a school or university, such as "Met in New York." Groups of people the participants could not identify, didn't know, or described as random were created by 14 participants, with an additional 10 participants creating groups called "other" or "miscellaneous."

Jones and O'Neill do not give a full breakdown of the labeled clusters in their paper, however they also reported commonly-created groups of close or best friends,

Table 2. Our participants created and named a total of 287 groups. Afterwards we classified the groups names into the above clusters, many of which contained subclusters.

Group Labels

Cluster	Total uses	% of users	Example labels
General friends	**85**	76%	
Location based	23	30%	Friends from China, France, Met in New York
Generic friends	16	35%	General, Good friends, Friends
Close friends	16	33%	Very good friends, Close set of friends, Best friends
Friends of friends	14	24%	Friends of my boyfriends, Friends' friends
College	**79**	76%	
General College	38	76%	College, Undergrad friends, University friends
Club or group	24	33%	Dancing club, Crew team, Band people, Fraternity, Freshman hall, Gaming/clan, Swimmers
Other education	**40**	70%	
High school	27	59%	HS friends, High school, Schoolmates
Grade school	9	11%	Middle school and below, Elementary, Junior high
Family	**35**	63%	Family, Cousins, Siblings, Relatives
Work	**15**	33%	Work, Work mates, People I've worked with, Office
Church	**5**	9%	Church, Bellefield church friends
Don't know	**17**	30%	People I hardly know, Never met in person, Unidentifiable, Random people who friended me

groups of weak ties such as acquaintances, "people I hardly know," or again "random" people. They continue to describe groups based on organizational boundaries (work, school) and specific times and places in their participants lives, all of which we also saw [8]. When comparing our clusters to those typically found in the social sciences research, for example [14], we see fewer family groups, which are almost mandatory, but we hypothesize that due to not everyone having family members present on Facebook, we do not see the same prevalence in our results. We do see that many of the common expected clusters are present, including work, religious affiliations, hobbies, school, childhood friends, and groups of friends met through other friends or relatives.

Our 46 participants created 287 groups (6.5 on average), ranging from four participants who made only 2 groups, up to GT3 who made 30 groups. The groups ranged in size from several single-member groups, to FB12's group of 394 high school friends. As expected, the more friends a participant had, the longer it took for them to group their friends.

6 Mechanisms Impact Groups

We observed that the mechanism that a user was given to group their friends impacted the resultant groups. While there were some expected differences (such as time taken to group), we also observed differences in the composition of the groups, specifically with friends who belonged to multiple groups, and also in the strategies used to create the groups.

6.1 Mechanisms

We briefly describe the trends we saw in each of the four mechanisms.

Card Sorting. All participants in the card-sorting group used an identical strategy of beginning at the top of the pile of cards and working their way through it in the order it was given to them (random). For each friend, they decided on a group to place them into, placed the card on the table in the area they had designated for that group, and continued. They created new groups only when they got to a card for a person who didn't fit into the groups that had already been created. Each card-sorting participant used the by-friend approach.

All twelve card-sorting participants placed each of their friends into exactly one group. One-third of our card sorting participants created single-person groups when there was not a better fitting group. Many participants seemed to hesitate with certain people, in some cases because they admitted they could not figure out who that person was, in other cases because they had a person who crossed multiple places, times, and contexts in their life. CS1, when asked if there was overlap between her high school and college groups said: "Yeah, a number of them. I put them in the high school group, because that is where I first met them." None of the participants in this condition placed similar people between two groups, or "duplicated" the card, although there were blank cards and pens within sight.

Grid Tagging. The grid-tagging condition did not lead to a common grouping strategy among participants. Some participants followed an approach similar to that used in the card-sorting condition, where each person was assigned a group before moving on to the next. Others attempted to list all of the groups they believed they would need before placing any friends, and others completed a page at a time, scanning the 20 friends per page before moving on, an artificial limit that our mechanism imposed.

Many participants did not place some of their friends into any group. 4 of our 11 participants in the grid-tagging condition had multiple friends who received more than one tag. While overlap did occur, many participants questioned if it was permissible. GT10 specifically asked, “Can you put someone into more than one category?”

File Hierarchy. While the main intent of the file-hierarchy condition was to provide participants with an easy and familiar method to create subgroups and organize friends into a hierarchical structure, we found participants did not take advantage of this. We did not explicitly say or demonstrate that subgoups were possible, as with all the other conditions. In some of the follow-up privacy scenarios participants did create subgroups, but even these events were extremely infrequent. More commonly participants would navigate into the groups, select or point at the users that were relevant to the given privacy scenario, and then navigate away. We did have some participants in the other conditions also try to use “subgroups” by creating groups that they later combined in response to certain scenarios.

Facebook Friend List Interface. The Facebook friend list interface was flexible enough that our participants used multiple strategies. Some participants created each of their ideal lists first before placing participants, others created lists as they went. Participants grouped both by group or by friend.

The edit friends page, where we had users begin, was not necessarily their preferred interface on Facebook. FB8 used the standard interface instead of the edit friends page, requiring her to recall and then search for friends, instead of paging through her entire network on the edit friends page (50 friends per page).

FB11 asked if duplicating members was allowed, and then added he had “probably accidentally done that,” as if it were a mistake (or possibly was prohibited in the study). Sometimes users were aware that there were overlapping groups. The number of friends placed in multiple groups may be particularly high for the Facebook interface because when adding a friend to a list, that friend’s current memberships are not shown, and a user may not remember where a given friend has been placed.

Although the Facebook interface was generally faster, some users spent considerable time with it. For example, FB12, with 889 friends, spent time reviewing many of his contacts’ profile pages to determine who they were. The Facebook interface was the only mechanism we tested that allowed participants to click through to view full profile pages, a frequently requested feature across the other conditions as many participants were unable to identify friends by just photograph and name.

6.2 Mechanism Comparisons

In general, the participants using the card-sorting method were much faster than the grid-tagging and file-hierarchy methods (average times in seconds per friend, card-sorting: 2.52, grid-tagging: 5.79, file-hierarchy: 5.32). The Facebook interface seems to be approximately as fast as card-sorting (average 3.11 seconds per friend), but many participants were not thorough and left many friends unplaced. Additionally, verifying all Facebook friends have all been placed is considerably more difficult using this mechanism than with the other three. Participants took on average 16.8 minutes to group their friends, yet, ranged in time from 9 seconds for CS11 to group her 6 friends into 3 groups, to the 89 minutes FB12 took to group his 889 friends. While these times give us an estimate of how long grouping takes, our participants were also told to think aloud, which can dramatically affect times. For statistical comparisons between mechanisms, non-think-aloud, larger-scale tests would need to be conducted.

Just under half of the participants in the grid tagging and the Facebook interfaces placed at least one friend in multiple groups, while only one file hierarchy participant duplicated a friend, and no one in the card sorting method did. Multiple placement, if used in future interfaces, may create rule conflicts, but it may remove decisions such as whether a friend really belongs in a high school or college group. To create an interface that aligns well with people's internal feelings about their friends, allowing multiple groups seems natural. However, even for the methods where group overlapping occurred, participants often asked the researchers if they could place friends into more than one group before they did it. So, although users may think of their friends as occurring in multiple groups, the interface must make it obvious that this is possible and encouraged.

7 One-Time Grouping Is Insufficient for Privacy

Immediately after the completion of the grouping task, we walked our participants through a series of Facebook information-sharing and related scenarios. We found that the groups that they had just created were wholly ineffective for meeting participants' privacy expectations in these scenarios. Many participants avoided using the groups they had created until prompted by the interviewer. Often when participants tried to use these groups they found them insufficient.

Even though participants had just spent an average of 17 minutes grouping their friends, nearly all of our participants attempted to work through the privacy scenarios without the use of the groups on the table or screen in front of them, opting instead to describe how an item of information would be shared with everyone or not placed on Facebook at all. Interviewer prompting was required to have the participants think in terms of the groups, and additional encouragement was needed to have the participants select certain groups they would or would not share items with.

For specific scenarios, such as hosting a party or event, many participants felt they would need to invite people on a case-by-case basis, and often did not consider the groups they created as a helpful starting point. CS4, when considering who to invite to a party said, "I would want a selective way of doing that, not just a group message." After CS4 eventually suggested a group that might work and the interviewer asked if everyone from that group could come, he added "No, no, not all of them."

On the other hand, in some scenarios participants stated they would provide the same access to multiple groups. FH2 responded to one question by allowing full access to four separate groups: “These four folders are OK, I have met them, talked with them.”

Additionally, for certain types of content sharing, such as religious and political views, participants needed to create completely new groups or tags to correctly identify a set of friends that they would share these more specific types of information with.

We set up the grouping task, as mentioned earlier, without explicitly mentioning privacy or sharing, just as Facebook’s friend lists currently are: under a vague Edit Friends settings page, without any explicit reason for their existence. However even doing this, we saw what we believe are the types of groups people will make even if prompted for privacy (Jones prompted users for privacy and the groups users created share many similarities to the groups in our study [8]). Until we have more information on what an ideal privacy/sharing solution is on Facebook, and a real-world study testing dissemination through groups over time, it is difficult to say which grouping-mechanism is the best, and therefore we attempted to gain an understanding of the difference between mechanisms, while also investigating how people already deal with privacy management on Facebook (without skewing the grouping task).

When participants did decide to indulge the interviewer and think about our specific privacy questions in terms of groups, the groups quickly proved insufficient. When a participant thoroughly reviewed a group, the group was frequently cleaved in two, or often required a few special cases to be removed before it would be suitable to share a specific photo gallery or vacation information with. When asked who GT4 would not have wanted to share her vacation information with she said, “Oh, my cousins.” When we asked if this was all or some of her cousins, she stated, “just a few” and proceeded to highlight which cousins should not have access to this information. A family group alone, which many users commonly create, could never have captured this specific event.

Additionally, participants often had difficulty relocating a particular participant, occasionally not being sure which group they placed them into a few minutes earlier. FH9 described not being able to locate a friend she had just placed: “This Teresa, oh wait, she might be in classmates. Or did I put her in acquaintances, maybe I put her in acquaintances....That’s weird. I don’t see her name.... I can’t find Teresa, this is so weird, it is like she disappeared.”

8 Privacy-Related Observations

Many participants appear to have developed strategies for protecting their privacy on Facebook without using groups.

For example, many of our participants seem to believe that regardless of the Facebook privacy settings, things they are putting on Facebook may be seen by anyone, and will thus only post information they consider public. People believed that they were able to use their discretion and not post any compromising pieces of information or anything that could be misconstrued as being compromising. When asked about possibly sensitive topics such as political and religious views, some participants stated that it didn't matter who saw these, and thus they were suitable for Facebook, while others said they would never put these topics on Facebook. This furthers the argument that users are unable or unwilling to utilize the full extent of Facebook's privacy settings.

Similar to previous studies [17], our participants varied in what they were comfortable posting. FH2 said of her vacation photographs: "Those pictures, they are just pictures, there is nothing about privacy." CS12 said of her cell phone number: "I let all my friends have it... if I am friends with them on Facebook I feel like they can talk to me anytime." This is further evidence that controls must be flexible enough to allow for users to specify desires that match their own personal preferences.

GT11 created two groups she explicitly had specific sharing requirements for, a group called "ex-boyfriends" and another created for a past stalker. Many other participants created groups of friends with whom they were not close, with labels such as "don't know very well" or "people I've grown apart from." In total, 11 of our participants made a group that we coded as "distant friends."

Even less closely related than distant friends, groups of people our participants could not identify, didn't know, or were described as "random people" were created by 14 of our subjects. As we discussed earlier, profile photographs significantly benefited users (names alone were often not enough). However, Facebook profile links were also desired (and used in the Facebook friend list condition), and without them some users admitted to being unable to identify a small percentage of their friends. GT10 said, "I don't know who this woman is—she just friended me, I replied. I can't remember who she was." FH2 stated of one friend: "I don't even know this person." GT4 said of her "random" group: "The random people I don't even remember where and who they are." CS1 had a group she described as: "These are people I cannot remember for the life of me."

Many participants simply believed they had too many friends. FB10 expressed this best, just a few minutes into the original grouping saying, "I should have cleaned out my friends list before I came here." GT3's comments were similar: "Wow I wish I didn't have so many friends." In general, participants had friends they didn't recognize, didn't know, and didn't want to expend the effort into grouping. Based on the post-survey 19 (41%) of our participants said they sometimes, often, or very often accept friend requests, "just to be nice."

In addition, participants were asked if they had any friends who sent too many application invites. If they answered yes, they were asked if they would consider placing these friends in a group to block them from sending application invites. What we found was that most would not do this and would continue to block the applications themselves. They didn't want to censor their friends in any way, as one user put it. This shows that users find applications themselves to be annoying, rather than the people who use the applications.

We also asked participants whether they had untagged themselves from photos. Most often, participants would tell us they did so because the picture was unflattering or perhaps inappropriate. This begs the question, what do people consider inappropriate enough that they would restrict others from seeing it? And more importantly, are there still people they would like to grant permission to see that content through privacy settings?

9 Recommendations

Our finding suggest some recommendations for designers of contact-grouping mechanisms in online social networks. These recommendations are not to be taken as a hard and fast list of requirements, but rather to provide a direction for improving interfaces to address real-world access-control needs. Our analysis has focused on privacy and security decisions and the actual content and sharing decisions that our participants make on Facebook.

Provide Ad-Hoc Grouping. One of our key findings above was that a priori user-created groups do not meet users' needs for use as access-control groups. While these groups provide a starting point for users to rationalize about their sharing decisions and understand the space of their online friend network, without allowing ad-hoc modifications they fail. We recommend that contact-grouping interfaces allow for usable, in-situ, temporary group modifications. This would allow a user preparing to post content they are aware is privacy-sensitive to select some number of pre-created groups, briefly review the summed memberships, and remove or add misaligned contacts.

Tagging and Filtering. In addition to ad-hoc grouping controls, it seems that because many groups are based on specific contexts, contacts overlapping between groups could be common. For example, Family, Church, and Hometown might all share certain members. To allow for this model without requiring the user to necessarily construct groups, the interface could allow searching based on information Facebook already has, such as friends with a similar current city, or contacts with the same high school or university listed. These filters could then be constructed into permanent groups (with user modifications made).

Additionally, we recommend that users be able to tag their friends with other freeform terms that could be further filtered or searched across. These tags could be similar in nature to Facebook's own friend relationships once requested upon friending; however, they should be private and not require external approval. Ideally these could also not only be added through a privacy/friends management interface, but also directly through a drop-down while viewing a friend's profile.

Assistance through Automation. Automated approaches for helping users create and modify groups could be enormously useful. In addition, approaches that would suggest access-control settings or remind users that they might want to restrict access to sensitive information could be helpful. These are areas where additional research is needed.

Holistic View. Previous work has suggested that access-control interfaces should contain a "holistic view" that allows a high-level, single-screen look at the entire policy [18]. Currently, the closest Facebook comes to that is the previously mentioned edit friends screen, which is still paginated and therefore does not actually provide a single complete view (nor does it scale particularly well). This is one of the greatest strengths of the physical card sorting mechanism: every participant was able to know that they had placed each of their friends, while understanding the relative sizes and relations of each group to the others. At the end of the grouping task, the table became their holistic grouping policy, which they could easily point to, gesture at, and modify throughout the privacy scenarios.

Iterative Management Tools. While our grouping task occurred for participants in a single session in our laboratory, the makeup of online social networks, and members' relations, change over time. While we have little understanding how users would like to and should modify their groups over time, we do know that the current interfaces do not assist this process. Indeed, nearly all (80%) of our subjects reported that they have either never or only once updated their friend lists.

Grouping is Not the Primary Task. While we believe all of these recommendations will lead toward better interface design, we want to emphasize that the act of friend grouping is not a primary task. Friend groups should be defined to assist users in managing their communications, content sharing, and general privacy on social networks. If the groups do not accomplish this task, then the privacy they seem to create is simply an illusion.

References

1. Acquisti, A., Gross, R.: Imagined communities: Awareness, information sharing, and privacy on the facebook. In: Privacy Enhancing Technology Symposium (2006)
2. Adams, P.: The real life social network v2, `http://www.slideshare.net/padday/the-real-life-social-network-v2`
3. Berteau, S.: Facebook's Misrepresentation of Beacon's Threat to Privacy: Tracking users who opt out or are not logged in (November 29, 2007), `http://community.ca.com/blogs/securityadvisor/archive/2007/11/29/facebook-s-misrepresentation-of-beacon-s-threat-to-privacy-tracking-users-who-opt-out-or-are-not-logged-in.aspx`
4. Binder, J., Howes, A., Sutcliffe, A.: Conflicting social spheres and experienced tension in social networking sites. In: CHI (2009)
5. Carr, A.: Facebook's New Groups, Dashboards, and Downloads Explained (October 6, 2010), `http://www.fastcompany.com/1693443/facebooks-big-announcements-dashboards-personal-information-downloads-friend-group-lists`
6. Davis, S., Gutwin, C.: Using relationship to control disclosure in awareness servers. In: Proceedings of Graphics Interface GI 2005, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, pp. 145–152. Canadian Human-Computer Communications Society (2005)
7. Gates, G.: Facebook Privacy: A Bewildering Tangle of Options (September 18 ,2010), `http://www.nytimes.com/interactive/2010/05/12/business/facebook-privacy.html`
8. Jones, S., O'Neil, E.: Feasibility of structural network clustering for group-based privacy control in social networks. In: SOUPS (2010)
9. Kwasnik, B.: How a personal document's intended use or purpose affects its classification in an office. In: Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1989, pp. 207–210. ACM, New York (1989)

10. Lampinen, A., Tamminen, S., Oulasvirta, A.: All my people right here, right now: Management of group co-presence on a social networking site. In: International Conference on Supporting Group Work (2009)
11. Lederer, S., Hong, I., Dey, K., Landay, A.: Personal privacy through understanding and action: five pitfalls for designers. Personal Ubiquitous Comput 8, 440–454 (2004)
12. Malone, T.W.: How do people organize their desks?: Implications for the design of office information systems. ACM Trans. Inf. Syst. 1, 99–112 (1983)
13. Marsden, R.: Facebook unveils its latest trick but the implications are worrying (September 18, 2010), http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-unveils-its-latest-trick-ndash-but-the-implications-are-worrying-2082708.html
14. Mccarty, C.: Structure in personal networks. Journal of Social Structure 3 (2002)
15. Mervis, C., Rosch, E.: Categorization of natural objects. Annual Review of Psychology 32, 89–115 (1981)
16. Ostrow, A.: FACEBOOK FIRED: 8% of US Companies Have Sacked Social Media Miscreants (August 10, 2009), http://mashable.com/2009/08/10/social-media-misuse/
17. Preibusch, S.: W3c workshop on privacy for advanced web apis. Journal of Interactive Advertising, 1213 (July 2010), http://www.w3.org/2010/api-privacy-ws/slides/preibusch.pdf
18. Reeder, R., Bauer, L., Cranor, L.F., Reiter, M., Bacon, K., How, K., Strong, H.: Expandable grids for visualizing and authoring computer security policies. In: CHI (2008)
19. Smart, D.: How do people limit their behaviour in response to their perception of potential social tension? A study of Facebook. Manchester Business School Bsc Interactive Systems Design (2009), http://intranet.cs.man.ac.uk/Intranetsubweb/library/3yrep/2009/7024300.pdf
20. Stangor, C., Lynch, L., Duan, C., Glass, B.: Categorization of individuals on the basis of multiple social features. Journal of Personality and Social Psychology 62(2), 207–218 (1992)

Privacy Concern and Trust in Using Social Network Sites: A Comparison between French and Chinese Users

Li Chen and Ho Keung Tsoi

Department of Computer Science, Hong Kong Baptist University
Hong Kong, China
{lichen,hktsoi}@comp.hkbu.edu.hk

Abstract. Though privacy and trust have been studied in the social network site (SNS), few have identified the relationships among users' privacy concern, trust and their actual usage behavior in SNS. Moreover, little attention has been paid to investigating the differences between users from different cultural contexts. In this paper, we have been engaged in addressing these concerns by surveying two typical user groups. The analysis of their answers showed that French and Chinese are not only significantly different regarding privacy and trust belief in SNS, but also act significantly different in disclosing personal information, posting messages, and developing new relationships. Furthermore, the effect of privacy concern and trust on users' visiting frequency and willingness to meet new people is also found different between the two groups.

Keywords: social network site, privacy concern, trust, usage behavior, cultural differences.

1 Introduction

With the increasing popularity of social network sites (SNS) worldwide, more and more attentions have been paid to understanding users' behavior when they are using the sites. One major concern is whether users are willing to disclose their personal info and tend to meet new people through the SNS. It is also interesting to know whether users actively post information regularly and whom they allow to access the info. These issues are in nature related to users' privacy concern and trust formation in the site. Indeed, trust has long been recognized as the primary factor leading to successful online transactions and interaction. Prior trust-related investigations in e-commerce sites showed that higher trust can prompt users to have more intention to purchase a product, and to return to the site for repeated uses [3]. On the other hand, it has been recognized that if a user worries his/her personal info will be used by the site for other purchases, s/he will unlikely trust the site and hence not be so active in disclosing her/his personal info. Thus, privacy concern seems to be a dominant factor in influencing users' trust building, and trust would further be the antecedent affecting users' actual behavior in the site.

However, though the concept "trust" has been extensively studied in SNS, the focus has been mainly on testing and enhancing trust relation between users [2], not

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 234–241, 2011.

on assessing user trust in the SNS itself. Moreover, though studies on users' motives and uses of SNS (e.g., Facebook) have been performed [6,7,8] and some even measured users' privacy perception in different SNS sites [1], few have experimentally identified the relationship among *privacy concern*, *trust*, and *users' actual behavior*. Another vacancy of related works is that few have assessed whether people from different nationalities would possess different privacy concern degrees, and whether their use of SNS would be also different as potentially impacted by their privacy belief and trust in the site.

Thus, in this paper, we report some interesting results from our on-going work in this direction. Specifically, we examined three aspects through the comparison of two countries' SNS users (56 French and 58 Chinese): *their privacy concern*, *overall trust*, and *actual usage behavior*. The reason of selecting the two countries is because they can respectively represent western and oriental cultures, so the discovered differences (if any) could in some sense reflect the two cultures' specialties. Indeed, according to Hofstede's cultural dimensions [4], western countries commonly have individualism and low context culture, whereas eastern countries have collectivism and high context culture. This hence suggests that users from eastern countries would be more addicted to use SNS because it creates a collaborative environment for them to enhance the social affiliation. If so, they would inherently have lower privacy concern and higher trust, in comparison to users from western countries. In the following, we will present experiment setup and results analysis with the objective of verifying this hypothesis.

2 Experiment Procedure and Instruments

The survey was launched in Dec. 2010 in the form of online questionnaire. It was announced through public channels in both China and France sides. In the survey, the subject was first asked to fill in her/his demographical information (including age, gender, profession, etc.). S/he was also asked to give one specific SNS that s/he has used most frequently in the past one year. Then, a set of questions was required to respond, which are mainly about their privacy concern when using internet and SNS, overall trust in SNS, privacy setting, types of personal info that they include in the profile, and whom (e.g., friends, family members, or anyone) they allow to access the profile. Another set of questions was requesting their actual uses, like usage period, daily usage frequency, amount of contacts, the information they have often posted, and so on. Some questions were multi-choices and some were 5-point Likert scale. The concrete questions are listed from Tables 1 to 6.

Till the end of Jan. 2011, 166 persons volunteered to fill in the questionnaire. We removed users who gave incomplete answers, ones whose indicated sites are not SNS, and ones who used SNS infrequently. Finally, we have 114 active SNS users for the analysis (active users mean that they have used SNS for minimal half of a year and used it on average 1 hour per day).

Among the 58 Chinese users, 39.66% are males and 60.34% are females. The average age is 22.9 (ranging from 18 to 49). Almost half of users are students in the university and the others work as engineer, researcher, etc. As for the 56 French users, the average age is 27.43 (ranging from 18 to 46), with 42.86% males and 57.14% females. Their professions include student, engineer, teacher, and officer. Regarding the sites that subjects have used most frequently in the past one year, all French subjects stated Facebook. Among Chinese users, RenRen.com was reported by most users (72.4%), followed by Facebook (17.2%), Kaixin001.com (3.4%) and other local variants. Implied from [5], these Chinese sites can be fairly comparable to Facebook, because they share the structural and functional similarities.

3 Results

3.1 Privacy Concern and Trust

To compare the answers from the two groups of subjects, we used the multivariate analysis of covariance (MANCOVA) because it can reveal whether the differences (if any) can be attributed to the nationality or others (e.g., gender). From Table 1, we can see that French respondents showed higher general privacy concern when using the Internet (*Q1*: $M_{\text{French}} = 3.81$ vs $M_{\text{Chinese}} = 3.48$, $p = 0.072$). The question about their SNS profile privacy setting indicated that they rated significantly higher (*Q2*: $M_{\text{French}} = 3.68$ vs $M_{\text{Chinese}} = 2.90$, $p < .001$). In addition, they felt less comfortable when giving personal information on SNS (*Q3*: $M_{\text{French}} = 2.02$ vs $M_{\text{Chinese}} = 2.69$, $p < .001$), less control in specifying and updating their profiles (*Q4*: $M_{\text{French}} = 2.97$ vs $M_{\text{Chinese}} = 3.77$, $p < .001$), and less agreed that their privacy is protected by the site (*Q5*: $M_{\text{French}} = 2.08$ vs $M_{\text{Chinese}} = 2.98$, $p < .001$). These differences are all significant. Regarding users' overall trust in SNS, it also showed that the trust level of French users is significantly lower than the level of Chinese (*Q6*: $M_{\text{French}} = 2.22$ vs $M_{\text{Chinese}} = 2.65$, $p < .01$).

In order to examine whether the cultural contexts significantly differentiate users' privacy and trust degrees, we used the six questions above as dependent variables, and

Table 1. MANCOVA results regarding privacy concern and trust: multivariate tests, adjusted means and standard errors

	Wilks' Λ		F		
Gender	.88		2.49^{*}		
Nationality	.64		9.87^{***}		
	France ($N = 56$)		China ($N = 58$)		F
	M	SE	M	SE	
Q1: How often do you concern about your privacy while you use the Internet?	3.81	.14	3.48	.13	3.29^{*}
Q2: How do you rate the privacy setting of your profile in SNS?	3.68	.13	2.90	.13	18.14^{***}
Q3: I feel comfortable giving personal information on SNS.	2.02	.13	2.69	.13	13.40^{***}
Q4: I feel having control in specifying and updating my profile in SNS.	2.97	.13	3.77	.12	21.01^{***}
Q5: I feel that the privacy of my personal information is protected by SNS.	2.08	.12	2.98	.12	27.41^{***}
Q6: My overall trust in SNS is high.	2.22	.11	2.65	.11	8.32^{**}

Note: All questions were responded on a 5-point Likert scale: Q1 from "very seldom" to "very often"; Q2 from "not private" to "very private"; Q3 – Q6 from "strongly disagree" to "strongly agree".

* Significant at $p < .1$; ** Significant at $p < .01$; *** Significant at $p < .001$ (the same notations used in the tables below).

gender as a covariate in MANCOVA analysis. The result showed that subjects with different nationalities did differ significantly (Wilks' Λ=.64, F=9. 87, $p < .001$), with the effect of gender being controlled (Wilks' Λ=.88, F=2.49, $p < .05$) (Table 1).

Moreover, Table 2 lists the multiple regression results that expose the causal relationship from privacy constructs (Q3 to Q5) to overall trust in SNS (Q6). For French users, the feeling that their privacy is protected by the site significantly leads to their trust in the site (β = .646, $p < .001$), whereas for Chinese, the comfortableness in giving personal information is a significant factor (β= .488, $p < .001$).

Table 2. Multiple regression analyses for predicting user trust

	France				China			
	β	t	R^2	F	β	t	R^2	F
"My overall trust in SNS is high"			.52	19.1^{***}			.22	5.07^{**}
Q3 (feel comfortable)	.157	1.46			.488	3.82^{***}		
Q4 (feel having control)	.014	.14			-.164	-1.28		
Q5 (feel privacy protected)	.646	6.18^{***}			.018	.15		

Note: *Q3 to Q5 are referred to Table 1*

At the next step, we analyzed the sorts of personal information that they included in their SNS profiles (as reported by them). This analysis revealed several significant differences between French and Chinese users (see Table 3). Specifically, Chinese subjects disclosed significantly more identifying information such as gender, email, phone number, and location. For example, 93.1% of Chinese subjects include their gender (vs. 62.5% French), 79.3% share email (vs. 51.8% French), 77.6% indicate location (vs. 46.4%), 55.2% give instant messenger account (vs. 17.9% French), 44.8% include biography info (vs. 10.7%), and 81.0% include education info (vs. 42.9%). On the other hand, more French users share single or married status than Chinese users (55.4% vs. 36.2%). As for other items, such as real name, nationality, self picture, birthday, they are popularly disclosed by both user groups (above 70%), and some items like home address are rarely included by both (below 10%) (due to space limit, these results are not listed in Table 3). In addition, Chinese users are more frequent in updating their profiles (i.e., "once or several times a week"), compared to French subjects who have updated profiles averagely "once or several times a month".

For the profile, we also asked the subject whom s/he allows to access it. Significant differences occur on the accessible right by friends, classmates, strangers and anyone (Table 4). In fact, more French subjects (96.4%) indicated that their profiles can be

Table 3. Personal information included in user profile

		Gender	Email	Phone number	Location	Single or married status	Messenger a/c	Biography	Education
French (N = 56)	Mean	62.5%	51.8%	5.4%	46.4%	**55.4%**	17.9%	10.7%	42.9%
Chinese (N = 58)	Mean	**93.1%**	**79.3%**	**31.0%**	**77.6%**	36.2%	**55.2%**	**44.8%**	**81.0%**
Sig.		$.003^{**}$	$.001^{**}$	$.000^{***}$	.000***	.041*	$.000^{***}$	$.000^{***}$	$.000^{***}$

Table 4. Other users' right of accessing the profile

Your profile can be accessed by..?		Friends	Friends of your friends	Colleagues	Family members	Classmates	Strangers	Anyone
French ($N = 56$)	Mean	**96.4%**	10.7%	21.4%	48.2%	19.6%	1.8%	3.6%
Chinese ($N = 58$)	Mean	86.2%	3.4%	29.3%	36.2%	**62.1%**	**10.3%**	**13.8%**
Sig.		.05*	.135	.338	.198	.000***	.05*	.05*

accessed by friends, than 86.2% Chinese users ($p < 0.05$). However, Chinese users are more open to allow classmates (62.1% vs. 19.6% French), strangers (10.3% vs. 1.8%) and even anyone (13.8% vs. 3.6%) to have the viewing right.

The above findings hence indicate that, when the users have less privacy concern and more trust in SNS (i.e., Chinese subjects), they are likely to disclose more in their profiles and enable more kinds of other users to access the personal info. We were hence motivated to further investigate whether users' actual usages of SNS were also similarly influenced.

3.2 Usage and New Relationship Development

As indicated in Section 2, we only analyzed answers from users who have used SNS for at least half of a year, because they can represent active users of SNS. The actual usage duration of Chinese group is 2.69 years on average (ranging from 1 to 4 years), and it is 2.7 in French group (also ranging from 1 to 4 years). Their overall visiting frequency is around "once or more than once per day" in both groups. As for the hours that they usually spent on SNS per day, 19.6% of French and 36.2% of Chinese subjects reported that they have used it for 1 to 2 hours, 12.5% of French and 13.8% of Chinese have spent 2 to 5 hours, 3.6% French have used it for above 5 hours, and 64.3% French and 50% Chinese have used it for less than 1 hour.

Table 5. Messages that users usually post to SNS

What do you usually post?		News	Hobbies	Personal life	Current event	Interesting observation	Thoughts/ opinions	Forward others' posts
French ($N = 56$)	Mean	**39.3%**	21.4%	32.1%	28.6%	30.4%	35.7%	16.1%
Chinese ($N = 58$)	Mean	24.1%	**41.4%**	**62.1%**	25.9%	**58.6%**	46.6%	**58.6%**
Sig.		.084*	.02*	.000***	.748	.002**	.243	.000***

Despite their similar visiting frequency, the number of contacts in their "friends" list was notably different. French users reported that they have on average 183.2 contacts (SD = 202.0), whereas Chinese have 299.1 contacts (SD = 342.4). In addition, the messages they have usually posted to SNS are also different (Table 5). Chinese users are more active in posting hobbies (41.4% against 21.4% French users), personal life (62.1 vs. 32.1%), interesting observation (58.6% vs. 30.4%), and forwarding others' posts (58.6% vs. 16.1%), while French are more active in posting news (39.3% vs. 24.1% Chinese). These differences all reach at significant levels.

In this survey, we also asked users whether they like to use SNS to meet new people, and whether they have contacted new people afterwards through other ways (e.g., telephone, email, instant messenger). In this regard, French rated significantly lower for using SNS to meet new people ($M_{\text{French}} = 1.73$ vs $M_{\text{Chinese}} = 2.71$, $F = 33.32$, $p < .001$). It is hence not surprising that most French subjects did not communicate with new people outside of the site (60.7% against 41.4% Chinese, $p < 0.05$; see Table 6). In comparison, Chinese users are more active in contacting new people off SNS by telephone and instant messenger.

Table 6. Other ways that users communicate with new people after they met her/him in SNS

Have you ever contacted new people after you met her/him through the SNS by the following ways?						
		Telephone	Face-to-face meeting	Instant messenger	Email	Did not communicate outside of the site
French ($N = 56$)	Mean	5.4%	17.9%	14.29%	19.6%	**60.7%**
Chinese ($N = 58$)	Mean	**15.5%**	17.2%	**50.0%**	31.0%	41.4%
Sig.		.07*	.931	.000***	.164	.038*

Table 7. Multiple regression analyses for predicting visiting frequency, amount of contacts and willingness to meet new people in SNS

	France				China			
	β	t	R^2	F	β	t	R^2	F
Visiting frequency			.259	4.447**			.089	1.297
Q3 (feel comfortable)	.178	1.291			.020	.130		
Q4 (feel having control)	.262	**2.009***			.285	**2.004***		
Q5 (feel privacy protected)	-.219	-1.263			.024	.176		
Q6 (overall trust)	.392	**2.242***			.002	.011		
Amount of contacts			.112	1.615			.013	.175
Q3 (feel comfortable)	.214	1.420			-.057	-.351		
Q4 (feel having control)	.142	.993			.054	.366		
Q5 (feel privacy protected)	.006	.029			.044	.309		
Q6 (overall trust)	.074	.387			.107	.693		
"I like to use SNS to meet new people"			.215	3.493*			.208	3.489*
Q3 (feel comfortable)	.231	1.629			.215	1.470		
Q4 (feel having control)	.042	.311			-.136	-1.029		
Q5 (feel privacy protected)	-.151	-.849			.317	**2.480***		
Q6 (overall trust)	.404	**2.244***			.134	.968		

Note: *Q3 to Q6 are referred to Table 1*

As shown in Section 3.1, various privacy factors were indicated to be driving forces for influencing French and Chinese users' trust in SNS. Based on it, we were interested in further elaborating the relationships between privacy/trust and users' actual usage behavior. For this purpose, users' overall visiting frequency, amount of contacts in their "friends" list, and willingness to meet new people were taken as dependent variables, to be predicted by privacy concern and trust (see Table 7).

In predicting the visiting frequency, control in specifying & updating profiles takes significantly positive effect in both groups ($\beta = .262$, $p = 0.05$ in French and $\beta = .285$, $p = 0.05$ in Chinese). Moreover, increased overall trust can also result in more visits among French subjects ($\beta = .392$, $p < .05$), which however is not significantly valid for Chinese. Regarding the amount of contacts, there is no significant predictor found in both groups. For the willingness to meet new people in SNS, the overall trust was shown to be a significant predictor among French ($\beta = .404$, $p < 0.05$), whereas for Chinese the feeling that their privacy is protected by SNS is the significant predictor ($\beta = .317$, $p < 0.05$).

4 Discussion and Conclusions

Thus, through this user survey, it is interesting to find that French and Chinese SNS users did possess significantly different privacy belief and trust. Specifically, French users are more concerned about their privacy while using the Internet. It hence seems being a natural consequence that their privacy setting in SNS is higher, they felt less comfortable in giving personal info, and they perceived the site less trustworthy to protect their privacy. Their overall trust in SNS is also significantly lower relative to Chinese users'. Such differences were further reflected in their profile disclosure and sharing. Chinese subjects disclosed more kinds of personal info in their profiles, such as gender, phone number, location, biography, education, etc., and allowed different types of users (e.g., including strangers) to access the profile. On the contrary, French users seem conservative in disclosing themselves, and they mainly allow friends and family members to view their profiles.

We further measured users' actual usage patterns. It showed that although both groups exhibit similar usage period and frequency, the ways they post messages and develop new relationships are significantly different. First of all, Chinese users have more contacts in their "friends" list. Secondly, they are more active in posting their hobbies, personal life, interesting observations, and forwarding others' posts, while French users are more active in posting news. Thirdly, Chinese users are more willing to meet new people through SNS and reported that they have contacted new people through other communication ways like telephone and instant messenger. In comparison, French users still behave conservative and most of them indicated that they did not communicate with new people outside of SNS.

The multivariate analysis of covariance (MANCOVA) further verifies that these phenomena can be attributed to the nationality difference. In addition, via the multiple regression analysis, we clarified the causal relation among these three factors: privacy concern, trust and users' actual uses. The analysis between privacy concern and trust first showed that for French users, the extent that they feel the site can protect their privacy is significantly correlated to whether they trust it, whereas for Chinese users, the significantly leading factor is whether they feel comfortable in providing personal info. Furthermore, some privacy factors and trust are found to significantly affect users' usage behavior. For instance, increased control in specifying and updating profiles can prompt both groups of users to visit SNS more often. Moreover, for

French users, increased trust also results in more visits and higher likelihood to meet new people. For Chinese, the perception with SNS's ability in protecting their privacy will likely result in their willingness to develop new relationships.

Thus, though the subjects' scale is limited in this survey, it revealed significant differences between Chinese and French SNS users from various aspects. We believe that the reason behind can be connected to how an individual internalizes the norms, rules and values in a society, which is in turn shaped by her/his cultural background. French, in a typical Western culture [4], are characterized by higher uncertainty avoidance, self-reliance, and emotional distance from in-groups. It might explain why they care more about their own privacy and attempt to keep distance from others (especially from new people). On the contrary, Chinese, as shaped by collectivistic culture which emphasizes family integrity, in-group membership and interdependence among people, should be in nature more willing to establish close social connections with others and hence be likely to disclose most about themselves. On the other hand, the country's privacy regulation may also influence users' privacy concern. That is, users from countries with omnibus privacy regulatory structure (e.g., France) would have higher levels of privacy concerns, than users from countries with sectoral or no privacy regulation (e.g., China) [9]. In the future, we will conduct more experiments to verify these explanations, and derive design guidelines for SNS so that it could better fulfill the expectations of their users with different cultural norms.

Acknowledgement. We thank Dr. Sylvain Castagnos and Prof. Anne Boyer from LORIA - Nancy University, France, for their help of recruiting French subjects. We are grateful to all survey participants for their time.

References

1. Dwyer, C., Hiltz, S.R., Passerini, K.: Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace. In: 13th Americas Conference on Information Systems (2007)
2. Golbeck, J.: Trust and Nuanced Profile Similarity in Online Social Networks. ACM Trans. Web 3(4) (2009)
3. Grabner-Kräuter, S., Kaluscha, E.A.: Empirical Research in On-line Trust: a Review and Critical Assessment. International Journal of Human-Computer Studies 58 (2003)
4. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations. Sage Publications, Thousand Oaks (2001)
5. Jiang, J., Wilson, C., Wang, X., Huang, P., Sha, W., Dai, Y., Zhao, B.Y.: Understanding Latent Interactions in Online Social Networks. In: Annual Conference on Internet Measurement, pp. 369–382 (2010)
6. Joinson, A.N.: Looking At, looking Up or Keeping Up with People?: Motives and Use of Facebook. In: ACM SIGCHI 2008, pp. 1027–1036 (2008)
7. Kim, Y., Sohn, D., Choi, S.M.: Cultural Difference in Motivations for Using Social Network Sites: A Comparative Study of American and Korean College Students. Computers in Human Behavior 27, 365–372 (2011)
8. Lampe, C., Ellison, N.B., Steinfield, C.: Changes in Use and Perception of Facebook. In: ACM Conference on Computer Supported Cooperative Work, pp. 721–730 (2008)
9. Privacy International Organization, https://www.privacyinternational.org

Privacy Concerns in Enterprise Social Travel: Attitudes and Actions

Netta Aizenbud-Reshef, Artem Barger, Yael Dubinsky,
Ido Guy, and Shiri Kremer-Davidson

IBM Haifa Research Lab
Haifa, Israel
{neta,bartem,dubinsky,ido,shiri}@il.ibm.com

Abstract. Privacy in travel refers to the way people manage their personal travel information and their willingness to share this information with others. Travel privacy concerns affect the amount and type of information people are willing to share within social networks, before or after their trip. Our study focuses on privacy and sharing concerns regarding business travel in an enterprise setting. We briefly present Voyage, an enterprise social travel application, and describe a study based on qualitative and quantitative data that inspects privacy and sharing concerns in business travels. We found that most employees are theoretically willing to share their business travel plans, but in practice they share less than expected. Further, most employees are less concerned to share their past travel information than their future plans. Based on our study, we suggest guidelines for the development of location based and enterprise travel applications.

Keywords: Privacy, business travel, information sharing, social activity.

1 Introduction

Sharing travel information is becoming very popular on the web. People typically share their opinions and experiences about places they visited, hotels they stayed in, and services they used; this is done on web sites such as TripAdvisor.com, Hotels.com, and yelp.com. Recently, people have also begun sharing their itinerary details with a select list of friends (defined on the site), through services such as TripIt.com and Dopplr.com. The benefits of sharing travel plans are clear as people can learn from other individuals' experiences and make better informed decisions [6, 11]. Ultimately, these decisions can affect their itinerary, cost, and level of satisfaction or enjoyment. Sharing travel plans also facilitates better coordination with other people and increases awareness of those who plan to be at the same place.

Privacy in the context of social applications that support information sharing has already been studied. Ahern et al. [1] examined privacy decisions in mobile and online photo sharing and identified relationships between the location of photo captures and photo privacy settings. Similarly, other studies [2, 3, 8] explored whether and what users are willing to disclose about their location. For instance, it was found

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 242–249, 2011.

that participants typically disclosed the most useful details about their location or did not disclose their location at all [2]. General studies of the impact of privacy concerns in social communities concluded that trust within the community and sharing norms affect information sharing [5, 9]. Because an enterprise setting can be readily perceived as a community, similar behavior is expected for business travel as well.

There is often a gap between what people want to get from other's experience and their willingness to contribute [7]. In an enterprise setting, this gap seems to be less pronounced. Inside the enterprise, people are willing to share more information than on a public site, because they are less concerned about the ramifications of sharing personally identifying information [4].

Our motivation is to study privacy concerns in enterprise social travel, the willingness of employees to share and consume travel data, and the main factors affecting their decision to share. As many business processes are turning social these days, we believe that enterprise social travel is one example of an enterprise business process that is transforming to be more social than before. The implications presented in this paper are relevant for such business processes as well.

The paper is organized as follows: first we present Voyage, a social enterprise travel application. Then we describe the research setting, including the methods we used for gathering data. Following this, we analyze the results and conclude with a summary and description of future work.

2 Voyage – Travel Meets Social

We experiment with Voyage, an enterprise travel reservation tool that aims at making the travel reservation process more social and collaborative. It extends a common enterprise web-based travel reservation application (such as Expedia.com, Orbitz.com, etc.) and adds various functional and social features, such as:

Learn from past travels – past reservations are aggregated to show statistical measures, such as the percentage of people renting a car in a specific city, the most frequently reserved hotel, or the flight route usually taken when traveling from one place to another.

Share feedback – users can provide feedback (comments and ratings) and see information entered by others for any travel-related item in their itinerary, such as hotels, flights, airlines, car agencies, airports, places, etc.

Share travel plans – users can share their itineraries and select the sharing level for each itinerary by setting its privacy level. Itineraries defined as 'Private' are not exposed to anyone but the user; those defined as 'Public' are visible to everyone. 'Confidant' itineraries are exposed only to a chosen set of users. This enables users to collaborate on travel planning, see who is going to visit the same place, see what reservations were made by others, and so forth. Users can also define a default privacy level to be used for all of their travel plans. The default privacy level set by the system is Public, in order to encourage sharing in the enterprise.

Travel information in one place – travel-related information is aggregated from both publicly available sources (such as weather sites and sites providing electricity information) and enterprise-specific sources (such as a meal limit database or a list of allowed hotels) and presented in one place, instead of sending users around the Internet and company intranet to collect it.

The figures below provide some examples of how the above features are presented on the Voyage site. Fig 1 shows the outcome of a hotel search within a specific period of time. It includes one hotel from the returned result list. Along with general hotel details such as name, address, and price, there are several social features: on the right, enclosed in a red rectangle, are three cues showing the number of comments for this hotel (1), the percentage of reservations made to this hotel out of all hotel reservations in the city (23%), and the number of people that already reserved this hotel during part or all of the selected time period (0). In addition, the stars on the left show the hotel average rating (3 of 5); more detailed ratings and comments are presented when opening the corresponding sections marked as (b).

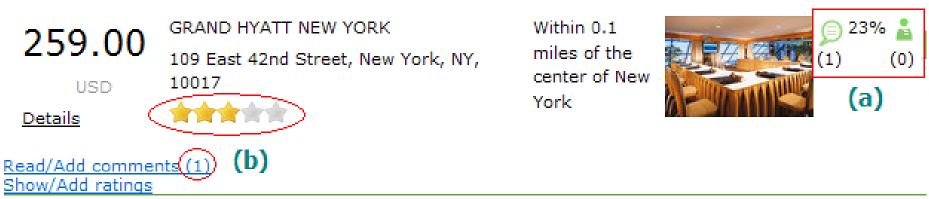


Fig. 1. Hotel options augmented with social cues

Fig 2 shows information on one itinerary (name, dates, and reservation types) and the people with whom it is shared.

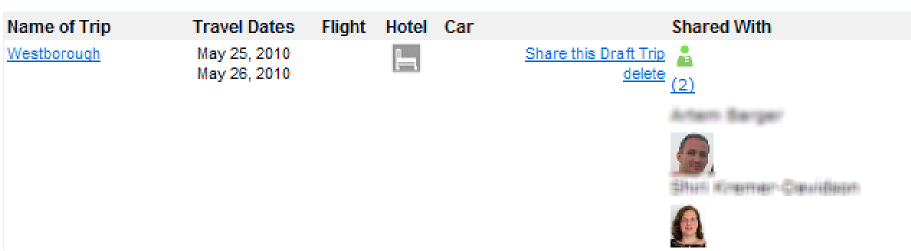


Fig. 2. Sharing travel plans

3 Travel Sharing and Privacy Concerns

Our research was conducted within a large distributed organization, with data collected from employees in Asia, Europe, the Middle East, and North America. We used four different methods to gather data about people's travel privacy/sharing concerns in order to increase the findings' trustworthiness: (1) Before Voyage was implemented, an initial user survey was sent to frequent travelers, querying their willingness to share their travel plans and the benefits they envisioned in knowing about others' past reservations and future plans. The survey included 15 multiple-choice questions and was answered by 1431 employees; (2) After Voyage was deployed and employees started to use it for their travel planning, we conducted a usage survey asking them about the actions they took while planning. This survey included 16 multiple-choice questions in total, but was adjusted to align with the selections made by the employee in the tool, for example, if the employee defined the privacy level to be 'Private', a question about the concerns that led him to this decision was included in the survey. Out of 92 employees who made their travel planning through Voyage, 54 answered the survey; (3) We inspected the tool logs to determine the actual sharing decisions users made throughout the process. Logs include information on the user actions, specific parameters for each action, the user name, and a time and date stamp; (4) Finally, we conducted semi-structured qualitative interviews with 13 employees, exploring their travel privacy/sharing concerns.

The analysis combines qualitative and quantitative results, which involve comparison of triangulated data sources and thick narrative to strengthen each finding [10]. We divide the results into three aspects. The first aspect deals with concerns and attitudes towards sharing travel information among employees. The second aspect explores the results related to sharing resources, rather than information. We examine sharing controls as the third aspect.

Information Sharing. From the initial survey we learned that employees value information about past travels of other employees: 81% of employees who answered the initial survey indicated it is useful for them to know where other people from the enterprise usually stay; 69% think it is useful to know who recently visited their travel destination; and 69% value ratings and comments input by other employees. From interviews, we learned that before they travel, employees often consult with their colleagues about hotels, directions, leisure activities, and even who they can get together with. One of our interviewees noted: *"If it is a place I rarely go or have never been before, I will ask others that have been there or the local people about hotel and means of transportation."*

Employees also found value in information on others' future plans, e.g., for coordination: in the initial survey 70% stated it is beneficial to know what flight a colleague is going to take; 70% want to know who else is attending a meeting/conference/event, etc. On average, around 70% of employees find this type of information very useful, while only 2% indicated it as not useful at all.

The benefits from knowing what other employees do with regard to travel motivate employees to share their travel plans in the majority of cases. Table 1 summarizes data that was gathered in the initial survey and depicts the willingness of employees to share their travel information with anyone in the enterprise (Anyone), their social network (SN), specific people (Specific), or not at all (No one). It is evident from the data in the table that employees are willing to share their travel information with different groups of employees in 96% of the cases.

In the interviews employees raised different reasons for sharing. Some perceive business travel as belonging to the enterprise, as one of the respondents indicated: *"What I do at work belongs to work; I get a salary for that, thus my business travel also belongs to the company".* Some see the benefits of coordinating plans: *"Maybe someone else is also traveling there in the same week. It's good just to know about it; maybe we won't share anything, like a taxi, but we could go out to eat together."*

Table 1. Sharing travel plans

With / Share	Anyone	SN	Specific	No one
Future plans				
Flight	32%	35%	28%	5%
Hotel	34%	39%	24%	4%
City	29%	37%	28%	6%
Past travels				
Flight	45%	34%	18%	3%
Hotel	48%	37%	13%	3%
City	47%	35%	14%	3%

Another noted "*if several of us travel together, it's important that we coordinate: flights and landing times, car rental pooling and selection of car size, hotels to stay in, and so on.*"

Focusing on the 'Anyone' column of the table, Fig. 3 highlights the difference in willingness to share future plans vs. past travels (on average 47% vs. 32%). We confirmed the significance of this difference using One-Way ANOVA ($p<0.01$). From interviews we learned that employees are mostly concerned about security issues, for example, that people will know where they are and when they are away from home. As one of the interviewees noted, "*The enterprise is very big, complex, and diverse; you can't trust all the people*". This concern relates to future plans only and may explain the difference in willingness to share past travels vs. future plans, that is noted before. A somewhat similar concern was expressed in connection with revealing personal information, such as a credit card number: "*What if a crook from the enterprise calls the hotel and asks for a copy of the receipt, and my personal information is there (e.g., who I called, my credit card number, etc.)?*"

Some interviewees noted that sharing information about their travel also reveals information about other people involved in the trip, such as accompanying people or the customer they are visiting, without checking the concerns these people might have: *"Sometimes the confidentiality of the project I am working on restricts the people I can meet."*

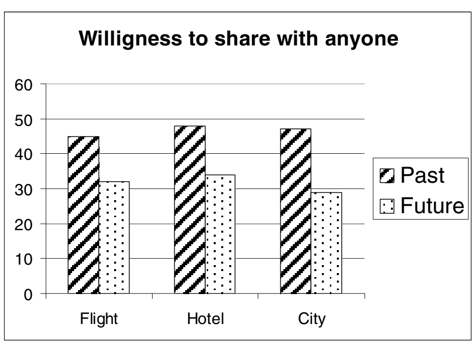


Fig. 3. Past vs. future sharing

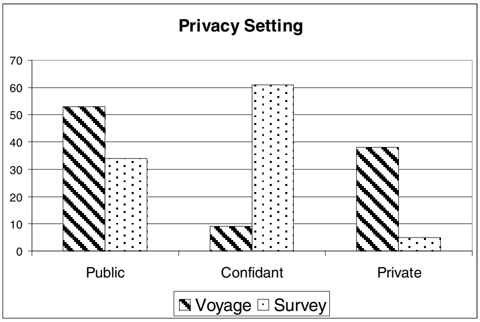


Fig. 4. Privacy setting

Inspecting actual privacy settings of travel reservations made in Voyage, we found that employees were less open to sharing than indicated in the initial survey. These gaps are depicted in Fig. 4. Fewer people than expected defined their trip privacy as 'Confidants'. It might resulted from them not being sure what is the meaning of this privacy level, being presented with the 'Confidants' concept at the first time. It could also result from laziness, since they had to manually define the list of confidants. The big gap between the percentage of people who did not want to share their travel plans according to the survey (5%) and those that defined their trip to be private (38%) is harder to explain. We can only hypothesize that when asked in a survey about attitudes, people tend to be more open than when they are confronted with the actual travel reservation and must make a real decision. We observed a similar phenomenon in one of the interviews. The interviewee first declared that he is willing to share all his travel plans with anyone in the enterprise. Thinking it over and raising some of the above concerns, he withdrew his original statement and changed it to sharing with only specific people.

Nevertheless, employees were willing to expose more about their travel details after they returned from a trip. In the usage survey, 60% of Voyage users who defined their trip as private, agreed to change it to public after they returned from the trip. This reinforces our findings about the difference between sharing past travels and future plans.

Resource Sharing. Sharing travel resources, such as rented cars or hotels, is treated differently by employees than sharing travel information. When asked about sharing a car rental or a taxi in the initial survey, 35% of the employees indicated they would share with anyone in the enterprise; 45% would share only with known people; and the other 20% are not willing to share at all. We observe that employees are more reluctant to share resources (20%) than to share travel plans (4%). It is very likely that this results from the feeling that sharing data is less personal than physically being at the same place with a stranger. Employees also feel that it introduces another difficulty to their travel planning: "*It's a burden when you have to coordinate a car rental between two people just because they are going to be in the same place"*. They are willing to share a car if it does not interfere with their plans: "*My experience of sharing a taxi has not always been positive, because sometimes you have to wait longer than if you had been alone."*

Sharing Controls. In the initial survey, we queried employees about the importance of controlling over what they are sharing. Table 2 presents the distribution of answers regarding both the control over sharing defaults and the ability to adjust sharing preferences on an item-by-item basis.

Table 2. Sharing controls

	Critical	Important	Slightly important	Not important
Control defaults	16%	45%	31%	8%
Control item-by-item	15%	46%	32%	7%

As is evident from the table, employees are keen to control the travel information they share. Currently, Voyage supports privacy control over a full itinerary and not for each itinerary item separately. This might explain why more employees than expected chose to define their travel plans as private.

The desire of employees to control the sharing defaults in the system is also clear from the table. Inspecting the behavior of Voyage users, only 30% opted to set their default privacy level. Out of those that cared do so, 2.5% set it to 'Private', 5% set it to 'Confidants', and the rest (92.5), although opted to set their default privacy level, selected to leave it 'Public' as it was defined by default. These findings strengthen previous research works that show that users rely on the system defaults. In our case, the default setting plays even a smaller role, as users can set the privacy level for each trip separately.

4 Summary

In this paper, we study travel privacy concerns and the willingness of employees to share their travel plans with others in an enterprise setting. We found that employees value other people's travel information. Although they theoretically agree to share their own travel information, in practice they share less than stated. Employees raised various sharing concerns related to privacy, security, and project confidentiality. We found that there is a difference between sharing past travels and future plans – employees are willing to share more of their past travels. To the best of our knowledge, this is the first study to point out this distinction. Future research should examine whether it applies to other business processes as well.

Based on the finding of this study, we suggest two high-level development guidelines for enterprise travel applications. First, employees stated that they would like to have the option to control the level of sharing allowed by the reservation application, and to some level used the options provided by Voyage. As only parts of the itinerary might be confidential, users should be provided with privacy settings to allow them to finely control what information is shared and with whom. Second, employees expressed their will to relax the privacy level of their trip after some time has passed, as the sensitivity of exposing to a larger audience fades over time. We believe that this approach applies not only to travel plans, but for any type of media in general. To encourage sharing, we suggest the application would actively offer users to relax the privacy settings of their media after some time has gone by since posting it (for example, issue an alert after a week).

In future work, we intend to further explore how sharing attitudes and actions change over time or in different situations, and the differences in privacy concerns and behavior between business travel and personal travel. We also want to research privacy aspects in other types of business processes and investigate whether the implications, as presented in this paper, are similar.

References

1. Ahern, S., Eckles, D., Good, D., King, S., Naaman, M., Nair, R.: Over-exposed?: Privacy Patterns and Considerations in Online and Mobile Photo Sharing. In: Proc. ACM Conference on Human Factors in Computing Systems, CHI (2007)
2. Consolvo, S., Smith, I.E., Matthews, T., Lamarca, A., Tabert, J., Powledge, P.: Location Disclosure to Social Relations: Why, When, & What People Want to Share. Proc. ACM Conference on Human Factors in Computing Systems, CHI (2005)
3. Church, K., Neumann, J., Cherubini, M., Oliver, N.: SocialSearchBrowser: A Novel Mobile Search and Information Discovery tool. In: Proc. IUI, pp. 101–110 (2010)
4. DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B.: Caring, Climbing, and Campaigning: Motivations for Social Networking at Work. In: Proc. CSCW (2008)
5. Dwyer, C., Hiltz, S., Passerini, K.: Trust and Privacy Concern within Social Networking Sites: A Comparison of Facebook and MySpace. In: Proc. AMCIS (2007)
6. Gretzel, U., Yoo, K.H.: Use and Impact of Online Travel Reviews. In: Proc. International Conference on Information and Communication Technologies in Tourism, pp. 35–46 (2008)
7. Gretzel, U., Yoo, K.H.: Influence of personality on travel-related consumer generated media creation. Computers in Human Behavior (2010)
8. Gross, R., Acquisti, A.: Information Revelation and Privacy in Online Social Networks. In: Workshop on Privacy in the Electronic Society. ACM Press, New York (2005)
9. Nov, O., Wattal, S.: Social Computing Privacy Concerns: Antecedents & Effects. In: Proc. ACM Conference on Human Factors in Computing Systems, CHI, pp. 333–336 (2009)
10. Patton, M.Q.: Qualitative Research & Evaluation Methods, 3rd edn. Sage, Thousand Oaks (2002)
11. Surowiecki, J.: The Wisdom of Crowds. Anchor (2005)

Online Games and Family Ties: Influences of Social Networking Game on Family Relationship

Jing Wen[1,2], Yong Ming Kow[1], and Yunan Chen[1]

[1] Department of Informatics
University of California, Irvine
Irvine, CA 92697-3440, USA
{jwen2,ykow}@uci.edu, yunanc@ics.uci.edu
[2] The State Key Lab of Software Engineering
Wuhan University
Wuhan, 430072, China

Abstract. How do online games impact intergenerational family relationships? To answer this question, we investigated QQ Farm, the most popular online game available on a popular Chinese social networking site. We conducted observations and semi-structured interviews with sixteen pairs of Chinese parents and their adult children. Of the sixteen pairs, seven lived locally; nine pairs lived remotely. The findings of this study suggest that online games provide common conversational topics among local family members and enrich family time. Online games with well-designed shared virtual spaces can also help enhance remote families' awareness of each other's real-life activity. For future intergenerational games, we recommend a game design that requires minimal effort, has a 'healthy' theme, and provides features, such as an activity log and message board, which help maintain game based family communications.

Keywords: Social Network Games, Parents-children Relationship, Domestic Play, Generational Gap.

1 Introduction

Online games allow hundreds of geographically dispersed players to perform shared activities in the same virtual world [1]. Players come to online gaming sites not only to play, but to socialize [2]. Formerly perceived as merely "games," online game has become a new environment for individuals to mediate social relationships [3][4][5][6]. While previous studies have focused on how games can strengthen or form new social relationships among players, game play among individuals with existing social relationships, such as intergenerational family members, have seldom been reported in previous literature [3][4][6][7]. In this article, we ask whether online games can be beneficial in reinforcing the social relationships between intergenerational family members with substantial age and value differences.

Intergenerational relationships are of special interest to China, a country where the traditional family structure is rapidly undergoing transformation. The traditional

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 250–264, 2011.

Chinese family is a co-residence among three generations of family members. However, co-residence among intergenerational family is becoming increasingly rare as young adults migrate to other regions seeking better employment opportunities [8]. Studies have shown that older parents continue to desire traditional family relations, such as the assistance of their children with housework, being able to care for their grandchildren [11], and keeping mutually close contact with their working children [10]. This raises the question how we can continue to maintain these traditional practices given the modern context where intergenerational relationships are practiced at a distance, rather than in the same household [11].

Given the widespread adoption of online gaming in contemporary Chinese society, we are interested in seeing if online games can serve as a new platform for mediating intergenerational family interactions. We examine QQ Farm, a popular multiplayer online game in China that at its height had 6.8 million users logon simultaneously. QQ Farm has a player pool that includes a large number of older parents and their adult children [12][13]. For our study, we conducted semi-structured interviews to examine how playing QQ Farm impacts the relationships among family members who live locally and those who live remotely. Sixteen pairs of QQ Farm players participated in the study, with a Chinese parent and one adult child in each pair. These pairs include any combination of father or mother with a son or daughter.

The results of the study have shown that social networking games such as QQ Farm can improve the communication among family members locally, and enhance the connectedness between family members remotely. In allowing online players to co-construct a shared social space from which family members can extrapolate into awareness of real world activities, online games can strengthen existing family ties. At the end of this paper, we will introduce opportunities for designing games that support intergenerational play.

2 Related Work

Recently, there is an increasing interest in studying social interactions in online game play in the HCI community [4]. Studies have found that social interactions in online games can improve social relationships among players [3] [6] [7], since online games allow large number of players to interact and accomplish shared goals together. Ducheneaut [6] highlights the designing strategy for creating a sense of social presence in online games. Nardi [7] argues that social interactions, particularly those that accomplished shared tasks, are especially desirable within a game environment.

Online game play is increasingly popular among the Chinese. Various reports and news articles often indicate that Chinese parents worry about the negative influence of video games on their children, in particular, the amount of time consumed in game play e.g. [14]. Interestingly, these older adults themselves often engage in and benefit from social gaming activities, such as chess, card, and mahjong [15]. For Chinese parents, social games, when played on computers, often have negative associations, such as game addiction and distraction from school work [14]. Lindtner *et.al.* [16] [17] found that many urban young Chinese players establish close social ties through game play in Internet cafés [16] and in game-related online communities [17]. These studies show that online gaming has become a meaningful way for youth to get to know friends and help each other in their offline lives.

While some digital games have been developed for families, these games have restrictive game spaces that only allow a handful of players within the same household. For instance, *Age Invaders* is an intergenerational game that is designed to bridge family ties by enabling older adults to engage in both in-person and remote play with their children and grandchildren [18]. Others games, such as the *Magic Box* [19], are developed to encourage the exchange of thought-provoking items between grandparents and grandchildren living remotely, indicating that intergenerational play can be associated with family history, culture, and fun. Voida et al. [20] studied console games and found mutual conversion between the individual-play and family activities during game playing, showing how players might respond differently according to the social relationship they have with other players.

In addition to games, intergenerational relationships can be promoted by other types of digital devices [21] [22]. *Digital Family Portrait* is designed to transmit iconic projections of older adults' daily activities to their remote family members. *Digital Family Portrait* allows family members to know whether the older adults are active and well on a daily basis [23]. Similarly, *Family Window (FW)* is designed as an always-on video media space to connect family members residing in two households through video cameras [24]. Using *Family Window*, family members can share their daily activities and increase their feelings of connectedness with each other. Devices such as *SPARCS* [25] and *Wayve* [26] are also designed to encourage information and photo sharing among family members.

Online games are becoming increasingly popular among the Chinese. Their penetration into intergenerational social space thus providing and the social-cultural landscape continues to transform Chinese families, understanding the possible impact of online gaming in family relationships could be vital. The literature suggests that games and other form of digital devices can be used to leverage intergenerational family interactions; nevertheless, there has not yet a study on the use of online game by intergenerational family members. In what follows, we report the findings of our study on QQ Farm and discuss possible implications for future game design.

3 The Game - QQ Farm

QQ Farm is an online game developed by Tencent, Inc. in 2009, and residing in Tencent's social networking site QZone. Social Networking Sites [27] are web-based services that allow users to communicate and construct social connections with other users. And Tencent is the provider of many popular Internet services including the QQ Instant Messenger. On March 31, 2010, it was reported that there were 568.6 million active user accounts on all QQ service platforms including QQ Farm, QZone, and QQ Instant Messenger.

On QQ Farm, players act as farm owners and manage their own farms. Figure 1 shows what QQ Farm looks like in the game environment. In addition to common farming activities such as cultivating, irrigating, and harvesting crops as real farmers do, players can also visit farms owned by their SNS friends. On the farms owned by their friends, players can perform limited actions—some helpful, some mischievous. Helpful actions include weeding and irrigating, whereas mischievous actions include stealing the other players' mature crops.

Fig. 1. Screenshot of the first author's QQ Farm

Each activity a player performs on her own farm is rewarded by experience points or, with virtual gold coins through the selling of mature crops. Accumulated experience points can raise the player's level in the game. The higher the player's game level is, the more privileges she has to expand her farming field**.** Similarly, the more coins the player has, the more "money" she has to purchase expensive crops and fields. Weeding and irrigating for other players are considered helpful in the gaming environment since it helps the other farm owners obtain additional experience points or coins. On the contrary, stealing other players' mature crops is considered a mischievous action since the owner is no longer able to gain virtual gold coins by selling her stolen crops.

In addition to game play itself, each player has access to a personal message board associated with their farm [Figure 2]. The personal message board shows log information about one's farm, including both helpful and mischievous actions. As shown on Figure 2, a player can see who came to visit their farm and when, and what they did to it. Players can also personalize their space through naming and decorating their own farms.

After its launch in 2009, QQ Farm became immediately popular among Chinese players. In general, people were intrigued by the fact that QQ Farm had attracted a large population of older adult players, and the family play of QQ Farm was soon reported by various newspapers and media sources as a new Chinese social phenomenon [12][13]. For instance, Net Ease 163.com, one of major news portal sites in China, published a story about a mother using QQ Farm as an education resource to help her child understand botany [12]. Another article described how a seventy-year old grandma played QQ Farm with her granddaughter remotely [13]. Following these reported stories, we conducted a qualitative study to examine the influence of QQ Farm game play on family relationships.

Fig. 2. Screenshot of a personal message board in QQ Farm

4 Methodology

We conducted a qualitative study from March 2010 to November 2010. We examine the impact of online gaming on intergenerational family relationships. As described in the literature review section, tensions arise when adult children move out of a home where the parents still hold traditional Chinese family values. We studied only parents and adult children because of this, excluding children below the age of 18, who were more likely to live at home.

4.1 Participants

Sixteen pairs of parent-adult children dyads participated in our. Participants were recruited using snowball sampling method. The first two participants were recruited from the university where the authors are affiliated, while the rest of the participants were referred by participants interviewed previously. The snowball sampling method appears to be most appropriate way to recruit family members who play online games due to the social property of QQ Farm.

Of the sixteen parent-adult children dyads we studied, seven pairs live in the same households, or in nearby neighborhoods in the Hubei and Sichuan provinces. We refer to these pairs as "local families" since they were able to meet in person frequently. The other nine pairs lived in different cities. While all the parents we interviewed live in China, their adult children lived remotely in America, Japan, Australia or Denmark. Table 1 summarizes the participants of the study based on their family relationship and local/remote living conditions. The age of parents ranges from 47 to 64. The age of adult children ranges from 18 to 36.

These sixteen pairs of participants contain three conditions: both members of the dyad are active players (9 pairs); only one member of the dyad (either the parent or the child) is a player (4 pairs); and both members of the dyad are non-players

Table 1. The parents-children relationships

Parents-children relationships	Number of pairs	Local families	Remote families
Father-Daughter	4	2	2
Mother-Daughter	4	2	2
Father-Son	3	1	2
Mother-Son	5	2	3
Total	16	7	9

(3 pairs). The non-playing participants served as a baseline for us to understand the behaviors of the "dual-play" pairs. By comparing intergenerational communication patterns between "dual-play" pairs with those having one or more non-playing participants, we found that online games could indeed improve intergenerational communications. It is worth mentioning that among the nine pairs of dual-play pairs, four pairs were local and the other five were remote families.

4.2 Data Collection

Human subject study approval was obtained from the university where the researchers are affiliated prior to data collection. We acquired verbal consent from every participant and recorded only their screen names used in the games. Nevertheless, all names reported in this paper are pseudonyms.

In the study, we first observed the game play behaviors of the six pairs of dual-play families at QQ Farm for two weeks. Observing game play helped us to understand the basic behaviors participants demonstrated online. We collected the game logs from the personal message board of each participant. These logs recorded all the players' interactions in the game space. The observation data were collected twice a day, once in the morning and once at night. 4132 messages were gathered during the two weeks. Excluding messages relaying the game play records between the farm owner and their other friends, 319 messages were found relating to interactions between family members, accounting for 8% of the total logged messages.

Based on the family play behaviors observed online, we developed semi-structured interview questions. The interview study was extended to single-play or non-play pairs in order to better understand the impact of game playing on family relationships. The interview questions were centered on game play and family relationships. Specifically, we asked:

-How often do participants play QQ Farm?
-How does participant play with their family members?
-How does the game impact their family relationships?

When possible, we interviewed subjects in-person or by telephone. When in-person or telephone interviews were impossible, we conducted instant messaging interviews. Of the thirty-two subjects, five were interviewed in person, twenty-three were interviewed by telephone, and the remaining four were interviewed through instant messaging. The length of the interviews ranged from 30 to 60 minutes. Data were voice recorded transcribed in Chinese, and then translated into English.

4.3 Data Analysis

Interview transcriptions were coded using grounded theory [28] to identify recurring themes. The coding process started with the extractions of themes that relate to family relationships. To ensure the dependability of the results, three researchers conducted an independent coding on a sample size respectively, which were four pairs of transcriptions chosen randomly from sixteen pairs of participants**.** The intergenerational relationships of theses four pairs include a mother and son, father and son, mother and daughter, and father and daughter. After the initial coding, we compared and discussed the four sets of themes coded by each individual researcher. The initial themes and relationships were identified gradually in a manner of convergence.

In the second stage, one researcher coded the rest of the data iteratively based on the initial themes. When all codes and memos were collected, the researchers grouped them in line with the categories or properties they represented. Based on this arrangement of data, this study examines the online game playing behaviors in both local and remote families.

5 Findings

In this section, we first describe general game playing behaviors demonstrated through observational data, and then introduce the diverse impact of game play on the local and remote families.

5.1 Online Game Playing

As mentioned earlier, we observed six pairs of dual-play families' game activities for two weeks and recorded 319 messages between the six pairs of participants during these two weeks. All families routinely logged onto the game space on a daily basis. Table 2 shows the quantitative data from log analysis. All participants had more than 6 months of gaming experience and reached at least the 15th level according to the game ranks designed in QQ Farm (0 as the lowest and 50 as the highest in the current game design). The participants logged on 1to 6 times per day, with 5 to 30 minutes spent on the game space each time. The log analysis indicates that the participants played QQ Farm in a relatively flexible manner as the game does not require players to play simultaneously over a long period of time.

Some of our participants played QQ Farm while engaging in their other daily activities. Figure 3 shows a participant studying while playing QQ Farm. This participant told us that he was afraid his mature crops might be stolen if he did not gather them in time. He took occasional glances at his farm while studying. Because QQ Farm does not require undivided attention, and allows players to perform other tasks while playing the game, most of our participants had considered the game as less intrusive than most other video games. Thus, our participants were more receptive to QQ Farm than other video games.

Table 2. Log analysis of six pairs of dual-play families

Relation pairs	Parents-children relationship	Age	Game ranks (Levels)	Logins (per day)	Playing Time (Minutes per day)	Interaction messages (2 weeks)
R1	Father	54	38	6	30	93
	Daughter	28	18	6	25	
R2	Father	63	19	2	5	58
	Son	28	15	5	20	
R3	Mother	63	22	3	10	47
	Son	28	15	5	20	
R4	Father	63	34	1	5	37
	Daughter	34	19	5	15	
R5	Mother	60	41	6	30	49
	Son	29	39	4	15	
R6	Mother	65	26	6	30	35
	Daughter	36	24	5	10	

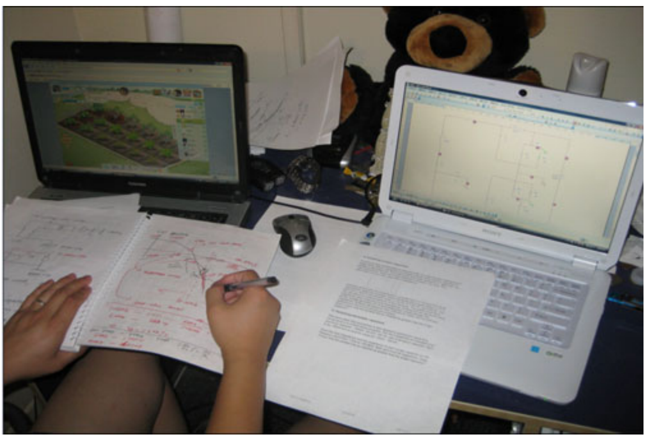


Fig. 3. Playing QQ Farm but not delaying the studying

Our study suggests that Chinese parents and their adult children had different motivations when playing online games. Most parents started playing because their children were playing, whereas the adult children played because they were interested in the game itself. As one participant said:

> It was amazing to watch the growth process of different plants...
> This game let me know what some regional plant species look like and how they grow.

Farming in the game space provides players with an opportunity to connect with a completely new environment and helps them relax away from their daily routines. The game space also serves as a social space in which family can interact. As a game-based interaction, QQ Farm greatly strengthens social ties and family relationships. The dual-play families frequently interacted in the game space, regardless of whether they resided at the same household or not. Family members visited each other's farm from 3 to 24 times a day. (Note: each game log in may be associated with multiple visits). The dual-play dyads admitted that they like to check on each other's farms; they liked to know what plants their parents or children were cultivating and whether they could steal or help each other in the game space. The other forms of interactions among family members, such as face-to-face talks, phone calls, or emails may only happen once or twice a week on average; but dual-play parents and children interacted on the game space multiple times a day. Noteworthy is that although we collected data on all types of family play behaviors, the comparison among dual-play, single-play and non-play families is not the main focus of the current paper; rather, the non-play group served as a basis from which we could understand dual-play behaviors.

5.2 Online Game Adding Joy to Conversations among Local Families

The local families live in the same household or in nearby neighborhoods and can see or visit each other frequently. However, the close physical proximity did not always correspondent to tight family relationships. Differences in education, culture, and values between the dyads often lead to conflict and tension in their relationships [9]. Counter intuitively, this tension may be amplified among local families due to the frequent face-to-face interactions.

In local family dyads, QQ Farm mediated interactions between family members such that their relationship may improve. For example, Su is the primary caregiver for her elderly father and lives in the same household with him. Their interactions were often intense due to frequent disagreement in care-giving activities. They found a way to show concern to each other through QQ Farm.

Su: daughter, 36 years old, living in Chongqing city:

> This game [QQ Farm] opened a door to us to become farmers--the role we couldn't be in the practical world, it was exciting that we can plant and harvest just like a real farmer...We found more things to talk about and found more joy in our conversation... when I found that my father's crops were mature, I would remind him to gather them. If I were at home, I would tell him in-person, else, I would give him a call...

In Chinese, "joy愉快" means happiness 欢乐, and lightheartedness 轻松. This example shows how online game like QQ Farm provides new ways which pleasant conversation can happen among family members.

At the same time, when daily conversation becomes intense, gaming activities can act as pretexts and reasons for making phone calls to family members. For local families, the barriers in maintaining close family ties are not due to the lack of face-to-face interactions, but rather due to differences in values between the two generations. In this situation, game space allows both parties to engage in conversations where these disagreements need not arise.

Even for local families that have well-maintained family ties, QQ Farm can enrich shared family time. Wang is a retired mother with two adult daughters. In Wang's experience, QQ Farm reinforces their existing family ties.

Wang, 63 years old, living in Wuhan city:

> Our family members were all playing QQ Farm... every time my daughters and their families came visit, we shared tips on how to play the game...my grandchild would sometimes join us [at her computer]to play together...this game fascinated everyone in our family!

In Wang's case, her children and grandchild played with her and shared the experiences that occurred in the game space in real-life interactions. The use of a shared computer contributes to the building of family ties since family members have reasons to occupy the same room and talk. Here the QQ Farm game was integrated into family time, and turned into a new family activity.

5.3 Online Game Serves as an Awareness System for Remote Family Members

Unlike local families, remote families are not able to see each other on a daily basis. Family members may also live in different time zones making it harder to make phone calls. In this case, family interactions may rely on either the common leisure time they have or on asynchronous communication that happens infrequently. Resonating with this study [29], non-playing family groups living at long-distances from each other also reported the challenge of staying connected and maintaining close family ties with each other.

Compared with the non-play families, dual-play families use the shared game spaces as another way to interact with family members. In other words, game play is turned into a family interaction that ties family members closer together. Zhang is a mother who was proud that both her son and daughter were attending graduate schools in the United States. Due to the time differences between the US and China As a result, Zhang could not make as many phone calls with her children as she had liked. When Zhang had free time, her son and daughter were studying in the schools; and when her children were free, Zhang was working at her company. Despite these barriers, Zhang managed to connect with her children daily through QQ Farm.

Zhang: mother, 57 years old, living in Wuhan city:

> Last year, my son and daughter went to the US. I missed them very much... Whenever we met in the farms, I feel that they were just next to me... At times I would be too busy to manage my farm and my children would help me in the game space, and that makes me happy.

Parents like Zhang find social connections with their children when their children visit their farms. The farms, as shared virtual spaces, allow players to perform activities that greatly affect each other beyond the game space. Zhang perceives the helpful actions in the game to be representative of how much her children care for her. In this case, family ties are strengthened through seemingly trivial game activities. Because of its asynchronous nature, this game play is able to enrich otherwise intermittent family interactions, when these other forms of interactions become more difficult.

The game space gives remote family members the opportunity to observe each other's online activities. And interestingly, this online observation had allowed many participants stay cognizant of their family members' activities in real-life. Jean, a Ph.D. student of computer science told us how she would know where her father was.

Jean, daughter, 28 years old, living in California:

> I love to guess at what my father is doing by examining his activity log...for instance, when he went on business trips, his farm would become unattended for a while... If later I found his farm gets cleaned up, crops gathered, and new crops sown, I know that he had come back home...

For Jean, the game space allows her to guess father's real-life daily activities based on his game activities. In this sense, the use of online gaming is turned into a system through which players are able to stay aware of each other's daily life [21]. Like *Digital Family Portrait* [23], our study suggests that the game space also has the potential to deliver awareness information amongst family members, although in a less direct manner. In the game space, a person can identify another player's daily game routines to discover atypical behaviors that may be influenced by an event that has happened in real-life. Our participants had not found such observations to be invasive of their private life. But rather, they were alternate ways for family members to care for each other.

Some studies have found that working children have treated communication devices such as the *Digital Family Portrait* as a replacement for the occasional phone call [23]. However, we have not found this to be a problem for online games such as QQ Farm. The way of interacting with family members can also be used purposely to deliver certain information. Yong, an elderly mother living in the countryside, created a personalized message on her farm to show how she missed her son.

Yong, mother, 64 years old, living in Hubei province:

> I believed this farm can convey emotional feelings... I changed the name of my farm to "come home whenever you can"看常回家看看. I hoped my son would come home as often as he can. I know that he is busy [with school work] so I do not want to tell him too directly [or he may get distracted].

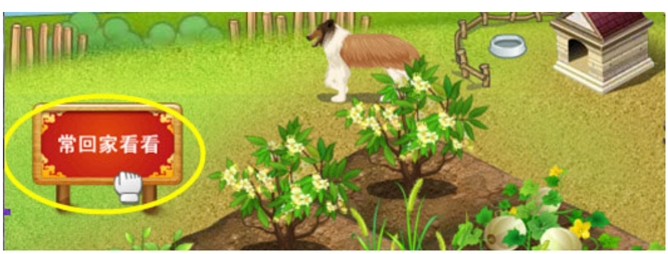


Fig. 4. The bulletin board in Yong's field

As shown in Figure 4, Yong's personalized message indicated that she had wished that her son would visit her more frequently. However, Yong had also wanted to avoid bringing up her feelings in telephone calls for it may become an emotional burden for her distant son. As a result, she used QQ Farm as a system to deliver this message subtly. In this case, the game space was used to deliver messages that may be too sensitive to bring up in direct conversation.

As described in this section, online gaming can connect distant family members. In addition, the game space also allows family members to stay cognizant of each other’s real-life activities. QQ Farm can also be used to convey information that is inappropriate in direct conversation.

6 Discussion

In summary, our research suggests that online games can bolster intergenerational relationships. Online gaming can support intergenerational relationships in two primary ways. First, it generates new conversational topics for local families. Second, it can be used to track the real-life activities of family living at a distance. In this section, we discuss how online games can reinforce existing family relationships.

Online games become common conversational topics for both parents and adult children. When game activities become a family conversation topic, the tension caused by generation gaps is hidden beneath the engagement of game play. When family members live apart, common conversation topics may be difficult to establish between intergenerational family members, since they may have very few common daily activities to share. Discussing online games often ignites fruitful conversations that eventually strengthen intergenerational relationships.

The second finding is that online games can act as an awareness system to keep family members informed about each other’s daily activities, since real-life daily routines are often reflected in online activity patterns, e.g., if Jean’s father changes a real-life activity, his online activity is likely to be affected as well. As we saw in Yong’s story, online activity can also be manipulated to express unspoken needs or concerns to family members.

Furthermore, online games provide awareness information for family members that are geographically distributed. It is critical for family members to co-construct a shared virtual space that keeps logging activities that can be seen by each other. A shared virtual space provides opportunity for players to leave traces of actions that are associated with their real-life activities. Online traces, such as when and what the other family member was engaged in, can be extrapolated into an awareness system of a player’s real-life activity.

The first finding suggests that the importance of designing a game space that is acceptable by both generations. Games with themes that are perceived by older parents as “healthy” are more likely to be adopted as a family play platform. QQ Farm is considered a healthy and educational environment since parents believe their children can learn agriculture and botany knowledge through game play. The playful activities embedded in the game arouse participants’ interests in the game and sustain their interpersonal interactions that are mediated through the game play.

In QQ Farm, two features allow players to trace game activities online: the activity log and the message board. The activity log automatically records activities associated with a certain farm; and the message board can be used as a proactive communication tool for players to pass on information to other players intentionally. Our findings suggest that both features can help family members become more aware of each other's activities.

For family members living apart, the game keeps them aware of some of their remote members' activities, but also augments and enriches existing communications, such as increasing the chances of telephone conversations as well the conservation topics. Beyond family members, closely connected players in other social relationships can also extrapolate each other's real-life activities from their virtual activity patterns, suggesting that online games may similarly augment other real-life relationships when social and cultural barriers hinder other forms of communications, such as in some corporate workgroups.

7 Conclusion

The paper highlights how online games can support intergenerational relationships. Online games can serve to strengthen social bonds underneath the pretense of online game play. Online games can also serve as an awareness system to help derive remote family member's real-life activities. But to achieve these benefits, designers need to develop a game that appeals to players of widely varied ages and backgrounds. Some of the design considerations that make a game truly intergenerational include: requiring minimal effort yet scalable to other committed players, containing themes that appeal to user of all ages, and providing features such as activity log and message board that capture online activities and communication.

References

1. Flew, T., Humphreys, S.: Games: Technology, Industry, Culture, New Media: an Introduction, 2nd edn., pp. 101–114. Oxford University Press, South Melbourne (2005)
2. Castells, M.: The Internet Galaxy, pp. 9–35. Oxford University Press, Oxford (2001)
3. Nardi, B., Harris, J.: Strangers and friends: collaborative play in world of warcraft. In: Proc.CSCW 2006 (November 4–8, 2006)
4. Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., Nickell, E.: From Tree House to Barracks: The Social Life of Guilds in World of Warcraft, Games and Culture, vol. 1(4) (October 2006)
5. Voida, A., Greenberg, S.: Wii all play: The console game as a computational meeting place. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2009), Boston, Massachusetts, April 4-9, pp. 1559–1568. ACM Press, New York (2009)
6. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: "Alone together?" Exploring the social dynamics of massively multiplayer online games. In: CHI 2006: Proceedings of the ACM Conference on Human-Factors in Computing Systems. ACM Press, NY (2006)
7. Nardi, B.: My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft. Digital Culture Books, Ann Arbor (2010)

8. Sheng, X., Settles, B.H.: Intergenerational Relationships and Elderly Care in China: A Global Perspective. Current Sociology 54, 293 (2006)
9. Goh, E.C.L., Kuczynski, L.: 'Only children' and their coalition of parents: Considering grandparents and parents as joint caregivers in urban Xiamen. Asian Journal of Social Psychology 13, 221–231 (2010)
10. Chen, X., Silverstein, M.: Intergenerational Social Support and the Psychological Well-Being of Older Parents in China. Research on Aging 22(1), 43–65 (2000)
11. Silverstein, M., Cong, Z., Li, S.: Intergenerational Transfers and Living Arrangements of Older People in Rural China: Consequences for Psychological Well-Being. J. Gerontol B Psychol. Sci. Soc. Sci. 61(5), S256–S266 (2006)
12. The report of QQ Farm from Chinese website NetEase (163.com), `http://tech.163.com/10/0730/11/6CR9G68A00094IL6.html` (retrieved September 18, 2010)
13. The report of QQ Farm from Chinese website duowan.com (retrieved September 18, 2010), `http://web.duowan.com/1004/136290607498.html`
14. Alex, G., Kate, L.: Just Like the Qing Empire. Internet Addiction, MMOGs, and Moral Crisis in Contemporary Chi, Games and Culture 3(1), 59–75 (2008)
15. Glei, D., Landau, D., Goldman, N., Chuang, Y.L., Rodriguez, G., Weinstein, M.: Participating in social activities helps preserve cognitive function: an analysis of a longitudinal, population-based study of the elderly. Int. J. Epidemiol 34, 864–871 (2005)
16. Lindtner, S., Mainwaring, S., Dourish, P., Wang, Y.: Situating Productive Play: Online Gaming Practices and Guanxi in China. In: Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I, pp. 328–341 (2009)
17. Kow, Y.M., Nardi, B.: Culture and Creativity: World of Warcraft Modding in China and the U.S. In: Bainbridge (ed.) Online Worlds: Convergence of the Real and the Virtual. Springer, London (2009)
18. Khoo, E.T., Lee, S.P., Cheok, A.D., Kodagoda, S., Zhou, Y., Toh, G.S.: Age Invaders: social and physical intergenerational family entertainment. In: Proc.CHI 2006, pp. 243–247. ACM Press, New York (2006)
19. Hilary, D., Frank, V., Peter, F., Martin, G., Steve, H.: "I Wish We Could Get Together": Exploring Intergenerational Play across a Distance via a 'Magic Box'. Journal of Intergenerational Relationships (2008)
20. Voida, A., Carpendale, S., Greenberg, S.: The Individual and the Group in Console Gaming. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, Georgia, USA, pp. 371–380 (2010)
21. Romero, N., Markopoulos, P., Baren, J., Ruyter, B., Jsselsteijn, W., Farshchian, B.: Connecting the Family with Awareness Systems, Personal and Ubiquitous Computing, vol. 11. Springer, Heidelberg (2006)
22. Tee, K., Brush, A.J.B., Inkpen, K.M.: Exploring communication and sharing between extended families. International Journal of Human-Computer Studies 67(2), 128–138 (2009)
23. Rowan, J., Mynatt, E.D.: Digital Family Portrait Field Trial: Support for Aging in Place. In: Proc. CHI 2005, pp. 521–530. ACM Press, New York (2005)
24. Tejinder, K.J., Carman, T.K., Neustaedter, C., Andrew, F., Kurtz, A.F.: The Family Window: The Design and Evaluation of a Domestic Media Space. In: Proc. CHI 2010: We Are Family, pp. 2361–2370. ACM Press, Atlanta (2010)
25. Brush, A.J., Inkpen, k., Tee, k.: SPARCS: exploring sharing suggestions to enhance family connectedness. In: Proc. CSCW 2008, Savannah, Georgia, USA, pp. 629–638 (2008)

26. Lindley, S.E., Harper, R.H.R., Sellen, A.: Designing a Technological Playground: A Field Study of the Emergence of Play in Household Messaging. In: Proc.CHI 2010: We Are Family, Atlanta, GA, USA (April 10–15, 2010)
27. Boyd, D., Ellison, N.: Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication 13(1) (2007)
28. Glaser, B.G., Strauss, A.L.: The discovery of grounded theory: Strategies for qualitative research. Aldine de Gruyter., Hawthorne, NY (1967)
29. Cao, X., Abigail, S., Brush, A.J., David, K., Darren, E., Ding, X.: Understanding Family Communication across Time Zone. In: Proc. CSCW 2008, Savannah, Georgia, USA, pp. 155-158 (2008)

The Influence of Customer Familiarity and Personal Innovativeness toward Information Technologies on the Sense of Virtual Community and Participation

Manuel J. Sánchez-Franco[1,*], José Antonio Carballar-Falcón[1], Francisco J. Martínez-López[2], and Juan Carlos Gázquez-Abad[3]

[1] University of Sevilla, Ramón y Cajal 1,
E-41018 Sevilla, Spain
majesus@us.es
[2] University of Granada, Campus Cartuja, s/n,
E-18171 Granada, Spain
[3] University of Almería, Ctra. Sacramento s/n,
E-04120, Almería, Spain

Abstract. The aim of this study is to investigate, on the one hand, the main effects of personal innovativeness and familiarity on perceived community support in the domain of the social network site Tuenti, and, on the other hand, the moderating effects of personal innovativeness on relationships between familiarity and perceived community support. A Structural Equation Modelling, specifically Partial Least Square, is proposed to assess the relationships between the constructs together with the predictive power of the model analysed. The results demonstrate that increasing PIIT and familiarity can be seen as important initiatives that promote perceived community support and avoid the consideration of competitive social network sites.

Keywords: Perceived community support, Social Network Sites, sense of virtual community, familiarity, personal innovativeness, moderating effects.

1 Introduction

Social Network Sites (SNSs) are virtual environments that allow users to connect with each other in order to seek and/or share content (*e.g.*, opinions or experiences), professional advancement, friendship, and entertainment. In particular, users have an essential focus on social interaction and mutual support. In this research, SNSs will thus be conceptualised as an online setting that emphasises (a) an individual's feelings of membership, identity, belonging, and attachment to others that is maintained by integration and fulfillment of needs (*i.e.*, a sense of a virtual community), and (b) virtual enduring participation [*cf.* 10, 35, 44].

Assuming that the ultimate success of SNSs derives from an enduring usage, Bhattacherjee (2001) emphasised the importance of continuance over initial

[*] Coresponding author.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 265–279, 2011.

acceptance. Long-term viability of SNSs and their eventual success will indeed depend on their enduring participation rather than situational participation, and, likewise, the formation of social capital –*e.g.*, support seeking and support providing [*cf.* 16]. Nevertheless, how to maintain and intensify the number of members and posts remains a problem. In fact, if not gratified and involved properly, members lose interest and eventually reduce their level of interaction. Perceived community support [hereinafter, PCS] can, then, be considered as an essential indicator of community success. Assuming that "the study of the community as a relevant source of support has been comparatively lacking" [26, p. 197], the purpose of this research will be to expand previous research of what contributes to PCS.

In particular, this paper is structured as follows. In the next section, we present a research model which connects PCS with two sets of factors: familiarity and personal innovativeness toward Information Technologies (hereinafter, PIIT). That is to say, identifying main determinants of PCS are the goals of this study. We will investigate the roles of familiarity and PIIT in building PCS. These two factors have not been examined in previous research on PCS. The hypothesised model will be validated empirically using data from a field survey of the most popular computer-mediated SNS among the Spanish college student population, Tuenti -created in January 2006 by Zaryn Dentzel. This paper will address these questions, and outline the managerial implications, and limitations.

2 Theory and Research Hypotheses

Previous research has shown individual differences play an important role in users' decisions to accept or reject innovations [38, 39, 41, 46]. In this regard, we propose two factors that can help identify such drivers of PCS to target those members that are highly enthusiastic during the acceptance of an SNS. The first is a personal trait (*i.e.*, a stable situation-specific individual difference) named personal innovativeness toward Information Technologies (PIIT), and the second is a psychological state (*i.e.*, a dynamic, situation-specific individual difference) named familiarity.

2.1 Perceived Community Support, PCS

Blanchard and Markus (2002) argue that "the success of community support platforms depends on the active participation of a significant percentage of the community members". Community participation will herein be conceptualised as community involvement, active participation in community activities, or social interaction [44]; e.g., "*I respond to calls to support my Tuenti community*". Participation is necessary for enduring interactivity and, consequently, an essential driver of successful SNSs. Secondly, the exchange of mutual support –*cf.* integration and fulfillment of needs- is a very important reason for the existence of virtual communities [4, 42, 49]; *i.e.* feelings of being supported by others in the community (*i.e.*, community organisation) while also supporting them. In fact, online contacts are perceived to be as helpful as support provided by offline contacts [*cf.* 34, 48]; e.g., "*I could find people that helped me feel better*". Thirdly, a main dimension -associated with community support- that distinguishes virtual communities from mere virtual groups is mainly their sense of virtual community (or community integration), defined as "as members' feelings of membership, identity, belonging, and attachment to a group that interacts primarily through electronic communication" [7; *cf.* 8, 35]; e.g., "*My affective bonds with my Tuenti community are the main reason why I continue to use its service*".

2.2 Personal Innovativeness toward Information Technologies, PIIT

Since SNSs represent an innovation, existing innovation theoretical frameworks from the innovation literature may be effectively applied to SNSs –as a new technology service. If users are more willing to accept innovative technologies in general, they may be more accepting of innovative technologies such as SNS, and more predisposed to having a primary focus on social interpersonal communication. Given the relative infancy of SNSs, it will, therefore, be appropriate to test personal innovativeness toward Information Technologies (PIIT) as an influencing variable.

In our research, as Agarwal and Prasad (1998), we shift the focus from observed adoption behaviour to underlying personal trait operating in differing technology acceptance realms. PIIT has a stable influence across situations involving information technology, while Rogers defines innovativeness as a behaviour -in the context of diffusion of innovations. In particular, PIIT is "an important individual trait for examining the acceptance of information technology innovations" [1, p. 206; *cf.* also 24, 26, 27], and it is conceptualised as an individual's willingness to try out an IT, seeking out new, mentally or sensually stimulating experiences. PIIT describes the extent to which the individual has an innate propensity toward adopting a new IT; *i.e.*, a form of openness to change. Innovative individuals will thus be more comfortable and confident when using a new system [*cf.* 51], this being a function of members' risk tolerance. That is, as Woszczynski *et al.* (2002, p.381) argued, "users who score high on PIIT may tend to be the first individuals to adopt a new technology product".

Initially we conceptualise and evaluate PIIT as a direct determinant of PCS. Indeed, we expect a positive relationship exists between PIIT and the formation and maintenance of social capital. Highly-innovative members will develop more positive perceptions about SNSs in terms of perceived usefulness and ease of use. In this regard, innovative members realise better the usefulness of SNSs for trying out a SNS and consequently, support seeking and support providing (*i.e.*, community organisation and participation). Rogers (1995) related innovativeness with social participation. Users who adopt innovative ideas more than others will tend to have a stronger desire to participate in and contribute to SNS. PIIT captures a positive dimension of individuals to experiment with SNSs. Likewise, as Chou (2010) confirms, online community members' PIIT is positively related to their perceived identity verification. "For most participants, identity —both the establishment of their own reputation and the recognition of others— plays a vital role" [15, p.30]. Highly-innovative members have more positive individual feelings of identity and attachment with the others (*i.e.*, community integration), showing a strong link between PIIT and PCS.

To sum up, innovative members will tend to be proactive in accepting SNS and in taking risks, solving a problem or generating an idea through online interactions –as instrumental benefits- or for fun and relaxation –as hedonic benefits- among others and predicting stronger feelings of being supported by the SNS. PCS will, therefore, be influenced by stable individual difference such as PIIT. Based on the previous arguments, this research proposes the following hypothesis: *PIIT positively influences PCS (i.e., community organisation, **H1a**; integration, **H1b**; and participation, **H1c**).* See Figure 1.

2.3 Familiarity

Familiarity about the functioning of an SNS will be defined here as the number of SNS-related direct or indirect experiences that have been accumulated by members, being considered one of the main drivers explaining their behaviour [*cf.* 42].

Familiarity is an important internal source of information; *i.e.*, the more familiar the SNS are to members, the less time and effort will be spent on information search. Familiarity is the knowledge of a product class [27] or "a specific activity-based cognizance based on previous experience or learning of how to use the particular interface" [19, p.727]; *i.e.*, role clarity, ability and motivation [30]. Familiarity therefore, gauges the degree to which prior experience has been understood.

Users' familiarity (and knowledge) will affect customer information processing activities in several ways [*cf.* 3], leading to a more elaborated cognitive structure [*cf.* 17, 36]. As users use a service more frequently, they are more familiar with it, reducing uncertainty in future decision situations (*i.e.*, perceiving less risk), and simplifying relationships with others by generating a knowledge structure [*cf.* also 6, 18, 19]. Familiarised members will believe that they know an SNS well [*cf.* 3], showing an individual judgment of one's capability to use an SNS (*i.e.*, self-efficacy) and more positive beliefs and attitudes about it. Familiarity could thus be used to describe the extent to which users know about an SNS.

A general principle is the easier the utilisation could be and the lower the entry barrier, the higher the possibility of using it is. Users' satisfaction and familiarity both have positive impacts on their voluntary participation in co-production and contribution to SNS [*cf.*, 23]. Hence, we expect a positive relationship exists between members' familiarity and the formation and maintenance of social capital; that is to say, the familiar virtual settings create a level of comfort such that members feel free to discuss and share their ideas. Increased familiarity among members will promote mutual aid and empathy. When the members are familiar with the others, they tend to be more informal and are more willing to share personal information, thereby (a) achieving greater PCS and (b) resulting in increased opportunities to deepen their knowledge about them [*cf.* 47]. Likewise, interpersonal interactions increase familiarity between community members, leading to a greater desire to interact in the future. Indeed, previous research supports that the more familiarised members normally exhibit targeted behaviour while the less familiarised opt for general exploration. Less familiarised users tend to engage in community participation but in a limited way, preferring to be readers rather than writers, and being reticent -or very seldom posting. As Casaló et al. (2007, p.6) proposed, "familiarity with a virtual community favours the individual skills in order to interact with other community members and, therefore, the participation in the virtual community may be increased".

To sum up, to promote PCS online moderators should use strategies to increase familiarity among members of SNS and develop favourable social relationships. Users will show more commitment once they could understand the value, norm and rule of the SNS. "As customers become familiar with an organization and/or its employees, relationships based on trust and reciprocity often develop" [23, p.13]. Based on the previous arguments, this research proposes the following hypothesis: *Familiarity positively influences PCS (i.e., community organisation, **H2a**; integration, **H2b**; and participation, **H2c**).* See Figure 1.

Thirdly, members with higher PIIT would require less information than the members with lower PIIT. Innovative members will be highly predisposed toward adopting SNSs, whereas others may prefer to continue exploring familiar community-tools. PIIT could then be a moderator of familiarity. Members with high personal innovativeness may not be familiarised members. Innovative users might be easily persuaded to try the new technology without any elaborated cognitive structure (or familiarity). We will, therefore, develop a third hypothesis maintaining that innovative members will not only hold different belief structures in terms of familiarity, but also that individual innovativeness will moderate the impact of familiarity on PCS. Higher levels of PIIT may diminish factors involving prior online experience (*e.g.*, SNS familiarity), and place more focus on new, mentally or sensually stimulating experiences. PIIT may, therefore, be viewed as a moderating driver affecting the relationship between familiarity and PCS. Based on the previous arguments, this research proposes the following hypothesis: *Overall, PIIT moderates (weakens) the relationship between familiarity and PCS (i.e., community organisation, **H3a**; integration, **H3b**; and participation, **H3c**).* See Figure 1.

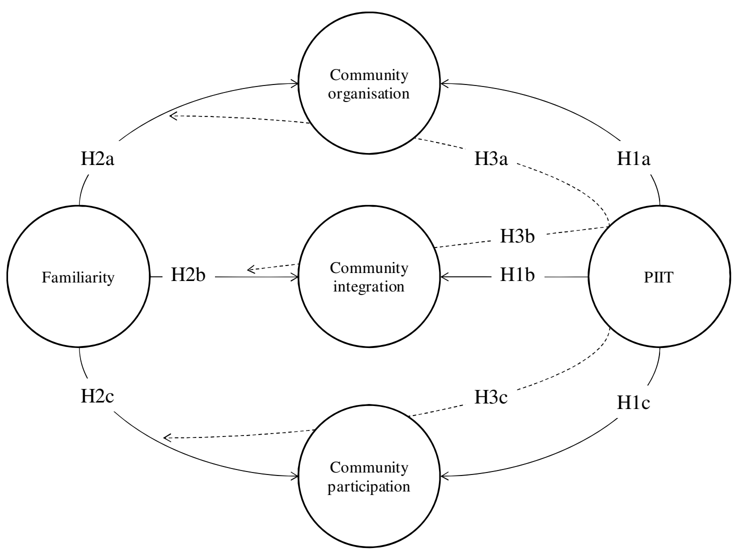


Fig. 1. Research model. Hypotheses.

3 Method

3.1 Participants

The structural model was validated empirically using data from a field survey of the most popular computer-mediated SNS among the Spanish college student population, Tuenti. Specifically, participants were recruited from social studies at a public University in Seville (Spain). The exclusion of invalid questionnaires due to duplicate submissions or extensive empty data fields resulted in a final convenience sample of 278 users. 42.8% were male respondents. The average age was 21.04 (SD: 2.403).

3.2 Measures

We have adopted the common latent model perspective, in which the latent variable is understood to be the cause of the indicators (i.e., reflective indicators). Fourteen items were used to assess community participation, organisation and integration (or identification with an SNS) -taken from Herrero and Gracia (2007), Geyskens *et al.* (1996), Loewenfeld (2006), and Sánchez-Franco (2009). On the other hand, a total of three items were employed to measure familiarity [18]. Finally, the self-report instrument for measuring the degree of PIIT has been operationalised by Agarwal and Prasad (1998) in the form of a four-item questionnaire.

A pretest assessed the suitability of the wording and format, and the extent to which measures represented all facets of constructs. All items are seven-point Likert-type, ranging from «strongly disagree», 1, to «strongly agree», 7.

3.3 Data Analysis

The hypotheses testing was conducted using Partial Least Squares (PLS), specifically, SmartPLS 2.0.M3 software [37].

Taking into account that hypotheses H3abc are based on interaction effects, one well-known technique has had to be applied to test these moderated relationships: product-indicator approach [24].

4 Evaluation of PLS Path Model Results

4.1 Assessing Our Reflective Measurement Model

Our measurement model proposed causal relationships from the latent variables to the manifest variables in their block. That is, as Henseler *et al.* (2009, p.285) summarise, "each manifest variable in a certain measurement model is assumed to be generated as a linear function of our latent variables and the residual ε".

In particular, the measurement model was evaluated using the full sample (278 individuals) -all items and dimensions- and then the PLS results were used to eliminate possible problematic items. On the one hand, individual reflective-item reliability was assessed by examining the loadings of the items with their respective construct, that is, in terms of correlations between a construct and each of its manifest variables (or standardised loadings). Individual reflective-item reliabilities were over the recommended acceptable cut-off level of 0.7. See Table 1. On the other hand, construct reliability was assessed using the composite reliability (ρc). The composite reliability takes into account that indicator shave different loadings, and can be (a) interpreted in the same way as Cronbach's α, and therefore, (b) used to check how well a construct is measured by its assigned indicators. The composite reliabilities for the multiple reflective indicators were well over the recommended acceptable 0.7 level, demonstrating high internal consistency. Moreover, we checked the significance of the loadings with a bootstrap procedure (500 sub-samples) for obtaining t-statistic values. They all are significant. See Table 1.

Finally, convergent validity on the one hand, involves the degree to which individual items reflecting a construct converge in comparison to items measuring different constructs. On the other hand, discriminant validity concerns the degree to which the measures of different constructs differ from one another. Convergent and discriminant validities were, therefore, assessed by stipulating that the square root of the average variance extracted (AVE) from its indicators should be at least 0.7 (*i.e.*, AVE > 0.5) and should be greater than that construct's correlation with other constructs. All latent constructs satisfied these conditions. See Tables 1 and 2.

Table 1. Measurement model. Main effects model.

a. Individual item reliability-individual item loadings. Construct reliability and convergent validity coefficients			
Latent Dimension	Loadings[a]	ρ_c	AVE
CO. Community organisation		0.9243	0.7097
CO1. I could find people that helped me feel better	0.8059		
CO2. I could find someone to listen to me when I felt down	0.8603		
CO3. I could find a source of satisfaction for myself	0.8722		
CO4. I could be able to cheer up and get into a better mood	0.8749		
CO5. I could relax and easily forget my problems	0.7955		
CI. Identification with virtual community (*i.e.,* community integration)	0.8059	0.9426	0.8042
CI1. My affective bonds with my Tuenti community are the main reason why I continue to use its service	0.8896		
CI2. I enjoy being a member of my Tuenti community	0.9103		
CI3. I have strong feelings for my Tuenti community	0.8799		
CI4. In general, I relate very well to the members of my Tuenti community	0.9069		

Table 1. (*continued*)

CP. Community participation		0.9109	0.6723
CP1. I participate in order to stimulate my Tuenti community	0.8787		
CP2. I take part actively in activities in my Tuenti community	0.7801		
CP3. I take part in social groups in my Tuenti community	0.7789		
CP4. I respond to calls to support my Tuenti community	0.7761		
CP5. I take part actively in socio-recreational activities in my Tuenti community	0.8786		
FAM. Familiarity		0.9306	0.8172
FAM1. I am quite familiar with my Tuenti community	0.9038		
FAM2. I am quite familiar with the services offered by my Tuenti community	0.9004		
FAM3. In comparison with the typical user of my Tuenti community, I believe I am quite familiar with it	0.9077		
PIIT. PIIT		0.9487	0.8222
PIIT1. If I heard about a new information technology, I would look for ways of experimenting with it	0.8995		
PIIT2. Among my peers, I am usually the first to try out new information technologies	0.8757		
PIIT3. In general, I am hesitant to try out new information technologies (R)	0.9313		
PIIT4. I like to experiment with new information technologies	0.9196		

(R) Reverse item

[a] All loadings are significant at p<0.001- (based on $t_{(499)}$, two-tailed test)

Table 2. Discriminant validity coefficients. Main effects model.

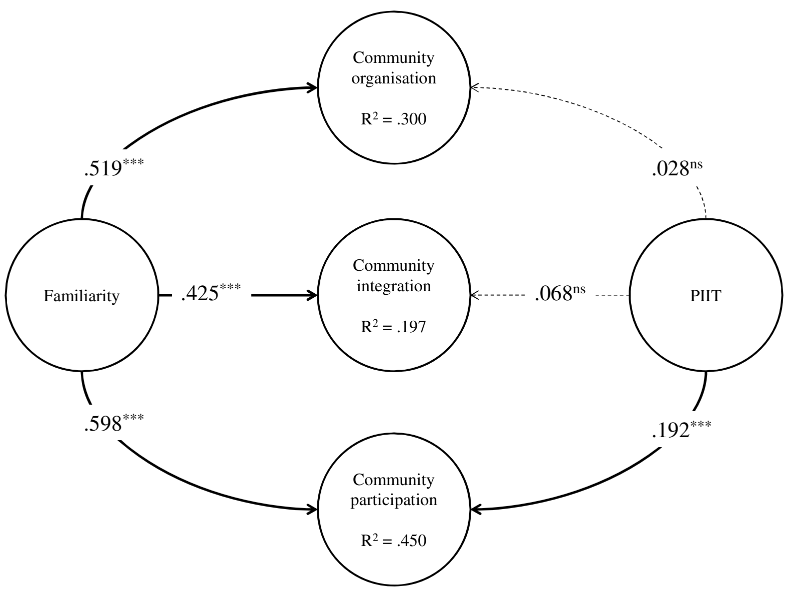
	CO	CI	CP	FAM	PIIT
CO	***0.8424***				
CI	0.3932	***0.8968***			
CP	0.5358	0.4248	***0.8200***		
FAM	0.5237	0.4371	0.6326	***0.9400***	
PIIT	0.1208	0.1438	0.2996	0.1790	***0.9070***

Note: Diagonal elements are the square root of average variance extracted (AVE) between the constructs and their measures. Off-diagonal elements are correlations between constructs

4.2 Assessing Our Structural Model

The bootstrap re-sampling procedure (500 sub-samples) was used to generate the standard errors and the t-values. Firstly, the research model appears to have an appropriate predictive power for endogenous constructs to exceed the required amount of 0.10 –R-square values. Furthermore, our structural model explains each endogenous latent construct by only two exogenous latent construct. R^2>0.19 may, therefore, be acceptable [*cf.* 12]. See Figure 2. A measure of the predictive relevance of dependent variables in the proposed model is the *Q2* test. A *Q2* value (*i.e.*, only applicable in dependent and reflective constructs) greater than 0 implies that the model offers predictive relevance. The results of our study confirm that the main model offers very satisfactory predictive relevance: community integration (*Q2*= 0.1507 > 0), community participation (*Q2*= 0.3636 > 0) and community organisation (*Q2*= 0.1930 > 0).

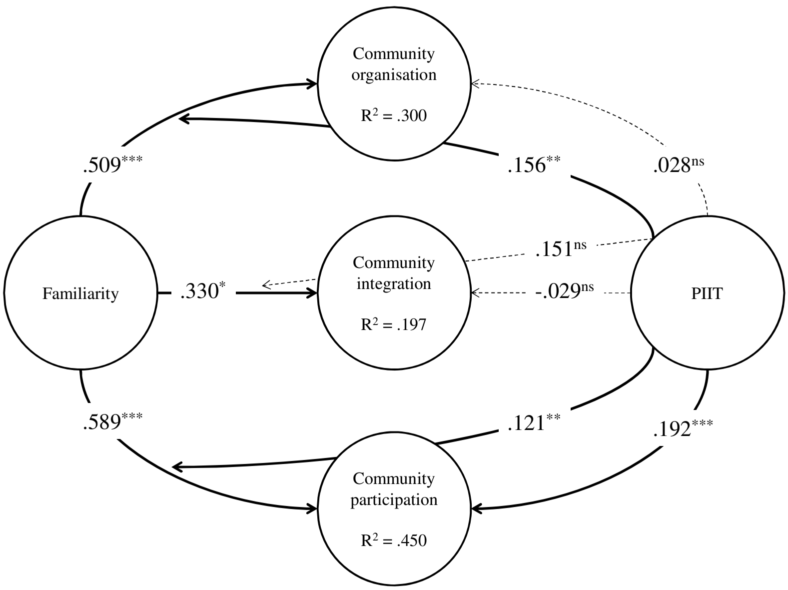
The data fully supported the main effects model and the hypotheses (excepting H1a and H1b) are supported on the basis of empirical data. As indicated in the main effects model, PIIT has a significant impact on participation, with path coefficients of 0.192 (p<0.001). However, due to the cross-sectional design of this study, no causation can be determined. Likewise, familiarity has a significant impact on organisation, integration and participation with path coefficients of 0.519 (p<0.001), 0.425 (p<0.001) and 0.598 (p>0.001). See Figure 2.



*** *p < 0.001, ns = not significant (based on t(499), one-tailed test)*

Fig. 2. Structural model. Main effects model.

The interaction effects were also included, in addition to the main effects model - see Figure 3. As in regression analysis, the predictor and moderator variables are multiplied to obtain the interaction terms. According to Chin *et al.* (2003), product indicators are developed by creating all possible products from the two sets of indicators and the standardising of the product indicators is recommended. However, in the presence of significant interaction terms involving any of the main effects, no direct conclusion can be drawn from these main effects alone [2]. In particular, the interaction effects were of 0.121 -familiarity * PIIT → community participation- (p<0.01), and 0.156 - familiarity * PIIT → community organisation - (p<0.01). Hypotheses H3a and H3c were supported. Higher PIIT increases the impact of familiarity on community organisation and participation.



*** p < 0.001, ** p < 0.01, * p < 0.05, ns = not significant (based on t(499), one-tailed test)

Fig. 3. Structural model. Interaction effects model.

5 Conclusions

Our research focused on the association between PIIT, familiarity and PCS by studying the moderating effects of PIIT –*i.e.*, the interaction effects model. We were unaware of previous research that examined interaction effects of familiarity and PIIT on the sense of virtual community and participation. In order to establish a social integration between users and SNSs, this study extended previous research.

In particular, the measurement models were valid, with acceptable convergent and discriminant validity. Our results provided strong support for the arguments that personal differences (*i.e.*, PIIT and familiarity) lead the Tuenti member into developing a growing PCS -by enthusiasm for a community activity and reducing the uncertainty of the relationship with it. PIIT directly influences on community participation, and familiarity influences on community participation, integration and organisation. Likewise, higher PIIT increases the impact of familiarity on community organisation and participation.

Community managers will, therefore, take account of familiarity and PIIT to customise their social strategies in order to (a) increase users' enthusiasm in SNSs, and (b) establish involvement and participation in community activities and, consequently, the individual's feelings of identity and attachment to them.

6 Discussion

Members with higher levels of PIIT (or willingness of an individual to try out any new Information Technology) (a) tend to be more risk-taking, and (b) are expected to have more positive intentions toward community participation. Innovative members will be *selective processors* who often rely on a subset of highly available and salient cues in place of detailed elaboration related to community integration or affective commitment. Innovative members will specially try the Tuenti community because it may provide productivity enhancement – even if they do not have a positive attitude towards using it or community integration –.

Likewise, it is suggested that PIIT could be a moderator of familiarity; *i.e.*, users with high personal innovativeness may not be familiarised users. Therefore, to some extent these relationships (between familiarity and PCS) could be influenced by an individual's enthusiasm or keen interest in Information Technology -which has been established as PIIT. We initially proposed that innovative users might be easily persuaded to try the new technology without any elaborated cognitive structure (or familiarity). Contrariwise, PIIT predisposed members to a higher influence of familiarity on community organisation, and participation. Hence, when familiarised members (*i.e.*, with fulfilling routines and procedures for dealing with Tuenti community) are more concerned about Information Technologies (and less deeply engaged in cognitive elaboration), they will thus more inclined to promote accurate mutual aid and empathy. They will tend to be more informal and are more willing to share personal information, thereby (a) achieving greater PCS and (b) resulting in increased opportunities to deepen their knowledge about them.

End-users' familiarity with using SNSs emphasises sense of virtual community and participation. Alternatively, innovative members will apply community features in innovative and novel ways that go beyond the familiarity -enabling them to explore the value potential of SNSs more fully. Accordingly, different members' segments (defined by PIIT) will play an interaction role in affecting the influence of familiarity (in terms of the knowledge members have of an SNS, based on their experience and previous contacts) on the success of community support SNS –*i.e.*, the active mutual exchange and integration of a significant percentage of the community members. These members will progressively overestimate members' feelings of being assisted by the online community in terms of support needs and resources available to the individual –*e.g.*, knowledge and ideas related to mutual interest. Furthermore, increasing PIIT and familiarity can be seen as important initiatives that promote integration and avoid the consideration of competitive SNSs. As Sørebø *et al.* (2008, p.326) note, "an important challenge, for both IS research and practice is to find ways of influencing users to initiate and engage in purposeful behaviour". A community manager should consider how to assist members fostering intrinsic motivation so that they can attain gains in social integration –e.g. novelty, flow, and increased perceived control, or providing community features where innovative members who have provided useful suggestions to other members are identified. In fact, "for individuals with high innovativeness in IT, improving their identity verification in an online community becomes very important" [14, p.848]. Community managers should also focus more on providing expressive and interesting features in SNSs, such as an attractive interface. For instance, as Sánchez-Franco and Roldán (2010, p.1455) concluded, "the influence of aesthetics on feelings of membership, identity, and attachment to a SNS was outweighed in routinisation conditions and personal innovativeness, reinforcing perceived community support". Moreover, these authors concluded that the visual originality itself is essential and inseparable from innovative users. Likewise, members need to be motivated so that they are willing to get familiar with the SNS. In this regard, community managers need to "understand the nature of customer expectations, how these are formed and how they change" (Kotzë and du Plessis, 2003, p193).

The authors recognise some limitations in this research. The model clearly did not include all the relevant variables. Future research should not overlook calculative integration (or commitment); that is, "the state of attachment to a partner cognitively experienced as a realization of the benefits sacrificed and losses incurred if the relationship were to end" [21, p.28]. Secondly, the sample selection process was non-random. The selection of the respondents might have been subject to a sample selection bias and therefore, possible refinements in this direction are advisable for future research. In particular, our respondents showed a bias on gender. As Green (1996) evidenced, previous research tends to indicate that female, older, and better-educated customers participate more in surveys. Therefore, it is difficult to generalise this quasimoderating model and extend the results to other social settings because members from only Tuenti community were surveyed. When studying users of one SNS, future research should exercise caution in generalising the findings to members of another social network site; i.e., certain profiles are more represented on some sites than others.

Acknowledgments. This research was supported by the Junta de Andalucía, Spain (Consejería de Economía, Innovación y Ciencia) -Proyecto de investigación de excelencia SEJ-5801.

References

1. Agarwal, R., Prasad, J.: A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Information Systems Research 9(2), 204–215 (1998)
2. Aiken, L.S., West, S.G.: Multiple regression: Testing and Interpreting Interactions. Sage Publications, Newbury Park (1991)
3. Alba, J.W., Hutchinson, J.W.: Dimensions of Consumer Expertise. Journal of Consumer Research 13(4), 411–454 (1987)

4. Baym, N.: Interpreting Soap Operas and Creating Community: Inside an Electronic Fan Culture. In: Keisler, S. (ed.) Culture of the Internet. Lawrence Erlbaum Associates, Mahwah (1997)
5. Bhattacherjee, A., Premkumar, G.: Understanding Changes in Belief and Attitude toward Information Technology Usage: a Theoretical Model and Longitudinal Test. MIS Quarterly 28(2), 229–254 (2004)
6. Bhattacherjee, A.: Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 25(3), 351–370 (2001)
7. Blanchard, A.L.: Developing a Sense of Virtual Community Measure. CyberPsychology & Behavior 10(6), 827–830 (2007)
8. Blanchard, A.L.: Testing a Model of Sense of Virtual Community. Computers in Human Behavior 24(5), 2107–2123 (2008)
9. Blanchard, A.L., Markus, L.: Sense of Virtual Community-Maintaining the Experience of Belonging. In: Proceedings of the 35th HICSS Conference, Hawaii, vol. 8 (2002)
10. Blanchard, A.L., Markus, L.: Technology and Community Behavior in Online Environments. In: 3rd International Conference on Communities and Technologies, June 28–30, Michigan State University, Lansing (2007)
11. Casaló, L.V., Flavián Blanco, C., Guinalíu, M.: Antecedents and Effects of Participation in Virtual Brand Communities. In: IADIS International Conference on Web Based Communities, Salamanca, Spain (February 1–10, 2007)
12. Chin, W.W.: The Partial Least Squares Approach to Structural Equation Modelling. In: Marcoulides, G.A. (ed.) Modern Methods for Business Research, pp. 295–336. Lawrence Erlbaum Associates, Inc., Mahwah (1998)
13. Chin, W.W., Marcolin, B.L., Newsted, P.R.: A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic Mail Emotion/Adoption Study. Information Systems Research 14(2), 189–217 (2003)
14. Chou, Sh-W: Why do Members Contribute Knowledge to Online Communities? Online Information Review (6), 829–854 (2010)
15. Donath, J.S.: Identity and Deception in the Virtual Community. In: Smith, M.A., Kollock, P. (eds.) Communities in Cyberspace, pp. 29–59. Routledge, New York (1999)
16. Ellison, N., Steinfield, C., Lampe, C.: The Benefits of Facebook "Friends": Social Capital and College Students' Use of Online Social Network Sites. Journal of Computer-Mediated Communication 12(4) (2007), `http://jcmc.indiana.edu/vol12/issue4/ellison.html`
17. Fiske, S.T., Kinder, D.R., Larter, W.M.: The Novice and the Expert: Knowledge-Based Strategies in Political Cognition. Journal of Experimental Social Psychology 19, 381–400 (1983)
18. Flavián, C., Guinalíu, M., Gurrea, R.: The Influence of Familiarity and Usability on Loyalty to Online Journalistic Services: the Role of User Experience. Journal of Retail and Consumer Services 13, 363–375 (2006)
19. Gefen, D.: It is not Enough to be Responsive: the Role of Cooperative Intentions in MRP II Adoption. The Data Base for Advances in Information Systems 31(2), 65–79 (2000)
20. Geyskens, I., Steenkamp, J.-B., Scheer, L.K., Kumar, N.: The Effects of Trust and Interdependence on Relationship Commitment: A Trans-Atlantic Study. International Journal of Research in Marketing 13(4), 303–317 (1996)
21. Gilliland, D.I., Bello, D.C.: Two Sides to Attitudinal Commitment: The Effect of Calculative and Loyalty Commitment on Enforcement Mechanisms in Distribution Channels. Journal of the Academy of Marketing Science 30(1), 24–43 (2002)

22. Green, K.E.: Sociodemographic Factors and Mail Survey Response. Psychology & Marketing 13(2), 171–185 (1996)
23. Groth, M.: Customers as Good Soldiers: Examining Citizenship Behaviors in Internet Service Deliveries. Journal of Management 31(1), 7–27 (2005)
24. Henseler, J., Fassott, G.: Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures. In: Vinzi, E.V., Chin, W.W., Henseler, J., Wang, H. (eds.) Handbook of PLS and Marketing, pp. 713–736. Springer, Berlin (2010)
25. Henseler, J., Ringle, C.M., Sinkovics, R.R.: The Use of Partial Least Squares Path Modeling in International Marketing. Advances in International Marketing 20, 277–319 (2009)
26. Herrero, J., Gracia, E.: Measuring Perceived Community Support: Factorial Structure, Longitudinal Invariance, and Predictive Validity of the PCSQ (Perceived Community Support Questionnaire). Journal of Community Psychology 35(2), 197–217 (2007)
27. Johnson, E.J., Russo, J.E.: Product Familiarity and Learning New Information. Journal of Consumer Research 11, 542–550 (1984)
28. Karahanna, E., Ahuja, M., Srite, M., Galvin, J.: Individual Differences and Relative Advantage: the Case of GSS. Decision Support Systems 32(4), 327–341 (2002)
29. Kotzé, T.G., du Plessis, P.J.: Students as "Co-producers" of Education: A Proposed Model of Student Socialization and Participation at Tertiary Institutions. Quality Assurance in Education 11(4), 186–201 (2003)
30. Lengnick-Hall, C.A., Claycomb, V., Inks, L.W.: From Recipient to Contributor: Examining Customer Roles and Experienced Outcomes. European Journal of Marketing 34(3/4), 359–383 (2000)
31. Lewis, W., Agarwal, R., Sambamurthy, V.: Sources Influence on Beliefs about Information Technology Use: An Empirical Study of Knowledge Workers. MIS Quarterly 27(4), 657–678 (2003)
32. Limayem, M., Khalifa, M., Frini, A.: What Makes Consumers Buy from Internet? A Longitudinal Study of Online Shopping. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 30(4), 421–432 (2000)
33. Loewenfeld, F.: Brand Communities—Erfolgsfaktoren und Ökonomische Relevanz von Markeneigenschaften. Deutscher Universitäts-Verlag, Wiesbaden (2006)
34. McKenna, K.Y.A., Green, A.S., Gleason, M.E.J.: Relationship Formation on the Internet: What's the Big Attraction? Journal of Social Issues 58(1), 9–31 (2002)
35. McMillan, D.W., Chavis, D.M.: Sense of Community: A Definition and Theory. Journal of Community Psychology 14(1), 6–23 (1986)
36. Mitchell, A.A., Dacin, P.F.: The Assessment of Alternative Measures of Consumer Expertise. Journal of Consumer Research 23, 219–240 (1996)
37. Ringle, C.M., Wende, S., Will, A.: SmartPLS 2.0 (Beta). University of Hamburg: Hamburg, Germany (2005), `http://www.smartpls.de` (accessed March 20, 2008)
38. Roehrich, G.: Consumer Innovativeness: Concepts and Measurements. Journal of Business Research 57(6), 671–677 (2002)
39. Rogers, E.M.: Diffusion of Innovations, 1st edn. Free Press, New York (1962)
40. Rogers, E.M.: Diffusion of Innovations, 4th edn. Free Press, New York (1995)
41. Rogers, E.M.: Diffusion of Innovations, 5th edn. Free Press, New York (2003)
42. Rothaermel, F.T., Sugiyama, S.: Virtual Internet Communities and Commercial Success: Individual and Community-Level Theory Grounded in the Atypical Case of TimeZone.com. Journal of Management 27(3), 297–312 (2001)

43. Sánchez-Franco, M.J.: The Moderating Effects of Involvement on the Relationships between Satisfaction, Trust and Commitment in E-Banking. Journal of Interactive Marketing 23(3), 247–258 (2009)
44. Sánchez-Franco, M.J., Roldán, J.L.: Expressive Aesthetics to Ease Perceived Community Support: Exploring Personal Innovativeness and Routinised Behaviour as Moderators in Tuenti. Computers in Human Behavior 26(6), 1445–1457 (2010)
45. Sørebø, A.M., Sørebø, Ø., Sein, M.K.: The Influence of User Involvement and Personal Innovativeness on User Behavior. International Journal of Human and Social Sciences 3(5) (2008)
46. Tornatzky, L.G., Fleischer, M., Chakrabarti, A.K.: The Processes of Technological Innovation. Lexington Books. Massachusetts (1990)
47. Tu, C.-H., McIsaac, M.: The Relationship of Social Presence and Interaction in Online Classes. American Journal of Distance Education 16(3), 131–150 (2002)
48. Turner, J.W., Grube, J.A., Meyers, J.: Developing Optimal Match within Communities: An Exploration of CMC Support Communities and Traditional Support. Journal of Communication 51, 231–251 (2001)
49. Wellman, B., Guilia, M.: Net Surfers don't Ride Alone: Virtual Communities as Communities. In: Wellman, B. (ed.) Networks in the global village: Life in contemporary communities. Westview Press, Boulder (1999)
50. Woszczynski, A.B., Roth, P.L., Segars, A.H.: Exploring the Theoretical Foundations of Playfulness in Computer Interactions. Computers in Human Behavior 18(4), 369–388 (2002)
51. Yi, M.Y., Fiedler, K.D., Park, J.S.: Understanding the Role of Individual Innovativeness in the Acceptance of IT-based Innovations: Comparative Analyses of Models and Measures. Decision Sciences 37(3), 393–426 (2006a)

Characterizing Interactions among Members of Deaf Communities in Orkut

Glívia A.R. Barbosa[1], Ismael S. Silva[1], Glauber Gonçalves[1], Raquel O. Prates[1], Fabrício Benevenuto[2], and Virgílio Almeida[1]

[1] Computer Science Department, Federal University of Minas Gerais, Brazil
[2] Computer Science Department, Federal University of Ouro Preto, Brazil
{gliviaangelica,ismael.silva,ggoncalves,rprates, fabricio,virgilio}@dcc.ufmg.br

Abstract. This paper presents a characterization of interaction phenomena among members of communities aimed at deaf people in an online social network system, namely Orkut. The results revealed that members of deaf communities are tighter than members of other communities. However, analysis of the interface indicates that it does not always address their specific needs.

Keywords: Social Networks, Online Communities, Deaf, Orkut, Accessibility.

1 Introduction

The challenges for deaf people to access or interact with social networks are often taken for granted, since most social networks do not involve audio elements in their interfaces. However, it is important to notice that the first language for deaf people is sign language and communication takes place, most of the time, in face-to-face situations [14]. Thus, having to interact through the written form of oral languages may be a challenge for deaf users [4], [10], [14]. Furthermore, deaf culture has many specific features that may impact how deaf users interact through social network systems.

One known feature of deaf culture is that interaction with other deaf people is extremely important and communities tend to be tighter [14]. Therefore, online communities can represent a space of autonomy, expression, communication and expansion of their relationship networks [15]. Thus, motivated by the importance that offline communities represent in the lives of deaf users and the benefits that online communities can offer them, the goal of this paper is to investigate and present the characterization of their friendship network, as well as the interaction phenomena among members of Orkut communities aimed at Brazilian's deaf users.

Orkut was chosen because it is the most popular social network system in the country [6]. In Orkut, a community refers to a virtual space to connect people who seek to discuss issues, express beliefs, values and share ideas freely [17]. Members of a community do not necessarily establish a friendship relationship among themselves [16]. The community can be defined as public or moderated and its members may interact through forums, polls and events.

The results revealed that the friendship network formed by community members aimed at deafs in Brazil is more connected than the network of users in communities

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 280–287, 2011.

with different focuses. However, when analyzing the interaction among them, mediated by available resources in the communities (e.g. forums, polls and events), the results indicate that communication among the members of deaf communities have often not been performed through these resources. The analysis identifies aspects of the proposed interaction which are challenging for deaf users and that may lead them not to use the available communication mechanisms.

This work contributes to the understanding of how deaf users are interacting through network systems, like Orkut. This understanding is important to the design of network systems which can be more accessible and usable by deaf users, allowing social network systems to become more inclusive.

2 Related Work

Currently there is a large number of works which focus on analyzing and characterizing phenomena in social networks (e.g. [1], [2], [3], [16] and [19]). Also there are some works regarding how deaf people use Information and Communication Technologies (ICTs) (e.g. [10], [12] and [15]) However, to the the best of our knowledge, an investigation aimed at characterizing interaction among members of deaf communities on social networks has not yet been published.

Papers that focus on analysis and characterization of social networks usually study their topology and/or interaction. In [1] and [16] the authors conducted a study in order to understand the structural characteristics that influence the evolution of online communities and social networks over time, whereas works like [2], [3] and [19] analyzed the user interaction in social networks in order to characterize it and provide relevant information for the development of future applications.

In papers that address the deaf and ICTs, the focus usually is on how ICTs can support communication of deaf users among themselves and with listeners. Works like [10] and [12] found that, as listeners, the deaf also want to communicate with a variety of people and quickly, but what differs is the language used in this interaction. Authors in [15] describe the important role of ICTs in the autonomy and socialization of the deaf, and note that not all the specific characteristics of these users have been considered in the interaction models of these systems.

None of the papers we have found analyze the use and interaction of deaf users through social network communities. In this work we present an analysis of the deaf culture in the use of communities in a social network system. Our goal is to characterize the interaction among members of network systems' communities aimed at deaf users in Brazil. The scope was limited to Brazilian´s deaf users, since other works (e.g. [5]) have shown that culture has an impact on how people use social networks. Since in Brazil Orkut is the most popular social network, the investigation was done in this system.

3 Methodology and Data Collection

The methodology adopted to conduct this study consists of five steps: (1) definition of communities of interest, (2) data collection, (3) friendship network analysis, (4) use and interaction analysis and (5) characterization.

Orkut does not allow its members to identify themselves as deafs, even if they wanted to. Thus, in order to analyze the deaf friendship networks we investigated members of communities aimed at deaf users. Communities that explicitly declared themselves as aimed at deaf people through their names or description were considered for the analysis. Examples of the selected communities names are: "*Orkut Deafs in Brazil*", "*Yes, I am deaf*", and "*Deafs of Brazil Unite*". We selected 14 deaf communities. In these communities 12% (average per community) of their members spontaneously declared to be deaf in the "*name*" field[1], in their profiles page (range 10-14% with 95% confidence).

For comparison we also selected other communities that focused on different subjects. We selected 6 communities aimed at homosexual users, 6 communities that connect people who have the same surname and 9 random communities. We collected in total data from 35 communities classified into 4 categories. Each community had between 100 and 500 members and all more than 4 years of existence. The data was collected from October 16th to November 20th, 2010.

The chosen communities have between 100 to 500 members due to two main reasons. First, the fact that Orkut does not display the profile of all members if the community has more than 1000 members. Second, in order to compare the different communities they should have a similar range of members. Thus, this range allowed for a more accurate analysis of the desired communities.

For the friendship network analysis we collected the information of the profile and friendship network of 10.957 members, distributed in 35 communities described above. With this data, we verified the number of friends that each member contains in Orkut and how many of these friends are also in the community. This analysis allowed us to identify aspects related to intensity of relationships (i.e. connection) among community members.

The use and interaction analysis in communities was performed in order to complement the results obtained from the friendship network analysis. This analysis intended to identify what was the impact of this connection on the interaction among community members within the community itself. The method used in this step was the Semiotic Inspection Method (SIM) [8], grounded on the Semiotic Engineering Theory [7]. The choice of this method was due to its ability to provide a theoretical based analysis of the observed user-interface-user interaction as meant by the designers, as well as the actual interaction that took place among users of a community [8].

In the next section we present the characterization of the interaction among members of the deaf communities, based on the analysis of their friendship networks, as well as their interaction through the communities.

4 Characterization

Group interaction is extremely relevant to deaf people [14] and their communities tend to have tight relationships. Thus, our hypothesis is that online communities

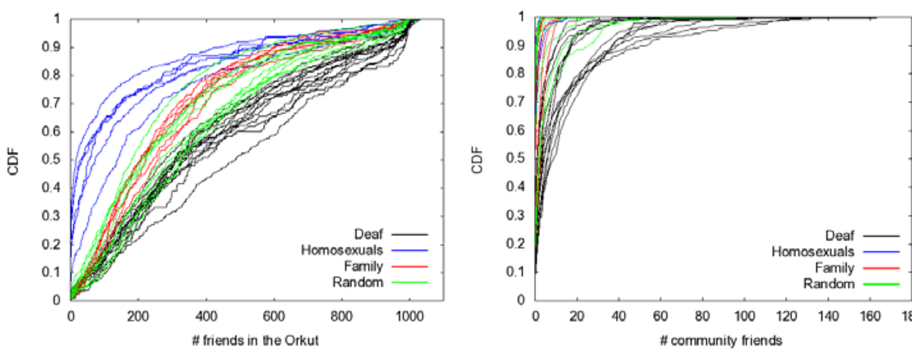
---

[1] Anytime the word "deaf" appeared in the "*name*" field the user was considered to be deaf. The information in the "*About me*" field was not considered since it would require a semantic analysis of the text.

composed by deaf users also tend to have tighter clusters than other communities. Therefore, in our investigation we intended to answer the following questions*: (1) Do deaf users have a closer relationship in online communities than other users? (2) If so, how does it impact their communication through Orkut communities?*

4.1 Characterization of the Friendship Network

The following graphics show the distribution of friends of the community members in Orkut, in their own communities, and the relation between these values. Each curve represents the Cumulative Distribution Function (CDF) [13] of an analyzed community: the X axis represents the number of friends of community members; the Y axis represents the cumulative fraction of members in each community with X friends. For instance, the point X=200 and Y=0.8 in a curve of the community "C" means that 80% of the members of "C" have at most 200 friends and that 20% of the members of "C" have more than 200 friends. The slower the CDF curves grow to reach 1 in the Y axis, the greater the number of friends of the community members.



(a) CDF of the total number of friends of the members of each community

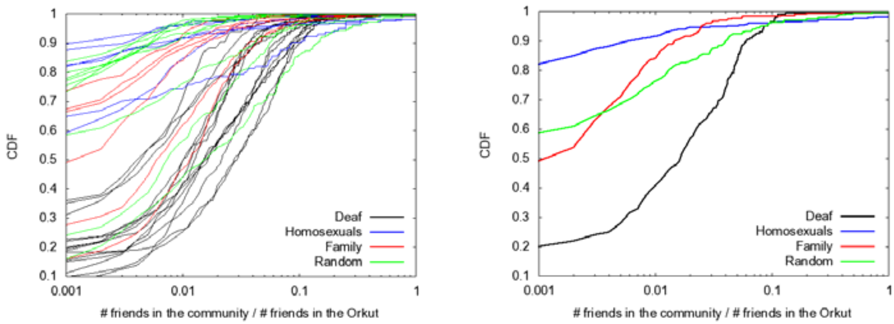
(b) CDF of number of friends within a community

Fig. 1. Number of friends that the analyzed Orkut community members have

The graphic depicted in the **Fig. 1(a)** presents the CDF of the number of friends that the community members have in Orkut and the graphic in **Fig. 1(b)** shows the CDF of the number of their friends who also participate in this community. We can note that among the 4 types of analyzed communities (i.e., deaf, homosexual, family and random selected), the members of communities related to the deaf have a higher number of friends in Orkut (see **Fig. 1 (a)**) and within their own communities (see **Fig. 1 (b)**). By comparing them, it is possible to see that members of deaf communities have more friends than members of the others (curves grow more slowly).

These results suggest that members of the analyzed deaf communities can be using Orkut to increase their friendship network. If they are also making new friends who are listeners, Orkut may be supporting the communication among deaf and listeners. This is an important finding, since face-to-face communication between deafs and listerners face the barrier of the different languages used [14], [10].

Next, we investigated the relation between the number of friends in the community and the number of friends in Orkut that each member of the analyzed communities has. These results are illustrated in **Fig. 2 (a).** To facilitate the visualization **Fig. 2(b)** shows only the most representative curves (i.e. median CDFs) of each analyzed group of community.



(a) CDF of rate of friends within the community that members of each analyzed community have

(b) Most representative curves of each group of community

Fig. 2. Relation between number of friends in community and total number of friends

Overall, we can note that members of deaf communities have a higher rate of friends within the community than the members of the other communities, since the deaf´s curves grow more slowly than other ones. For instance, while 60% of members of the most representative deaf communities have more than 1% of their friends participating in the same community, only 25% of random community members, 15% of family community members and 8% of homosexual community members have the same proportion of friends within the community. If we look at other points of this graphic **Fig. 2(b)**, the observed results will show that deaf users have a higher rate of friends who also participate in the same communities.

These results represent an evidence of a phenomenon among members of deaf communities called *homophily by selection* [9]. The establishment of relations motivated by *selection* indicates that people establish a relationship because they have "immutable" similar characteristics (e.g., disability or ethnicity) [9].

This observed phenomenon (i.e, *homophily by selection*) could be a reflection of offline behavior of the deaf community. According to [14] it is in the contact with their peers that deafs identify themselves and find stories similar to their own. The search for this deaf identity motivates the emergence of communities and associations which discuss the right to life, culture, education, work and welfare of all deaf [14].

The next section shows how this tighter relation among users of deaf communities impacts their use of the system.

4.2 Characterization of the Use and Interaction

The tighter relationship among deaf community members in Orkut raised the issue of how it impacted their interaction, that is, whether they used the available communication mechanisms differently or more often than other users. To investigate this issue the Semiotic Inspection Method [8] was used to analyze the possibilities made available by the system to its users; and how users actually use what is offered.

The analysis of what Orkut communities offered to users showed that there are 3 mechanisms – forums, polls and events – made available to whomever creates the community to decide which ones (if any) will be used by the community. All 3 mechanisms allow only for textual communication (no images or videos can be used). In communicating within a community, a member can only send messages to the whole community, no private messages can be sent. In order to check for updates in the ongoing communication, members must enter the community and access the thread of interest (users cannot request to be informed of updates).

These 3 aspects (text only, public messages and no informs) can be potential problems to all users. However, limiting communication to text can be a challenge for deaf users since their first language is usually sign language [4], [10], [14], which is visual and not textual. Thus, for deaf users there is an overhead of communicating in a language which is not their first language.

The analysis of the 35 communities considered showed that there were no relevant differences on the use of the communication mechanisms by deaf communities or the other ones. All the communities made available the 3 communication mechanisms (i.e., forums, polls and events) to its members (for creation and posting), but none of them had polls that had been opened in the year of 2010 or events listed for that year. Thus, the only mechanism actually used by community members was the forum. Looking specifically at the use of forums, in all communities new topics had been created in their forums in the previous month[2] . Nonetheless, the number of topics and posts were low[3], especially considering the number of members in these communities.

The analysis showed that although members of deaf communities have a tighter friendship network, this closeness does not impact the way they interact through the communities. In order to understand whether one of the reasons for this was the limitation to only communicate through text, we analyzed how members of the deaf communities who spontaneously declared themselves as deafs communicated in their personal scraps page (an environment outside the community).

Through this inspection it was possible to notice that deaf users receive many *scraps*[4] from their friends and the messages usually included videos, images or text. Most of them contained animations and short messages. The scraps that were mainly textual tended to be short (at most 2 sentences, each containing about 5 words). Another aspect that came to our attention was that often spam content received was not deleted from their profile. These spam messages were usually long texts, and it is probable that deaf users could not make sense of them and chose to leave them there.

Our analysis allows us to raise two possible explanations for the fact that members of deaf communities have closer relationships in the friendship network. The first is

---

[2] Analysis was performed from 18th to 25th of November, 2010.

[3] In the 2 months previous to the analysis, only 4 communities which had new topics added to the forums in that period received more than 10 posts relative to the new topics.

[4] An Orkut functionality that allows one user to send a message to a friend in which different media can be used.

that it is really a cultural trace of deaf communities, in which members tend to associate themselves to people who are similar to them (as suggested by our analysis of the networks). The inspection of the interface allows us to raise another hypothesis which is that due to the limiting communicative means in the communities, when members meet each other, they also include these new people as their personal friends in order to communicate with them more efficiently. Determining what role each of these possibilities in fact plays in the phenomenon of deaf communities being tighter than the others must be further investigated.

5 Final Remarks and Future Work

In this paper we investigated, analyzed and presented a characterization of the interaction among members of 14 Orkut communities aimed at connecting Brazilian´s deaf users. The research of related literature led us to pose two questions that were investigated: *(1) Do deaf users have a closer relationship in online communities than other users? (2) If so, how does it impact their communication through Orkut communities?*

The answer to the 1st question was obtained by the analysis of the deaf communities' friendship network, which showed that members of these communities are more closely connected than those of other communities with different focuses. To answer the 2nd question we inspected Orkut community interface, as well as how members of deaf communities used them. This analysis showed that no relevant differences could be noticed between the interaction of members of deaf communities and those of the other communities considered.

The analyses allowed us to raise two possible explanations to the phenomenon of deaf communities having a closer connection: (1) it is a cultural trace of deaf communities; (2) it is caused or fostered by interface decisions that limit possibilities of communication through community mechanisms. The role played by these possibilities has to be further investigated. However, the identification of the closeness of the deaf communities, as well as the new raised questions contribute to HCI research since it characterizes relevant aspects of the interaction Orkut community members aimed at deaf users in Brazil. Moreover, the methodology adopted can contribute to researchers interested in investigating the use of network systems by users with deficiencies or other special needs.

The next steps in this research include interviewing deaf users of Orkut communities; investigating how quantitative analysis along time can help us better understand if there is a specific relationship pattern associated to this deaf members' closeness; and analyzing deaf users´ relationships in other network systems to identify what observed aspects are culturally determined, as opposed to those that are a consequence of the interaction offered by the system.

Acknowledgments. This work is partially supported by the the INCT-Web (INWeb) (MCT/CNPq/ grant 57.3871/2008-6), and by the authors' individual grants and scholarships from FAPEMIG, CNPq, CAPES and UOL.

References

1. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. In: Proc. of ACM SIGKDD (KDD 2006), pp. 44–54. ACM, New York (2006)
2. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: Proc. of ACM SIGCOMM. IMC 2009, pp. 49–62. ACM, New York (2009)
3. Burke, M., Marlow, C., Lento, T.: Feed me: motivating newcomer contribution in social network sites. In: Proc. of CHI 2009, pp. 945—954 (2009)
4. Cavender, A.C., Otero, D.S., Bigham, J.P., Ladner, R.: Asl-stem forum: enabling sign language to grow through online collaboration. In: Proc. of CHI 2010, pp. 2075–2078. ACM, New York (2010)
5. Chapman, C.N., Lahav, M.: International ethnographic observation of social networking sites. In: Proc. of CHI 2008 Extended Abstracts, pp. 3123–3128. ACM, New York (2008)
6. comSCORE.: Orkut Continues to Lead Brazil's Social Networking Market, Facebook Audience Grows Fivefold (2010), `http://goo.gl/amrO0`
7. de Souza, C.S.: The semiotic engineering of HCI. MIT Press, Cambridge (2005)
8. de Souza, C.S., Leitão, C.F., Prates, R.O., Bim, S.A., da Silva, E.J.: Can inspection methods generate valid new knowledge in HCI? The case of semiotic inspection. International Journal of Human-Computer Studies 68(1-2), 22–40 (2010)
9. Easley, D., Kleinberg, J.: Networks, crowds, and markets: reasoning about a highly connected world. Cambridge University Press, Cambridge (2010)
10. Fajardo, I., Abascal, J., Cañas, J.J.: Bridging the digital divide for deaf signer users. In: Proc. of ECCE 2008, vol. 369, pp. 1–6. ACM, New York (2008)
11. Foucault Welles, B., Van Devender, A., Contractor, N.: Is a friend a friend?: investigating the structure of friendship networks in virtual worlds. In: Proc of CHI 2010 Extended Abstracts, pp. 4027–4032. ACM, New York (2010)
12. Henderson-Summet, V., Grinter, R.E., Carroll, J., Starner, T.: Electronic communication: themes from a case study of the deaf community. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 347–360. Springer, Heidelberg (2007)
13. Jain, R.: The Art of Computer Systems Performance Analysis. Wiley Computer Publishing, John Wiley & Sons, Inc. (1991)
14. Lane, H., Hoffmeister, R., Bahan, B.: A journey into the Deaf-World. Dawn Sign Press (1996)
15. Martins, S., Filgueiras, L.: Evaluating interaction models for communication of deaf users. In: Proc of IHC 2010, pp. 193–196. 4 pages. SBC (2010) (in Portuguese)
16. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proc. of IMC 2007, pp. 29–42. ACM, New York (2007)
17. Orkut.: Help (2010), `http://www.google.com/support/orkut/bin/topic.py?topic=29543`
18. Preece, J.: Online Communities: Designing Usability, Supporting Sociability. John Wiley & Sons, Chichester (2000)
19. Wilson, C., Boe, B., Sala, A., Puttaswamy, K.P.N., Zhao, B.Y.: User interactions in social networks and their implications. In: Proc. of EuroSys 2009, pp. 205–218. ACM, New York (2009)

The Role of Music in the Design Process with Children

Ruut Tikkanen and Netta Iivari

Department of Information Processing Science, University of Oulu
P.O. Box 3000, 90014 Oulu, Finland
{Ruut.tikkanen,netta.iivari}@oulu.fi

Abstract. Music and other art based methods should play a significant role in the HCI field when designing with children, taking the developmental stage of the children into consideration. Music has been neglected in the design process in HCI research, while there is a lot of meaningful research in social and educational studies. HCI research has concentrated more on technological products and technological needs of special education, bringing up music as an important part of media too. In this paper we emphasize the versatile role of music during the design process with preschool children. We describe three different workshops identifying three different roles for music: a contextual role, music as a trigger and music as content. The roles demonstrate numerous possibilities for using music in design workshops and show the usefulness of music in collaborative design with children. HCI research should utilize music in more varied forms during design.

Keywords: Preschool children, music, design process.

1 Introduction

There are a number of different elements, which should be considered in the child centred interface design. The Human Computer Interaction (HCI) literature concerning children contains a number of perspectives. The literature maintains that children differ from adults. Their world is different. Their skills are still in the developmental stage. On the other hand, children's imagination can offer for adults something that they themselves could not imagine. (See e.g. [9], [10], [12], [20], [42].) Although the development of a child is an individual process, development areas such as motoric, cognitive, social and emotional development can be identified. For example, according to children’s cognitive developmental stage theory (see e.g. [36]) children from 5 to 6 years old live early childhood and are in a preoperational stage. Particularly play and imagination as well as different kinds of activities with others of the same age have a big meaning in children’s life. According to Hanna and colleagues [20], children of this age are happy to show you what they know, and what they can do on the computer independently. When assessing appeal or engagement, it is important to observe children’s behaviour such as sighing, smiling, or sliding under the table. Children of this age often have difficulties in expressing their likes and dislikes in words. [20.]

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 288–305, 2011.

Methods used with children usually differ from those used with adults, although some methods intended for adults are applied also with children. It is, nevertheless, recommended that the methods should be such or modified so that they are appropriate for child participants and their developmental stage (e.g. [11], [12], [18], [20], [30], [31], [42]). In the literature on child's development one of the central ideas, from Piaget [36] and Vygotsky [44] is that children perceive the world by playing and singing. That is one scientific basis for using art-oriented methods, which are considered to provide interesting dimensions to user interface design with children (see e.g. [17]). Only very few articles have touched upon the role of music in the design process with children. Furthermore, the studies describe the design process with children, the music being related to the product being designed, not to the design process itself (see [3], [32] [33]). In this study, instead, music plays an important role during the design process. We also argue that music should be utilized in technology design more broadly. Design processes remain only too silent.

This paper examines the variety of roles music may have in the design process with children. This study is an exploratory study of the roles of music, revealing some of those, while more rigorous experiments on the effects of music can be carried out later on, based on and extending the results of this study. The topic is examined in a research project developing a music application for children, together with children. In the research project, numerous innovative workshop sessions have been organized with children (see e.g. [32]), music being in different roles in the workshops. In this paper we describe three different workshops, through which we outline what kind of meanings music may have in the design process with children. To summarize; music had mainly a contextual meaning in the first workshop. The aim was to determine what kinds of ideas children have about musical instruments and producing sounds: what do they build, if they are free to compose their own musical instrument or device? In the second workshop music was combined with emotions. Two music samples were played for children and they drew their thoughts and feelings, listening to music in the background. In the third workshop music had a substantive role. The purpose was to find out what kind of sounds children connect to different pre-given landscapes of the music application being designed.

The paper is structured as follows. The next section reviews the extant literature on children, technology and music. The third section present the research setting and procedure involved in this study. The fourth section outlines the roles of music we identified in our design sessions. The final section summarizes the results, outlines their implications and limitations and identifies a number of paths for future work.

2 Literature Review

A short review to different journals and articles on children's technology and music field illustrates that music and children is an important research topic in connection to technology. Technology has provided an opportunity to revise working with music and music teaching already in 1990s. Technology like computers, synthesizers and hypermedia seems to be conceived like a new tool in relation to old craft [43].

There are only few music related articles concentrating on describing design process with children in this area. Bar-Ilan and Belous [3], Mazzone and colleagues [32] and McKnight and Read [33] enlarge this topic from three different viewpoints. In the experiments of Bar-Ilan and Belous [3], children act as architects of web directories. The authors advise researchers to include children in the design process, not only designing the interface but designing the content structure as well. McKnight and Read [33] concentrate on the design of a record button, their work being part of designing a musical interface with children. Mazzone and colleagues [32] create a rich background on a general level related to designing with children in the context of a music application, because they have actively worked with children during the design process, but they do not concentrate on the meaning of music in the process. Although music is mentioned in these experiment descriptions, it appears more in the subtext of the design research with children. Also Hartnett and colleagues [22] describe the design and development process of an audio device to stimulate exercise and fitness in children, but the device is designed for children, not with children. None of the articles discuss directly music. Music is addressed in connection to the product, not during the design process.

On the other hand, there is an extensive body of research addressing music, technology and children more generally. This research can be broken up into different areas. First of all, children's technologies including different kind of software and web-based-systems seem to be central in the scientific discussions as well as in practical work. Children are widely perceived as users of different kinds of music software. Music is one area of concern when presenting children's future technologies. There are several kinds of music software, in most cases educational software, discussed, aimed at e.g. for composing [6], [24], for making music [16] for having fun [40] for playing [1], [6], [39], [40], [45], for moving [2],]22], [46], for interactive communicating [1], [2], [7], [27], [46], for exercising and learning [16], [35] or for performing [45]. There are also several media systems, where music is a big part of the technological entity. Music is also related to other arts or creative media like painting, dancing or architectural planning [2], [3], [40].

Furthermore, music and children are increasingly in the focus of human brain research concerning, for example, music, speech and language processing [25], [26], [29]. Especially children's musical development is taken into consideration in these studies. Closely in relation to technological brain research there is also a research domain of music therapy, which concerns different kinds of technological solutions for the impairments of children in the area of speech and music [5], [21], [25].

Music is generally treated as an important matter. It is a central focus in many research papers as such, or treated as one part of media [27]. For example, the central focus of the research by Beck [4] is the confidence of children to be achieved through nurturing them with music. The significance of music has been showed in many ways and it has significance in many kinds of technologies, although they are not primarily music technologies. Music and its influence to the human brain is an important part of children's development. For example Bresler [8] and Rusinek and Rincon [41] use musical performing when presenting the importance of music. The purpose of most technologies is to support in an appropriate way certain stages of child's development.

Although there is some research found concerning design process or part of it carried out together with children [3], [32], [33], the designed music product has been the main issue in most of the studies.

Table 1 summarizes the literature review results.

Table 1. Different meanings of music in connection to children and technology

Aspects	**Music and Design**	**Children's Music Technologies**	**Music Education**
RESEARC-HERS	Bar-Ilan & Belous [3], Mazzone et al [32], McKnight & Read [33]	Akiyama [1], Aucouturier & Ikegami [2], Berry et al [6], Bobick et al [7], Fischer & Lau [16], Hartnett et al [22], Jennings [24], Li [27], Miura & Sugimoto [35], Reynolds [39], Rosenbaum & Silver [40], Weinberg et al [45], Zigelbaum et al [46]	Beck [4], Benveniste et al [5], Bresler [8], Fischer & Lau [16], Harris & Summa-Chadwick [21], Jennings [24], Jentschke et al [25], Koelsch et al, [26], Magne et al [29], Rusinek & Rincon [4]

Why music and sound should be introduced as an important issue in the design process? Best knowledge is going to be found undoubtedly from the area of music research. Several studies especially in the area of music education argue for the significance of music for children, being based on the developmental research theories. According to Flohr [17] sound is among the earliest stimuli to have some form of meaning for the human being. According to him, music clearly belongs in the young child's life as a distinct discipline and as an art form. The value of music for children has been divided into intrinsic and extrinsic musical benefits. According to several music educationis [14], [15], [34], [37], [38], the intrinsic benefits of music education include a big amount of important aspects for children's development such as self-expression, as well as emotional expression and aesthetic enjoyment. The extrinsic music benefits for young children include fostering motor development, promoting cultural heritage, providing release of tension, teaching language development, using music as a carrier of information or as a cue for naptime, and using music to help teach other subjects such a reading or mathematics [17]. These characteristics support the idea to take music into focus in the creative design process.

When thinking especially the role of music as a trigger, there are lots of supporting theories for this role, especially for listening to music. Firstly, listening is considered as an important musical experience [13], [17]. According to Dura [13] musical experience is different from visual. Music is heard within the entire body [13]. According to him, music-listening experience is one of immersion in sound. Secondly, listening has a comprehensive character, children are dedicating to music, body and soul. According to Flohr [17] listening skills permeate all music activities [17]. Regarding research on music listening, it is important to note that when researching children it is often difficult to tell with certainty if a child is actively listening, but usually this can be determined by the child's movements or expressions [17]. Regarding the results that can be gained from children when listening to music, in a study of preschoolers it is argued that social interactions increased after music listening [19].

Even though our literature review presented three roles for music in connection to children and technology, instead of the whole categorization, we are going to focus only on the ‘music and design’ part of it. In this article the purpose is to concentrate on the roles of music in the design process and describe all the benefits we got during the design process when using music.

3 Research Setting and Procedure

The research setting is a multidisciplinary, international research project developing a music application for children, together with children. The researchers in the project have expertises in the information technology (IT) field, HCI research and/or music education. The target user groups of the music application are preschool (aged 3-6) and school-aged (6-12) children. The application in question is a learning application including games for singing, composing, instrument playing and improvisation.

The initial project objectives and the requirements for the music application were based on music education expertise. HCI experts entered the scene through evaluating the initial requirements and designs produced by the music educators, and through encouraging children to provide feedback as well as to create new ideas and designs for the application. Therefore, the HCI experts invited children into the design process as testers, informants and design partners [12]. This paper focuses on the design activities carried out during the project’s first year of functioning in Finland, with the age group of 5-6 years old children. These activities were initially defined by a HCI research group in the University of Central Lancashire, the HCI research groups in Finland and UK both independently and collaboratively experimenting with and refining the activities (for more information, see [32]). The requirements for the activities came partly also from the music educators, who wished to gain insights and feedback related to their initial ideas and designs. The design activities examined in this paper were carried out in one nursery in Finland with the child group consisting of 9-13 5 and 6 years old children.

The design activities included the activities of “The Magic Music Toy”, “If I were there I would play this” and “Cheers me up, puts me down” (for detailed description, see [32]). The “Magic Music Toy” activity allowed children to design their own magic music toy, first drawing it, and then producing a prototype of it with art craft materials. The activity aimed to understand children’s concepts of music playing and to gather their ideas related to an ideal music toy. In the “Cheers me up, puts me down” activity, on the other hand, children listened to music and at the same time drew the feelings that the music triggered. This was done with two different music samples. The aim of this activity was to examine how children were able to relate music with emotions. This was requested by the music educators who planned to include a related feature to the music application. Finally, “If I were there I would play this” activity was about evaluating and inquiring how children relate music and places, which was also requested by the music educators since they planned to place different music samples to different landscapes in the music application. Children were asked to refine the original landscape pictures by adding instruments, sounds and any kind of general sensations they associated with the landscape. At the beginning of each workshop the researchers played and sang with the children and checked the situation of the day. Afterwards, the children were given the assignment. During the workshops the researchers walked around the room, discussing with the children and encouraging them if needed. At the end of the workshops, the children were interviewed.

All these design activities not only contributed to the specific research question mentioned in connection to them, but they all also produced a lot of useful and inspiring material to be utilized in the further application design: e.g. the children's drawings and prototypes provided highly useful information and ideas. The children were also continuously observed and interviewed during several occasions during these sessions, and their stories provided valuable insights for the research team.

All in all, during the design activities a lot of research data was gathered. The material was gathered by observing the children during the activities, by interviewing them during the work and afterwards, and by saving all the drawings and prototypes created by the children. The material includes video recordings, field notes made by the research team, photographs and the artifacts created by the children. In addition, the research team members were interviewed later on.

The data analysis concerning the meaning of music during the design process included a rich amount of material. The analysis process was data driven and inductive. The roles of music emerged from the data, i.e. there were no predefined hypotheses tested nor control groups involved to verify the effects of music, but instead this was an exploratory, data driven study of the roles of music. More rigorous, quantitative experiments can be carried out later on, based on and extending the results of this study. The data driven analysis process in this case started through collecting together all the information concerning sound and music and the equipments for producing sound from the video transcriptions, interviews, research notes and diaries, and children's productions including descriptions. The analysis involved expertise both in HCI and music education. All the data were connected together through careful analysis and transforming the findings into textual form and placing them into tables. The three roles of music inductively emerged during this process. Related to each role, supportive evidence was recorded. The results will be presented in the following empirical section.

4 The Role of Music in the Design Process

This section will demonstrate the meaning of music as context, as a trigger and as content in the design process. Although music and sound are important part of almost all software, the product being a music application increases the importance of music, especially in relation to designing content. The following descriptions will concentrate on illustrating the three different roles of music; therefore many details related to the whole design process in the examined research project in question are omitted. The first workshop was organized to have a general view about musical conceptions and children's musical world. It was like an inspiring start for the process. In the second workshop the purpose was to wake up the children to draw their sensations through carefully chosen music. In the third workshop, children were asked to fill up landscapes and at the same time to create musical material and soundscape for the user interface of the composition game part of the upcoming music application. Next the roles of music are discussed, as identified from these three workshops.

4.1 Music as Context

In this paper 'context' refers to the children's surrounding musical environment and to the concepts that clarify the meaning of music and music application among children. In a sense, it can be connected to the old claim of HCI research and practice to 'know the user and her context of use'; i.e. to understand the users' skills, knowledge and characteristics and the (physical, technical, social, organizational, cultural) context of use (e.g. [23]). In the first workshop the idea was to figure out the children's musical context and to find out their own music conceptions, preferences, abilities and knowledge, as well as to discover how they perceive the concept of "magic music toy": How their own magic music toy would look like? What kind of music and sounds would it produce? What would they like to do with it?

Sketching and making a prototype of the music toy was very inspirational work for the children. When observing their behavior (see e.g. [20]) they seemed to be enthusiastic with recycling materials and familiar with the topic of a music toy. The children also had clear ideas about their music toys and how they should look like. When asking the children questions about their sketches or devices they corrected erroneous impressions and had detailed explanations of the function and the structure. They were influenced by those instruments they had played or seen in the nursery or at home. Although the original form of the magic music toys was based on basic instruments like glockenspiel, drums or local instruments, the children added a lot of different kind of details and also ideas from the unmusical world like petrol and tank or the idea that instruments could walk or fly. Their ideas had even human faces and body combined with instruments like a man zither. Therefore, we claim that their ideas were much richer and imaginative than adults.

The children also had clear ideas on how to produce sound with the music toys; on the techniques how to play them and on the kinds of possibilities one has for producing sound. They had parts from wind instruments, string instruments or percussion. Although there were no keyboard instruments found in the children's prototypes, they used a lot of touchable buttons: "*Over here you should give a whistle and at the same time you should press the buttons*". They showed to the researchers how to play: "*like this*". Some of the children showed by voice what kind of sounds they produce. Children uttered different sound like: "*pam pam, ti tit tii, tin tin*". They described also very imaginative sounds such as white and black sounds coming out from zither. According to one boy some of the sounds are still secrets: "*You can listen the most funny sounds of the world, but you can not know what kind of they are before you have played and heard them*" Another boy added: "*You can just guess them*". The children had also clear ideas on what kind of equipments one needs for producing the sound. In addition to their hands and fingers, the children played with their feet and elbows, even guitar was played by heels. They had also additional equipments to produce the sound like a mallet. Those instruments were also designed for social playing: "*the other person is whistling and other is pressing the buttons at the same* *time*". Also Godeli and colleagues [19] have shown that music increases social behavior.

The children were also familiar with some musical concepts. They used during their work some concepts, which referred to volume (louder-softer), pitch (higher-lower) and timbre, because of different sources of sound like strings, music pipes and drums. One example of timbre changes is in the black and white music of the zither. There was one girl who had a mechanism for producing sounds from different pitch levels and one boy who had a bass drum. Some children had clear volume buttons: "*here you can put softer, and here louder*". They had a conception about sound volumes "*here you listen the sound and stop the sound, and here you can get the sounds quieter and here you can get them louder*".

This workshop showed the children's ability to produce, by themselves, a logical completeness related to a music toy, which was constructed through a sketch, a prototype and the child's functional presentation. Children's imagination and playful stories about those constructions differ from those that would have been created by adults. It was very natural for the children to connect, for example, a musical instrument and an element from their daily life, like a starting motor. In their developmental stage it was also very natural to connect some human elements to the instruments, like a face, eyes, mouth or the whole body. Music offered an extra dimension for the work. Different kind of resounding methods and musical world inspired the children to create fictitious combinations of different elements. Their fantasy produced much more than the research group could note. According to a nursery maid this task has lived a long time in the games and plays in the nursery.

Musical context was inquired also in the second workshop related to emotions and music. When listening to music in the second workshop, the children drew a lot of familiar elements that came to their mid when listening to music. Children's intuitive visual representations of the listened music contained general contextual information about children's intuitive responses. Their drawings showed the musical world surrounding them. They drew what kind of instruments they had and played at home or in the nursery: "*I drew a violin, because I have a violin at home*" When interviewing the nursery staff, it was revealed that the children drew a lot of instruments they played in the nursery, like local instruments or rhythm instruments. The listening task showed also the children's ability to recognize different instruments from music. One six years old girl in particular recognized special rhythmic instruments like a quiro and a cabasa and blow instruments like a corn. In addition to real elements, the children also created imaginative elements through listening to music. Although some of the children had very imaginative, creative stories about their creations, most of the children draw, in addition to the instruments, familiar things for them like family members, home and nature.

4.2 Music as a Trigger

Trigger in this paper refers to an impulse, whose purpose is to support or shore up the design process, in this case children's creation through drawing. In the HCI field there seems to be a lack of research related to this area. Most references are from the educational and social studies. When describing music as a trigger, music is seen as a trigger for design. In this case music listening is intuitive and listening is like an experience [17] that the children are expressing mostly through drawing and unconsciously also for example through dancing.

In the second workshop on music and emotions, children listened to two different kinds of music samples and draw all the sensations that the music had given them and all the things that came to their mind. The music was chosen carefully, considering the developmental stage of the preschool children [17]. Even though children need to be considered as individuals, there are general principles concerning the developmental stages. For example, it has been argued that 5 to 6 years old children react to tempo changes and volume in music [17], [36], [44]. The first music sample was changing musical tempo from fast to slow, back to fast and slow again and so on. The other sample was changing from happy to sad music and back. All the children liked the music samples when asked about them: ”*music was fun*”.

All children were very keen on listening to music and drawing. When working with the children and when analyzing the video materials, in addition to settling down to listen, the children seemed to concentrate very well on drawing when listening to music. Music gave clear starting and ending points for the drawing task when compared to the drawing task that was carried out in the third workshop without music. The children started immediately and put their souls fast into creating the pictures. Engaging children's musical minds in an intuitive way seems to be a profitable way to produce material. The children also concentrated for a relatively long time to the task. When compared with the third workshop, in which the children were drawing on the prefilled landscape pictures, it seems that the children did not concentrate to drawing for so long time without music. When listening to music, the children worked intensively the whole time. They did not ask: "*when we are going to stop?*" They did not seem to be tired.

Like Flohr [17] describes, it is often difficult to tell with certainty if a child is actively listening, but usually this can be determined by the child's movements or expressions [17]. Although children were asked to sit down and to draw, one boy started to dance when he heard music. According to Dura [13], dance should be taken as an enticement to bodily listening. According to him, musical experience is different from visual, and music is heard within the entire body [13]. There were lots of examples found in the workshop authenticating those influences of music. Although children were asked to listen and to draw silently, there were also some children moving their feet with music or imitating the pipe sound through making different expressions with the mouth. One of the girls took immediately both hands in front of her mouth and like played a pipe with them when listening to march music. One of the boys slid his pen through a tunnel he had drawn at the same time when listening to the music. The movement of the music led one boy to draw a water wall and a bags throwing machine. Two of the children were drawing different windows. Perhaps this could also be interpreted through the rhythm of the music.

The workshop also showed that music supports children's creative imagination. Some drawings could be interpreted in a way that music gave some rhythm for those drawings, like for the machine which was throwing bags that was created when listening to march music. Two of the boys were drawing ways for the music: there was the water wall, the tunnel and a labyrinth. They also put sounds coming from the elements, like from the water wall or the machine: "*machine is making the sound and the bags are flying and they join to the other picture*" In addition, when being ready with drawing, as mentioned, one boy moved his pen through the tunnel when listening to the music. However, the music giving rhythm for the creations is a matter that requires more research. In addition, there is not that much evidence for understanding the influence of musical mode: the changes from happy to sad and back to happy. When listening to sad music some children drew also sad music like a sad violin. However, for most of the children sad music seemed to be happy. Most of them did not recognize the sad part of music. When asking children: "*music was happy*".

When having music as a trigger in the design process, individual creativity should be taken into consideration. It does not function in a given time. Although music gave also the ending point for the task, some children should have had more time for listening and drawing. Time refers to the moment-to-moment changes that occur within and between musical events, creating an experience of time as "lived or virtual time" as opposed to "real or clock time" [13]. In addition to time for drawing, there should be time for telling about those drawings and discussing then with adults. Music inspired some children not only to draw, but also to create imaginative stories. Often in the opinion of an adult those stories do not have very clear logic, but for the children the ideas are very clear and they just need time to explain them to the researchers in the discussion or in some kind of child's own presentation: "*these are views, landscapes [shows a tractor and place where people are living], between them there is a water wall, which is at the same time a machine…*" The boy was continuing his story through playing "*smoke, put, put, put [showing with his pen the way of smoke coming out from the tractor pipe]"* and he would have continued his story even for a longer time if he had been given the possibility. The children also liked to tell their stories to the researchers. Music was really like a trigger or impulse for creating material from their imagination.

4.3 Music as Content

Content means the musical substance, which is going to be contained in the given surrounding in the upcoming music application. The meaning of the content is significant in every design project. In this project the purpose was to produce suitable material for the composition part of the music application for preschoolers. Especially when designing a music game, music plays a big role during the process. The original requirements of the game design included three different landscapes: the jungle, the castle and the city. The children were asked directly to draw and add different kinds of sensations to the ready drawn landscape backgrounds. One important point of the task was to inquire the musical content; to note all the musical sounds children were combining with the pictures that were to act as user interface pictures in the upcoming music game.

Although the jungle is not a familiar surrounding for the children in Europe, they had well-defined opinions on what kind of sounds there should be. Some children liked silence, but most of them combined different nature sounds like wind, water, brooks, bird singing and trees to the picture. They added also instruments like a flute, a trumpet and a drum to the jungle. One of the girls heard ballet music in the jungle. They also described the character of the sound, how the sound was coming out: trees were vibrating, wind was blowing, and water and brook were producing sounds too. Animals were also very lively. The monkeys were shaking coconuts, and birds were singing. Animals were behaving like humans: ”*Jee, I got the banana*”. The children again created very imaginative picture combinations. One girl told that there was no sound in the picture, although a butterfly was carrying string instruments in her tail. There were also trumpets and flutes combined with the jungle.

In the castle and the city landscapes the sounds of the musical world were poorer, likely because of the children's ability to concentrate. Completing two pictures was fine for the preschool children, but three was too much. In the castle surrounding, a favorite source of sound was a ghost. They were hooting in almost all pictures. Children were putting themselves also in the picture. Humans were uttering sounds like: “*iik, help, böö, uuu, ooo*”. In the surrounding it was raining and thundering. There was also steam coming out from fire. Some of the children added also trumpet sounds to the picture. One of the boys thought the castle to be a church and he heard a priest speaking in a microphone and a lot of different instruments like a guitar and a cello.

In the city surrounding the most popular sound was the car. There were different kinds of cars whirring in the children's drawings. One boy had written “*milk shop car*”. However, the children could not concentrate any more to the third picture. This was seen, for example, in the mixing of the ideas of the different pictures. For example, there were also coconuts in the city picture and those drawings were less detailed than before. Despite of this, the children found celebrating sounds and speaking people between apartment houses. There were also a thunderstorm and fire in the city. Some of the children found violin sounds and one of them told about secret sounds, but he could not tell what kind of sounds they were. Someone could hear “*surr*” in the picture.

This workshop showed the remarkable role of music in designing content, especially for a music game. Although this workshop did not include any live music, music and different sounds were an important part in those drawings. Especially all this information is important when the game designers are producing the preliminary sounds for the game. The children found natural sounds for the surroundings, like trees and brooks. They also found imaginative sounds like speaking animals and hooting ghosts. In addition to different instruments, there were a lot of interesting sources of sounds in the pictures like cars, animals, different kinds of people, even a singing flower, which had an influence on a color changing tree. There were also silent instruments, like those zithers in the tail of the butterfly.

In the first workshop the children also created sounds. They produced different kinds of sounds when creating and demonstrating their magic music toys. In addition, as mentioned, they told about secret, unheard sounds: according to one boy some of the sounds are secrets: “*You can listen the most funny sounds of the world, but you can not know what kind of they are before you have played and heard them*”, another boy adding: ”*You can just guess them*”. These sentences bring again up the imaginative perspectives that adult designers probably would not think of.

5 Discussion

This paper examined the variety of roles music may have in the design process with children. Table 2 summarizes the results of our empirical examination. The results are presented according to the three roles of music we have identified, but also according to the insights gained through the use of music, categorized to two viewpoints: the insights gained related to the music application to be developed, and the insights gained in relation to children.

Table 2. The roles of music in the design process with children

Role of Music	Context	Trigger	Content
Application	-appearance and structure -functionality -sources of sound -playing techniques and additional tools and instruments -quality of sound and timbre	-extraordinary, surprising viewpoints	-sounds from the imaginative and real world -imaginative combinations of sounds -secret sounds -silence
Children	-familiar music applications, instruments, sounds, playing techniques -musical concepts: volume (louder-softer), pitch (high-low), timbre -imaginative elements and combinations from the children's world - important issues in the children's life (in addition to music) - social behavior related to music	-giving information about children's intuitive responses -enabling musical experience -showing expressions, in addition to drawing, through moving and discussing - supporting children's developmental stage and imagination - supporting concentration and intensive working - giving structure for the task	- ability to create and describe sounds to a given surrounding - ability to create imaginative musical worlds - ability to express clear opinions - problems in concentrating for very long

Through the use of music the researchers gained inspirational background information related to children's favorite music toys and applications, and directly utilizable musical material for the music application as well as useful, inspiring material for the user interface design. One interesting detail in the results is the issue of secret sounds that even the children themselves had not heard or created yet. Clearly, a lot can be learned from children when designing together with them, not only music applications, but different kinds of applications in general. Related to the children, the researchers gained a lot of insights of their musical world, of the concepts they use, of their imagination related to music, and of their preferences, abilities and knowledge. The results revealed the importance of musical experience also in the context of IT design. It aroused children's sensations and imagination. In addition, the researchers learned that the children were able to create imaginative musical worlds and to create and describe sounds to a given surrounding. They were able to refine the pre-defined musical landscapes and to express their opinions about the musical content. It was also learned that their ability to concentrate is quite limited, but music acting as a trigger may diminish, to a certain extent, this problem.

The workshops carried out also had a clear impact on the design of the application (see also [32]). All the findings and ideas that could be somehow utilized in the application design were gathered from the research material, and discussed with the developers and with the persons responsible for the application design. It was revealed that the material collected while working with the children contained ideas regarding the structure of the application and the design of certain icons, and even some pictures that could be directly used in the user interface design. Paper prototypes were afterwards built, utilizing numerous ideas gained from these workshops, and evaluated together with children. Thereafter, the developers built functional prototypes relying on those evaluated designs. Also the actual implemented application still contains ideas derived from children during the workshops.

All in all, the different roles of music we identified demonstrate numerous possibilities for using music in design workshops and show the usefulness of music in the collaborative design process with children. Especially music should be utilized while designing music applications. However, we argue that HCI research, altogether, should be broadened to utilize music in more varied forms, relying on the already established interdisciplinary results related to music. So far, HCI research has focused on the design of interactive products emphasizing their usability, fun and learning related aspects. The numerous design methods created for working with children within the HCI field do not pay any attention to the use of music. However, important to note is that in our research project there are also music education specialists involved. The utilization of music in the design process requires the participation of the domain specialists related to music. However, also HCI researchers played a substantive part in organizing the workshops and in analyzing the results. We recommend multidisciplinary collaboration for the researchers interested in utilizing music in their design sessions.

We bring up four important matters connected to using music in HCI research. Firstly, music constitutes a big part of the life and education of children. Music gives strength for a child and his/her actions and development. Music clearly belongs to the young child's life, like Flohr [17] describes. The meaning of sound and music comes up already in the prenatal phase of human being [17]. Developmental psychologists such as Piaget [36] and Vygotsky [44] underline the value of music as a supporting element in children's development. Music is a natural part of children's life and education, like this research connected to the different tasks concerning music has also shown. Music is one of the important elements among other subjects such as art and drama in nurseries. It is very easy and artless for a child to immerse in the world of music and to show her expertise in different tasks, such as in those described in this paper.

Secondly, music is comprehensive and it interacts with different levels of sensation. Music also penetrates easily through human being to the stage of emotions, like several specialists [14], [15], [34], [37], [38] have confirmed in their studies. Through that characteristic it is easier to understand children's behaviour when listening to music. Listening is an experience [13], [17], which is different from visual. Music is heard within the entire body. Children's concentration, movements, dancing or spontaneous reacting to the sounds or music illustrates the multidimensionality of music. Through music it is possible to reach something, which otherwise would not be possible.

Thirdly, music enriches children's creations and there is a possibility to gain surprising results. Like several researchers [14], [15], [34], [37], [38] have discovered, music supports children's expressions and for example aesthetic enjoyment. Music supports children's creative process and gives them inspiration and new, imaginative ideas. From the point of view of design, especially the listening and drawing exercise with music acting as a trigger brought up new, interesting ideas.

Fourthly, music inspires and enables children to participate in a design process in a natural way. According to Flohr [17] extrinsic musical benefits for young children include using music to help teach other subject such as reading or mathematics. It is easy to subscribe to this and to pronounce the idea of music as an important supporter of design process with children, too. Music provides a natural way for a child to approach new issues such as design. Therefore, music can be used for making it easier for children to engage in technology design.

Altogether, this study offers some confirmation to the previously presented theories and demonstrates the usefulness of music in the design process with children. The major arguments supporting our findings concerning the usefulness of music in the design process could be summarized as follows:

- music is a natural part of children's life and education
- music is comprehensive and has an influence on all sensations
- music enriches and supports children's creations
- music inspires and supports children's participation

6 Conclusions

This paper reviewed the role of music in the design process with children. When taking the developmental stage of the children into consideration, we claim that music and other art based methods should play a significant role in designing with children. A review of the related research was carried out. Based on the review it was argued that the role of music in the design process has been neglected in the research on music, technology and children, while there is a lot of meaningful research in the areas of social and educational studies carried out. These studies indicate that music can and should appear in the design process in many different forms. The review of the literature on music, technology and children showed that the research has concentrated on technological products and technological needs of special education, mentioning music also as an important part of media. We, instead, examined the versatile role of music during the design process with preschool children in three different workshops. We identified three different roles for music: a contextual role, music as a trigger and music as content. The contextual role offers significant background information for the design process and insights into the world of the participating children. It also gives a preliminary conception about children's music and the elements of the music application. The role as a trigger shows the importance of musical experience, in this case of music listening, in the creative process like designing. It helps to arouse children's sensations and gives information about intuitive responses. Because of the support for creativity it is possible to get extraordinary results for the design process. The meaning of music as content increases especially when designing a music application. Children have clear opinions about musical content and they can create both real and imaginative sounds.

Regarding the practical implications of this study, we claim that other researchers interested in designing with children can now start planning to use music as part of their design sessions. Similar kinds of design sessions as the ones described in this paper can be organized, but the three roles of music can be integrated into other kinds of design sessions as well. Musical context can be inquired through interviewing, observing, drawing and prototyping, as has been demonstrated in this paper, in line with the results presented e.g. by Druin [11]. Treating music in the contextual role enabled the research team to gain information about children's understanding about different kinds of instruments, musical applications and music toys - about their structure and functionality - and about playing techniques and musical concepts they use. This all provides highly useful information for the adult designers making decisions related to the future music application: related to its structure, functions, appearance and content. The results of this study also indicate that preschool children are fully capable to sketch, build and present imaginative musical prototypes.

In addition, this study shows the usefulness of using music as a trigger. The benefits of this are connected to the imaginative results gained and to the ways music inspired children and supported their concentration and intensive working. Music samples played in the design session also gave structure for the task. Music can be integrated into numerous kinds of creative workshops organized with children. However, we recommend multidisciplinary cooperation in planning these sessions. Music education expertise was valuable in our research project especially during this phase.

Furthermore, we recommend also other researchers to invite children to create, together with adult designers, the content of the application, i.e. to act as design partners [12]. In our session children created imaginative musical worlds and brought up also surprising perspectives (e.g. the secret sounds). They were capable to create and describe sounds to the predefined musical landscapes and well equipped to express their opinions. However, related to this session it was also noticed that the preschool children's ability to concentrate is limited, which should be taken into account by all researchers working with young children.

There are several limitations connected to this study. More extensive presentation of the whole development process as well as of the research material gathered during the three workshops would bring up more varied picture of this field and give to the reader more possibilities to apply these ideas in practical design sessions with children. On the basis of this research we highlight the importance of charting the different roles of music and its several utilization possibilities. It is useful to specify the character of different roles and the methods used with children in design in practise. In the future it is worthwhile to continue this kind of work and to describe the different possibilities of music, not only as a listening experience, but also as a singing, composing and improvising experience.

Finally we conclude that in order to create meaningful, inspiring design experiences for children, designers in the HCI area should take into consideration the meanings of music. This area has many meaningful and creative possibilities to utilize music. This article shows a way, through music, towards more creative, inspiring design process where children are active participants designing the target application.

Acknowledgements. This research has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 22456. We thank the UMSIC consortium for enabling this work. We also thank SILKKIUIKKU -project group for carrying out the design activities in the nursery, as well as the children and teachers from the nursery for contributing to the design sessions.

References

1. Akiyama, Y., Oore, S.: PlaceAndPlay: a digital tool for children to create and record music. In: CHI 2008: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing, pp. 735–738. ACM Press, New York (2008)
2. Aucouturier, J., Ikegami, T.: The Illusion of Agency: Two Engineering Approaches to Compromise Autonomy and Reactivity in an Artificial System. Adaptive Behavior 17(5), 402–420 (2009)
3. Bar-Ilan, J., Belous, Y.: Children as architects of web directories: An exploratory study. Source: Journal Of The American Society for Information Science and Technology 58(6), 895–907 (2007)
4. Beck, L.: Nurturing your child with music: How sound awareness creates happy, smart, and confident children. Library Journal 125(4), 118–118 (2000)
5. Benveniste, S., Jouvelot, P., Lecourt, E., Michel, R.: Designing wiiimprovisation for mediation in group music therapy with suffering from behavioral disorders. In: Proceedings of the 8th International Conference on Interaction Design and Children, IDC 2009, pp. 18–26. ACM Press, Como (2009)
6. Berry, R., Makino, M., Hikawa, N., Suzuki, M., Inoue, N.: Tunes on the table. Multimedia Systems 11(3), 280–289 (2006)
7. Bobick, A.F., Intille, S.S., Davis, J.W., Baird, F., Pinhanez, C.S., Campbell, L.W., Ivanov, Y.A., Schutte, A., Wilson, A.: The KidsRoom: A perceptually-based interactive and immersive story environment. Presence-Teleoperators and Virtual Environments 8(4), 369–393 (1999)
8. Bresler, L.: Teachers as Audiences: Exploring Educational and Musical Values in Youth Performances. Journal of New Music Research 39(2), 135–145 (2010)
9. Druin, A.: Place called childhood. Interactions 2(5), 12–15 (1996)
10. Druin, A., Stewart, J., Proft, D., Bedeson, B., Hollan, J.: KidPad: a design collaboration between children, technologists, and educators. In: Proceedings of the SIGGHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, United States, March 22-27, pp. 463–470 (1997)
11. Druin, A.: Cooperative inquiry: Developing new technologies for children with children. In: Proceedings of CHI 1999, pp. 1–25. ACM Press, New York (1999)
12. Druin, A.: The Role of Children in the Design of New Technology. Behaviour and Information Technology 21(1), 1–25 (2002)

13. Dura, M.: The Phenomenology of the Music-Listening Experience. Arts Education Policy Review 107(3), 25–32 (2006)
14. Elliott, D.J.: Musicing, Listening, and Musical Understanding. Contributions to Music Education (20), 64-83 (1993)
15. Elliott, D.J.: Putting Matters in Perspective: Reflections on a New Philosophy. Quarterly Journal of Music Teaching and Learning 7, 20–35 (1997)
16. Fischer, T., Lau, W.: Marble track music sequencers for children. In: IDC 2006 Proceedings of the 2006 Conference on Interaction Design and Children, pp. 141-144 (2006)
17. Flohr, J.W.: The musical lives of young children. Pearson Prentice Hall Music Education, Upper. Saddle River (2005)
18. Guha, M., Druin, A., Chipman, G., Fails, J., Simms, S., Farber, A.: Working with Young Children as Technology Design Partners. Communications of the ACM 48(1), 39–42 (2004)
19. Godeli, M.R., Santana, P.R., Souza, V.H., Marquetti, G.P.: Influence of background music on confused and agitated elderly patients. Archives of Psychiatric Nursing 7, 284–291 (1996)
20. Hanna, L., Risden, K., Alexander, K.: Guidelines for Usability testing with Children. interaction, pp. 9–14 (September + October 1997)
21. Harris, B., Summa-Chadwick, M.: A computerized system for Neurologic Music Therapy. Journal of Computing Sciences in Colleges 21(2), 250–257 (2005)
22. Hartnett, J., Lin, P., Ortiz, L., Tabas, L.: A responsive and persuasive audio device to stimulate exercise and fitness in children. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 1837–1842 (2006)
23. ISO 13407: Human-centered design processes for interactive systems, International Standard (1999)
24. Jennings, K.: Hyperscore: A Case Study in Computer Mediated music Composition. Education and Information Technologies 10(3) (2005)
25. Jentschke, S., Koelsch, S., Sallat, S., Friederici, A.: Children with specific language impairment also show impairment of music-syntactic processing. Journal of Cognitive Neuroscience 20(11) (2008)
26. Koelsch, S., Grossmann, T., Günter, T., Hahne, A., Scröger, E., Friederici, A.: Children Processing Music: Electric Brain Responses Reveal Musical Competence and Gender Differencies. Journmal of Cognitive Neuroscience 15(5) (2003)
27. Li, W.: Musical box: draw it yourself. In: IDC 2010 Proceedings of the 9th International Conference on Interaction Design and Children (2010)
28. Lugmayr, A., Risse, T., Stockleben, B., Laurila, K., Kaario, J.: Semantic ambient media-an introduction Source: Multimedia Tools and Applications, vol. 44(3), pp. 337–359 (2009)
29. Magne, C., Schön, D., Besson, M.: Musician Children Detect Pitch Violations in Both Music and Language Better than Nonmusician Children: behavioral and Electrophysiological Approaches. Journal of Cognitive Neuroscience 18(2) (2006)
30. Markopolous, P., Bekker, B.: On the assessment of usability testing methods for children. Interacting with computers 15, 227–243 (2003)
31. Markopoulos, P., Read, J., MacFarlane, S., Höysniemi, J.: Evaluating Interactive Products for and with Children. Morgan Kaufmann, San Francisco (2008)
32. Mazzone, E., Iivari, N., Tikkanen, R., Read, J., Beale, R.: Considering context, content, management, and engagement in design activities with children. In: IDC 2010 Proceedings of the 9th International Conference on Interaction Design and Children, pp. 108-117 (2010)

33. McKnight, L., Read, J.: Designing the 'record' button: Using children's understanding of icons to inform the design of a musical interface. In: IDC 2009, pp. 258–261. ACM Press, Como (2009)
34. Merriam, A.P.: The Anthropology of Music. Northwestern University Press, Evanston (1964)
35. Miura, S., Sugimoto, M.: Supporting children's rhythm learning using vibration devices. In: Proc. of CHI, pp. 1127–1132. ACM, New York (2006)
36. Piaget, J.: Play, Dreams, and Imitation in Childhood. Norton, New York (1962)
37. Reimer, B.: A philosophy of music education. Nortwestern University, USA (1970/1989)
38. Reimer, B. (ed.): World Musics and Music Education: Facing the Issues. Reston, VA.MENC – The National Association for Music Education (2002)
39. Reynolds, N.: The Computer as Scaffold, Tool and Data Collector: Children Composing with Computers. Education and Information Technologies 10(3) (2005)
40. Rosenbaum, E., Silver, J.: Singing Fingers: fingerpainting with sound. In: IDC 2010 Proceedings of the 9th International Conference on Interaction Design and Children (2010)
41. Rusinek, G., Rincon, C.: Attending Musical Performances: Teachers' and Students' Expectations and Experiences at a Youth Programme in Madrid. Journal of New Music Research 39(2), 147–158 (2010)
42. Scaife, M., Rogers, Y.: Kids as informants: Telling us what we didn't know or confirming what we knew already? In: Druin, A. (ed.) The Design of Children's Technology, pp. 27–50. Morgan Kaufmann, San Francisco (1999)
43. Upitis, R.: Technology And Music - An Intertwining Dance. Computers & Education 18(1-3), 243–250 (1992)
44. Vygotsky: On the Zone of Proximal Development. In: Mind in Society (Trans. M. Cole), pp. 84–91. Harvard University Press, Cambridge (1978)
45. Weinberg, G., Orth, M., Russo, P.: The embroidered musical ball: a squeezable instrument for expressive performance. In: CHI 2000 Extended Abstracts on Human Factors in Computing Systems (2000)
46. Zigelbaum, J., Milner, A., Desai, B., Ishii, H.: BodyBeats: whole-body, musical interfaces for children (2006)

ToCoPlay: Graphical Multi-touch Interaction for Composing and Playing Music

Sean Lynch, Miguel A. Nacenta, and Sheelagh Carpendale

University of Calgary,
Calgary, Alberta, Canada
{sglynch,miguel.nacenta,sheelagh}@ucalgary.ca

Abstract. With the advent of electronic music and computers, the human-sound interface is liberated from the specific physical constraints of traditional instruments, which means that we can design musical interfaces that provide arbitrary mappings between human actions and sound generation. This freedom has resulted in a wealth of new tools for electronic music generation that expand the limits of expression, as exemplified by projects such as Reactable and Bricktable. In this paper we present ToCoPlay, an interface that further explores the design space of collaborative, multi-touch music creation systems. ToCoPlay is unique in several respects: it allows creators to dynamically transition between the roles of composer and performer, it takes advantage of a flexible spatial mapping between a musical piece and the graphical interface elements that represent it, and it applies current and traditional interface interaction techniques for the creation of music.

Keywords: Multi-touch, collaboration, composition, music, musical instrument.

1 Introduction

Musical instruments can be considered interfaces between a musician and the acoustic space. It is, therefore, not surprising that music has been one of the main areas of application for the advances in interface science and engineering of the last few decades. The advent of new interface paradigms such as multi-touch and tangible computing have spawned a series of remarkable prototypes and commercial systems that enable new ways of creating and experiencing music, often beyond what is possible with traditional instruments. For example, new interfaces for musical expression currently allow people without music backgrounds to explore new sounds [4, 19, 21] and groups of people to perform together closely through interactive tables [14, 15, 19, 21].Computer and interface technology has changed how music can be performed, how sound can be generated, and also how it can be composed; however, with few exceptions [12, 20,19], composition tools do not take advantage of new interaction possibilities, and still are restricted to mouse and keyboard, sometimes with interfaces to electronic versions of traditional instruments [6, 9].

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 306–322, 2011.

In this paper we present ToCoPlay (Touch-Compose-Play), a musical interface designed to support playing and composing music. By enabling seamless transition between performing and composing we explore facilitating musical actions that are already part of musicians' activities, since composing music requires listening and changing what is being created, and, conversely, there are many forms of music where composition is part of the performance process (e.g., jazz). Our system is based on a horizontal multi-touch surface, and is intended for people with or without musical backgrounds. During the design of this new kind of musical interface, we made design decisions to enable control, configurability, a straightforward mapping between the visual and acoustic space, and to support a minimal set of interface elements and operations.

The structure of the paper is as follows: in the next section we provide relevant background on interfaces for musical expression. Then we discuss our goals and strategies for the design of the ToCoPlay interface, followed byits detailed description. We present four examples of ToCoPlay in use, followedby a discussion of the results of our exploration, andfinish with our conclusions.

2 Related Work

There exist many applications for musical expression on digital surfaces, and these vary widely in interaction methods and purpose. Most applications support a combination of the following musical activities: playing sounds, synthesizing sounds, sequencing sounds in real-time, and composing. In the following paragraphs we provide a brief summary of prior work, and how it differs from our own.

There are a large number of previous systems that focus on creating new sounds through the modification of streams or samples. Some of these are based mainly on tangibles, such as the seminal work by Jordà et al. on the Reactable [14,15], Audiopad [21], Pin & Play & Perform [27], and Condio [5]. These systems rely mostly on physical blocks that represent certain sound synthesizing operations (e.g., filters, tone generators, echo, samples) that are placed on the table and manipulated in real time to produce new sounds and rhythms. Other related work has the same approach, but is based on touch. For example, Akustich [1] is a direct-touch interface in which participants manipulate sound by performing multi-touch gestures on a table, where each gesture changes the sound and the visuals in a different way. In a similar way, performers using Sound Storm [23] stand in front of a vertical display and use touch to modify sound waves that come across the screen. The Synthesis and Control on Large Multi-touch Displays [7]demo by Perceptive Pixel provides a multi-touch knob and slider interface for the Synthesis Toolkit (a sound library for C++).

These systems resemble ToCoPlay in that they allow the real-time modification of sound by human performers, which makes them *virtual instruments*. However, they differ from ToCoPlay in several ways. First, ToCoPlay is based on standard note scales and sounds; therefore it does not focus on the generation of new sounds and rhythms. In this respect, ToCoPlay is simpler. Second, these systems are not designed to support composition: the temporal evolution of the music is difficult to reproduce because sound is mostly dependent on the state of the interface and the specific user actions. ToCoPlay is designed to store, modify, and reproduce created sequences, which supports performance, composition, and all combinations of the two.

A second group of related systems do not support sound synthesis; instead, they focus on supporting music performance that resembles how we play traditional instruments. For example, SurfaceMusic [8] enables playing three instruments on an interactive tabletop (a drum, a string instrument, and a wind instrument) by using a specific multi-touch gesture on each (tapping, strumming, and air-pushing). The Roots [12] application generates ambient sounds when a user touches a particular region on the multi-touch horizontal table (the Bricktable [11]) and produces visually pleasing, spiraling trees. The Jam-o-drum [4] is a series of external controllers that lay on a horizontal table display, where multiple people can play sounds by tapping their own pad; the Jam-o-drum is primarily for playing games. These systems are closer to ToCoPlay in the simplicity of the basic sound elements that are combined, but they do not generally support composition any more than traditional instruments.

A third group of interfaces enable composition. Traditional applications such as Garage Band [9], and Cubase [6] are single user, rely on traditional notation, and are meant to be used from a desktop PC. The MusicTable [24]allows people to place cards that represent pitch and instrument on a table. The cards are detected by the system and played in sequence from left to right. BlockJam [19]lets people combine attachable active physical blocks that have a button to generate a sequence of sounds. The rules of how the sound is produced depend on the type of block and its current state. The resulting sound cycles through the blocks and visual feedback is provided through a monitor. InstantCity [13]is an art music automata that plays different ambient sounds depending on how blocks are distributed on the table. Xenakis [2]is a non-deterministic system that uses tangible blocks to generate music according to probability distributions based on the distance between blocks and their similarity.In Noteput [20]participants place musical note-shaped physical blocks onto a digital horizontal surface which are played left to right according to their position, creating a physical staff.Spaces [12]is an application for composing ambient sounds on a multi-touch table, where participants touch a region of a table to make that region a warmer or cooler colour, with matching sound. Stereotronic Multi-Synth Orchestra [25] is a Microsoft Surface [17] application in which participants place notes into pre-defined concentric rings, and notes are triggered when the rotating element of each concentric ring goes over that particular note.

These systems resemble ToCoPlay in their ability to easily set up temporal sequences of sounds, but their design does not facilitate the performative component. ToCoPlay, similarly to all of these systems, allows setting up music and then triggering its performance without much human intervention; however, ToCoPlay focuses on supporting seamless transition between configuring the music, changing it, and performing or improvising with the available components.

In summary, ToCoPlay situates itself as being both a composition tool and an instrument, but not a sound synthesizer. ToCoPlay's goal is different from the goals of systems mentioned above, since we set out to achieve seamless transitions between the playing and composing activities. It also has several features that distinguish it from previous systems, most notably a flexible and precise spatial mapping (contrary to the mostly linear ones described above) and its configurability to create configurations that adapt to the person and the situation.

3 Design Goal and Strategies

As illustrated by the related work section, there is considerable activity in the development of software with musical capabilities. However, most of these focus on strengthening either the performance potential or concentrate on supporting musical composition. Our goal is to, as part of a musical software interface, investigate the integration of these activities: composition and performance. Thus, we intend to both keep interactions as simple possible and to explore ways of using simple interactions to create powerful methods for introducing complexity. In ToCoPlay, we offer an interface that can be played as an instrument, but that also invites composition. To achieve these goals we apply four main design strategies: enabling control, enabling configurability, creating a formative visual-acoustic mapping, and providing a minimal set of interface objects and operations.

3.1 Enabling Control

For ToCoPlay we opted for a discrete paradigm, which is more closely related to how a piano is played (by pressing keys) than to the continuous model used by most modern synthesizers, which are commonly controlled by turning knobs and adjusting continuous parameters. In ToCoPlay individual sounds are made by pressingkeys. Complex sounds are developed through grouping, sequencing, and nesting. Grouping, sequencing and nesting are all interactively created and visually explicit.

3.2 Enabling Configurability

The physical characteristics of traditional physical instruments, such as the arrangement and shape of keys, have evolved into their current forms through many small mutations that often took centuries. Current electronic interfaces have not yet enjoyed the popularity or the time to evolve in this way. For ToCoPlay we introduce a model in which the location of keys and their grouping is fully configurable in real time. It is reconfigurable during the design of the sounds or melodies and it is reconfigurable during play. These freedoms can enhance the creation of a personalized interface that fits one's own hands and, thus, is easier to play. Through reconfigurability it can also be adapted for different people and compositions; for example, one can group notes together so that they can be played without moving one's hand (e.g., by placing them directly under the fingers when they are in a natural posture), or one can move a certain set of notes closer when they are to be used.

3.3 Visual-Acoustic Mapping

For most people composing music requires at least two basic elements: a way to try out the sounds, melodies, and harmonies as they are being created, and a way to record how these are produced (including their temporal relationships), for later performance. Many of the current music generation systems that we describe in Section 2 are inadequate for composition, not only because they lack features to record how interaction must take place to reproduce a certain sound, but also because the continuous nature of the interaction makes it difficult to create a notation that accurately and effectively relates the almost infinite variety of gestures that can take place in a multi-parametric continuous space to the way they sound.

Therefore we designed a visual-acoustic mapping that provides a straightforward relationship between what is displayed on screen and what is going to be reproduced, so that actions on the interface can be interpreted as changes in the music. Note that this is not a visualization of the sound but a spatialisation of the relationships between sounds. One such possible mapping is traditional music notation; however, this notation did not evolve to facilitate real-time interaction, is not generally accessible to novices and, most importantly, it imposes a rigid spatial mapping of time (along the staff, moving horizontally through time and then down through the scale) that directly contradicts our configurability goal. In other words, using a staff metaphor for our interactive system would not allow us to make an interface that is easily playable; it would encumber collaboration (e.g., due to its implicit orientation), and would not allow us to group elements according to other groupings that are not strictly temporal. Most of these arguments apply as well to other existing sequencing and composing applications (e.g., [2]). Again, we keep this spatial/visual relationship simple. Sequencing is specified through explicit directed links, where the spatial distance is mapped to temporal spacing.

3.4 Minimal Objects and Operations

Making an interface accessible to novices requires simplicity in the number and complexity of object types in the interface and their operations. We set out to create an interface that was based on a minimalistic approach. Unlike many existing musical interfaces that have large lists of objects, filters, types of connections and parameters, our interface needs only a few objects and a few rules about the ways the objects can relate to each other.

It might seem counter intuitive to restrict the design of the system to a small set of primitives in order to enhance expressivity and encourage playfulness. However, there is a wealth of experience in other fields that shows that minimalistic systems can lead to complex and expressive results; for example, a von Neumann machine, the Game of Life [10], SWARM intelligence [3], or the game of Go.

4 The ToCoPlay Interface

The interface of ToCoPlay is composed of five atomic elements: keys, key fountains, containers, links, and dummy keys. These elements can be combined to enable the properties outlined by the design section above. The following subsections provide a detailed description of each of the atomic elements, their interactions, and how their design, interactions and combinations support our design objectives.

4.1 Keys

The base atomic element in ToCoPlay is the key, a group of which are shown in Figure 1. A key is a circular interface element that plays a specific note of a specific instrument when touched. The note assigned to a key is one of the twelve pitches of the Chromatic Scale, which are represented visually by a change in value of the key's color; that is, lower notes are represented by darker colors than higher notes. The hue of the key indicates the instrument (or timbre) of the note. At this moment, two synthetic instruments, a sine and a square wave generator, are represented by pink and yellow key hues respectively.

Fig. 1. Keys in ToCoPlay, representing the C Chromatic Scale

Keys are circular to facilitate being easily touched while having a small footprint, which allows many keys to be visible simultaneously on the interface. Note that, unlike with most traditional interfaces, ToCoPlay can have many keys for the same note which, as we will see, supports the generative nature of the system. The one-to-one mapping between the colour of the keys and the sound that they emit is designed to enable a direct relationship between the visual and acoustic spaces.

Tapping a key instantly plays the corresponding note and makes it transparent, establishing a clear connection between the visual and the acoustic feedback of the system. A key can be moved and repositioned anywhere across the interface. Repositioning keys allows the musician to arrange them in ways that support playing the interface as a customizable instrument; for example, the keys necessary to play different leit-motifs of a piece can be placed so that they correspond to the tips of the fingers when the hand is in different areas of the table. The mobility of keys on the interface also enables clustering of notes in different visual layouts which, as described under 'links', further supports the compositional aspects of the interface. Keys can also be flicked around in the interface to quickly move them away from an area of interest, or to remove them from the interface if flicked out of the visible area of the interactive table.

4.2 Key Fountains

New keys are created by dragging them out of *key fountains*. Key fountains are represented by a large circle surrounded by the smaller, circular keys (see Figure 3). The keys are arranged around the circle clockwise from C to B in half-tone intervals (12 notes). When a key is removed from a fountain, another copy appears in its place, which enables the creation of an unlimited number of keys of each type, as shown in Figure 2.

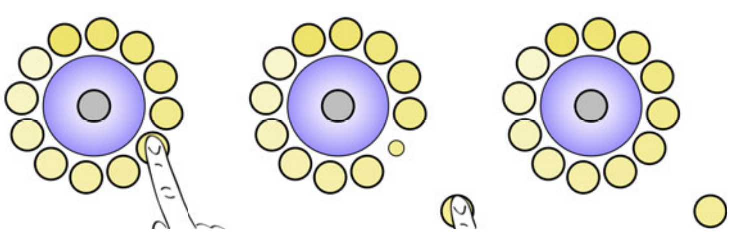


Fig. 2. A key is removed from a key fountain, and is replaced by another key

The interface contains a key fountain for each instrument (timbre). Although fountains initially appear at the center of the application, they can be moved around by dragging them from the central circle to make space for other kinds of interaction. It is also possible to rotate and scale fountains through the standard two-point RST gesture [18] to accommodate the needs of people located in different areas of the table. Consistent with the key interface elements, a fountain can also be flicked away out of the screen, which clears it from memory. Note fountains can be duplicated by double tapping on their central circle, again to accommodate the needs of multiple people around the table.

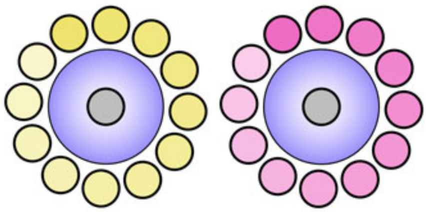


Fig. 3. The two different key fountains, each representing a different timbre. Each key fountain consists of a large circle surrounded by keys, representing the notes in the chromatic scale, from C(darkest) to B(lightest).

4.3 Containers

Containers, which are used to hold elements, are regions defined by free-form touch-traces on the table (see Figure 4). Containers are created by starting a touch-trace on an unoccupied place on the table and closing off the trace. A line indicates the outline of a container while it is being drawn, allowing for the precise definition of regions. When the finger is released and the container is completed, it appears as a translucent blue shape, and it automatically contains the objects that the trace enclosed.

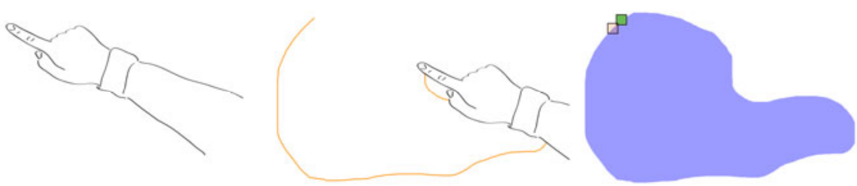


Fig. 4. To trace a container start in an unoccupied space and create a closed form

Containers can group heterogeneous collections of any interface objects except key fountains(see Figure 5). Containers can also contain other containers, which supports the generative nature of composition by allowing nesting. They enable the individuals to work with their own pieces of a composition, or of an instrument, and then to combine them at a later point in time.

Similarly to key fountains, containers can be moved, scaled or deleted (i.e., they can be dragged, flicked or manipulated with two-finger RST). Objects such as keys and other containers that the container currently holds stay with the container as the container is moved or rotated. This allows musicians to manage the screen area efficiently by manipulating meaningful groupings of objects and supports socially-based use of the space [22]. An object (e.g., a key) can be added to a container by moving it from outside the container to inside the container, or by moving the container so that it fully contains the object.

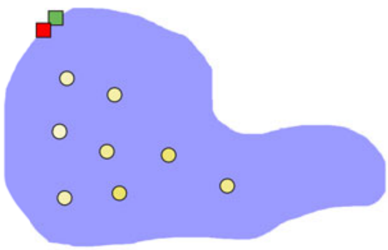


Fig. 5. A container with keys placed inside

Containers also have two buttons. The container and all of the elements and structure it contains can be copied or duplicated by tapping on its green button. The copy of the container appears at a small offset from the original (see Figure 6). The copy also oscillates for a brief period of time to highlight that it is a new object. Next to the copy button, there is a red button that is used to control the lock state of the container. A container can be in one of two states: locked or unlocked. An unlocked container can be moved, resized, or rotated and also allows its contents to be moved or removed. When locked, the container prevents accidental movement or transformation of its contents and itself, keeping all elements in place. When unlocked, containers and their contents can be manipulated to find different internal and external configurations. If a container is unlocked, an element can be moved out of a container; once it is fully outside of a container, and is released, it no longer belongs to that parent container. The opacity of the red lock button reflects the current state of a given container: opaque red means locked, translucent means unlocked. Containers can be deleted in one of two ways: either by throwing them off screen like keys or key fountains, or by a gesture that crosses the container boundary twice.

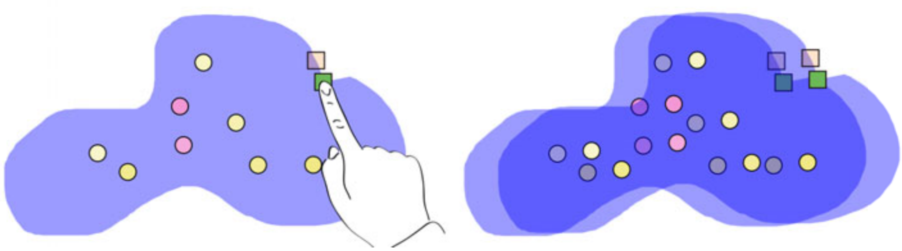


Fig. 6. Pressing the container's green button creates a duplicate at a slight offset

Together, duplication and repositioning provide the freedom to design personalized performance arrangements. Container and key groupings can become ‘instruments’ that can be shaped according to the specific anatomical or musical needs of the performer (e.g., according to the hand shape) and placed by the musician in different areas as required by the performance (e.g., storing instruments in distant regions when not required and bringing them close when being used). Simultaneously, the grouping of keys supported by containers, combined with nesting, enables the natural organization of musical compositions in different hierarchies such as themes, choruses, bridges, sections, etc. Container manipulation also supports reducing and organizing containers in space according to their current relevance (shown in Figure 7).

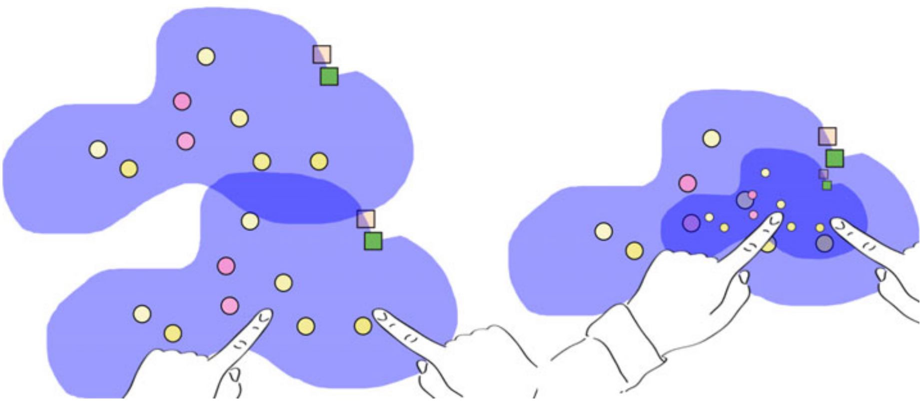


Fig. 7. A container is rotated, scaled and translated to fit inside of another container

4.4 Links

The elements described thus far support playing notes, grouping keys so that they form ‘instruments’ of arbitrary shapes, sizes, and orientations, and distributing these groupings in meaningful layouts over the table. However, the nature of the interaction with these elements does not enable composition or playback, since each note still has to be directly played by the musician. In order to support composition it is necessary to provide a mechanism that allows sequencing of notes at different times and intervals. We provide this mechanism through our link interface element.

Links are one-directional connections from one key to another that denote a sequence in how the notes of the corresponding keys are played (see Figure 8). Links are created by tracing a line that connects two keys when the original key is inside a locked container. Providing that the original key is within a locked container, any two keys can be connected, including keys belonging to different containers, keys that do not belong to any container or even keys that belong to contained sub-containers. Links can be deleted by tracing a line that crosses the link’s representation within a locked container or in blank space.

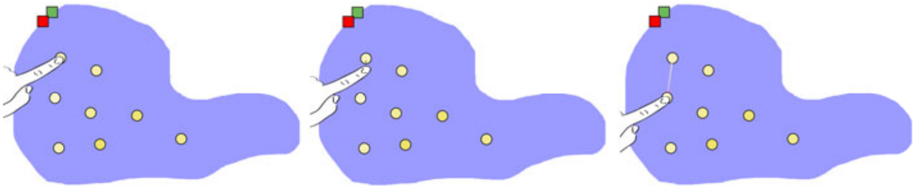


Fig. 8. A key, inside a locked container, is pressed and then connected to another key, forming a link

If a key is tapped and it has an outgoing link to another key, the other keys will automatically be played when the original key has finished playing its note. If the subsequent keys have links, the subsequent notes will be played and so on, until the chain ends, thus providing a way to sequence music events. The direction of a link is indicated by an arrow at the end of the link, indicating which key is theinitial key and which is the subsequent key. The length of the link determines the timing between the tapping or signal reception of the initial key and the start of the note of the subsequent key. The initial key's duration depends also on the duration of the following link, since its note will play until the next note in this particular chain starts to play. Timings created with links are not fully continuous and instead snap to multiples of 125ms.

A key can have any number of incoming or outgoing links; every time that a link sends a signal to a key, its entire set of outgoing links will fire signals to the corresponding keys. This branching enables creating chords and counterpoint melodies –common elements of music in most cultures– as well as cyclic patterns and recursive music as we will show in the examples of Section 7.

4.5 Dummy Keys

Dummy keys are special keys that do not play any sound by themselves, but serve as proxies to activate other keys. A dummy key's incoming link length is ignored. By placing a dummy key that connects distant keys A and B (in that sequence) close to B, the note of key A will be played for a time proportional to the distance between the dummy note and key B instead of by the distance between keys A and B (see Figure 9). This allows for flexibility in the connection and placing of groups of keys that need to be played close in time but need to remain in non-adjacent locations. Similarly, a combination of dummy keys can be used to create instruments with keys

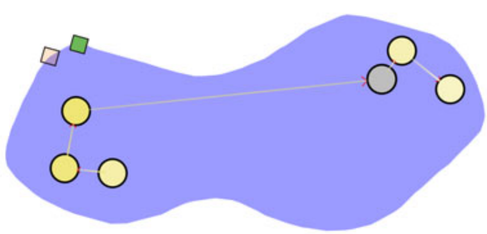


Fig. 9. A dummy key connects two keys at a large distance. As a result, the key with the outgoing link will have a shorter distance to the outgoing link of the dummy key.

located in anatomically convenient locations, as we will show in Section 6.1. Dummy keys can be created by dragging them from the center of a key fountain, and their inner colour is grey.

5 Implementation

We implemented ToCoPlay to work on the Microsoft Surface™, and used Microsoft's WPF™ libraries for Surface development. As a result, every visual object in ToCoPlay receives events when it is touched. For sound, we used Microsoft's MediaPlayer framework, and triggered wav samples. The wav samples were generated to have two distinct timbres: that of a chromatic scale with sinusoidal tones, and of a chromatic scale with square tones. We used a simple tone generator application to create these, and loaded multiple instances of each file so as to have the ability to play a sample more than once at any point in time. For the visual elements of ToCoPlay, prototypes were built using Microsoft Expression Blend™, and the final versions were built using a combination of Microsoft's C# and XAML.

6 Use Scenarios

6.1 Chord and Scale Instrument

This use case scenario illustrates the configurability of the system through the use of shaped containers, key layouts, and dummy keys. It corresponds to Example 1 in the video figure. The video shows how to build two instruments. The first one is a simple instrument with the five keys of a blues pentatonic scale, which are arranged in space to correspond to the tips of the fingers of the performer. Keys are made larger to make them easier to play, and adjusted for angle of comfort for the performer.

The second is an ad-hoc chord instrument that can play three basic chords of a simplified blues progression (C7, F7, and G7). Each chord is formed by four keys, which are activated by a dummy key simultaneously. Each of the chord's dummy keys are activated (with no delay) by a dummy key located within their respective containers, to facilitate playing with the left hand (see Figure 10 and video figure). The performer shown on the video (one of the authors) does not have any piano experience.

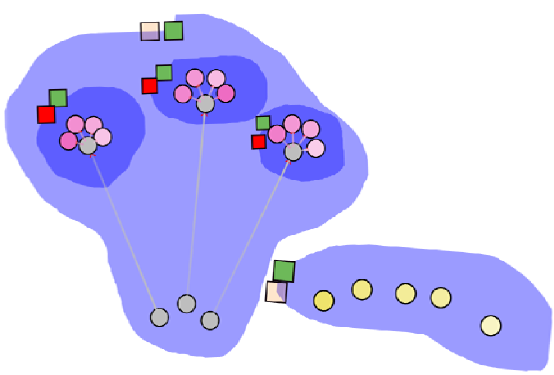


Fig. 10. A chord instrument (left) and a scale instrument (right)

6.2 Harmonious Loops

This scenario illustrates how simple musical loops can produce a more complex musical texture through straightforward copying of a container, and the real-time rearrangement of the keys. In this example we part from a single loop, then copy it, modify the copy's loop timing slightly, and play both loops simultaneously. Because both loops are similar but not identical in length, the harmonies being played change continuously, in a sequence that repeats itself on a period much longer than the duration of either loop, creating a complex musical texture out of two simple elements.

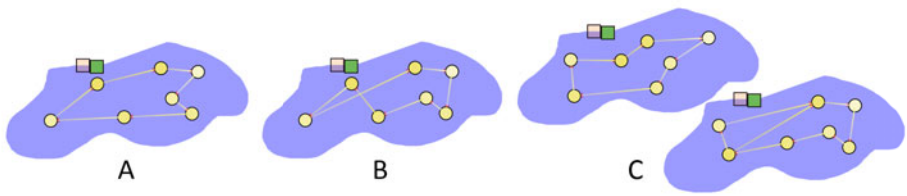


Fig. 11. A) A simple loop; B) A slightly modified copy of the loop in A);A) and B) played together will form a complex musical texture; C) Both loops can be modified in real time to further evolve the texture created

6.3 Canon

In music, a Canon is a musical form that is based on the repetition of the same melody by several instruments or voices, starting at different points in time. Because each instrument or singer has a delay with respect to the others, different notes sound simultaneously, resulting in chords. This example is a more sophisticated evolution of the harmonic loops, with a more sophisticated output. The triangular structure on top (see Figure 12) triggers the start of a full cycle of the melody (bottom), at times that are delayed by exactly one third of the length of the full melody. The cycle ends when one of the links of the triangular loop is cut (see video figure, Example 3).

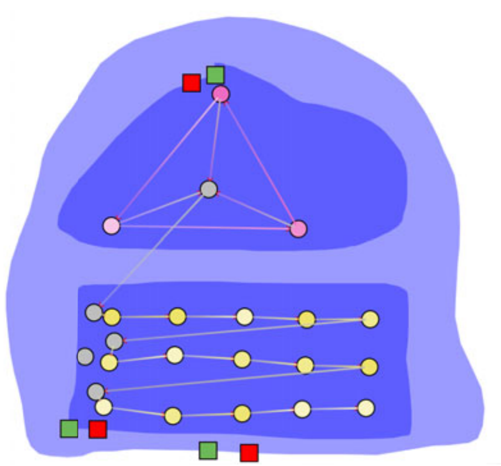


Fig. 12. A Canon composition

6.4 Making Music with Shapes

This scenario takes advantage of the flexibility of our visual-acoustic mapping and the configurability of the interface to show how visually creative configurations can be created. In this case, two participants are collaborating in creating music to illustrate Antoine de Saint-Exupéry's famous story from the Little Prince. Figure 13 and the video figure shows how two participants can create shapes, modify them, populate them with music, and then combine them to generate a small sound composition. This example illustrates how ToCoPlay can also be playful, and suitable not only for serious composition, but also for musical games. The creation of shapes to embed musical themes might also prove useful to improve recall and identification of musical sections when the piece is complex, even if not used with visual-aesthetical goals.

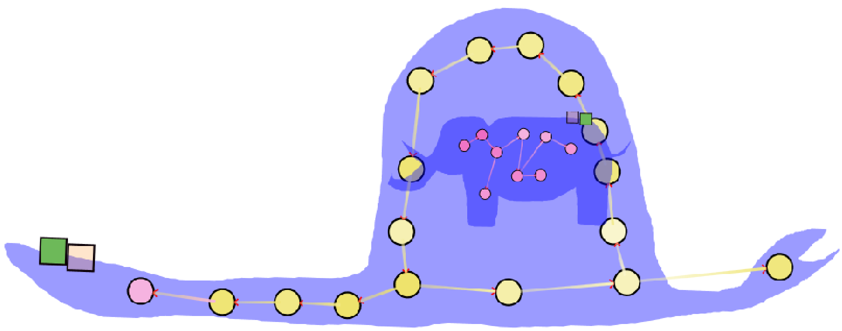


Fig. 13. Two shapes are created with connected links inside: an elephant inside a snake; or is it a hat?

7 Discussion

7.1 Continuous versus Discrete Control

Designing music interfaces based on multi-touch interactive surfaces provides strong advantages; for example, a horizontal surface can enable collaboration, the input and output space coincide and are fully configurable, and a virtually infinite number of parameters and levels can be controlled through touch manipulations, traces and gestures. These manipulations are often of a continuous nature; for example, in the Reactable the rotation of an element is often mapped to a continuous parameter such as pitch or volume. Our system intentionally diverts from this model and instead offers a small discretized set of alternatives: a key can only play one of twelve notes, and time intervals are multiples of 125ms.

We fully recognize the importance of continuous control, especially for the manipulation of a complex, continuous control space such as sound synthesis. However, our design is more discrete because our goals are different from those of synthesizer and computer instruments. We prefer instead to make the results of actions easier to control and predict, and therefore easier to learn for people without a background in music. In exchange for this simplification of interaction, we are reducing how parameters that can be manipulated through our system; for example, it is not possible yet to produce vibrato or tremolo (small variations in pitch or volume) with ToCoPlay.

7.2 Homomorphic Relation between Space and Sound

Many of the existing synthesizers and computer-based instruments described in the related work, as well as most VJing interfaces (e.g., [26]) provide visuals that depend mostly on the sound that is generated at that moment (or shortly before) and/or the input being provided. Since our main goal to support seamless transitions between performing and composing, we provide a tool that supports the broader view of the piece required by composers. We achieve this through our straightforward mapping that relates the visuals of the interface (the position, colour and lengths of the links) to sound. In particular, the mapping between the length of links and the duration and sequence of notes was of great help and impact when creating the examples described above, and it was often understood in demonstrations before it was explained. This relationship between the visual and the sound spaces is also useful in enabling shifts from composition to performance, since looking at the interface could allow performers to intervene and perform modification of a composed element in real time.

Although it would have been possible to create one-to-one relationships between the spatial and the acoustic spaces, we preferred to allow some freedom. Unlike traditional musical notations, which are generally isomorphic, our spatialisation of time allows several patterns to generate the same sequence of notes; for example, by changing the angle between keys without changing their distance or using dummy keys. This makes space for layouts that are not exclusively following musical requirements, but also accommodate interface considerations (e.g., collaboration, shape of the hand as in Examples 1 and 4), and aesthetic or storytelling possibilities, as shown by Example 4. We believe the visual aesthetic elements can be complementary and important for performing and composing, especially for people without musical background and for children; in fact, notations that cross media boundaries in art are not uncommon (e.g., Apollinaire's concrete poetry).

7.3 Few Building Blocks Can Go a Long Way

Our current version of ToCoPlay is certainly limited in several aspects, and does not provide many of the expressive channels that other music applications offer. Nevertheless, we found our decision to keep the interface to a few elements and rules useful and rewarding. We believe that the small set of elements will make the interface easier to learn, and we showed through our examples how initially simple combinations of a few elements can result in fairly complex sound textures (e.g., Example 2).

This kind of approach is particularly suited to a multi-touch interface, and generally difficult to apply to tangible interfaces such as the Reactable [15] and Audiopad [21], which would require many physical elements and can be quite expensive. In multi-touch systems have advantages in that we are free to replicate elements until the screen is full. However, touch detection and touch feedback do not achieve the same tactile and haptic quality of tangible interfaces.

7.4 Limitations

ToCoPlay is an exploration of musical interfaces and, as such, has helped us uncover new challenges and presented new difficulties. Probably most obvious is the need to improve the quality and quantity of samples for instruments to improve the quality of the sound. The sinusoidal- and square-wave generators used as initial instruments were adequate for our initial explorations, but we are aware that new, richer sounds can substantially help increase the musical value of the interface. Adding an external sound generation engine (possibly on a different machine) can help improve the quality of sound and, at the same time, solve some of the problems derived from dealing with an interface that has many elements and momentarily freezes for some operations such as large container creation.

Similarly, we have contemplated adding new controllable parameters to increase the expressivity of the system. For example, changing the volume (dynamics) of individual notes or containers can be valuable for composing. We also do not currently provide a method of creating pauses and delays in a composition. Future changes in this direction must, however, be weighed against the increase of interface complexity that contradicts our strategy of maintaining a simple set of basic elements and interactions.

Some of the aspects of our visual-acoustic mappings also present space for improvement. For example, we have noticed that the mapping from the HSV colour space of keys to their note's pitch is difficult to read and compare, especially if notes are distant from each other. We are already considering other kinds of mappings based on symbols and non-linear mappings of color to pitch, and mappings that would facilitate the inclusion of more than one octave.

Another issue with our spatial mapping is the ability to read complex scores of music.We are interested in finding out whether spatial relationships could become problematic (e.g., with a lot of containers embedded within one another) and how the interface can be modified for large scale compositions.

Finally, our current evaluation of the system has been constrained to just a few people from our department. Although reactions have been positive, we intend to validate our design by placing it into the real world and gather more impressions to produce a new, better version.

8 Conclusion

This paper presents ToCoPlay, a multi-touch tabletop musical interface that allows multiple people to create music. The main goal of ToCoPlay is to provide an interface that enables the performance of music while also encouraging composition. We applied four main strategies to design a system that allows seamless transitions between the performance and composition activities of participants: enabling control, enabling configurability, creating a simple visual-acoustic mapping, and relying on a minimal set of interface elements and operations. We also provide a set of four examples that demonstrate different aspects of the operation of the interface and its expressive capabilities. Finally, we discuss the implications of our design decisions and how they make ToCoPlay unique.

Acknowledgements. We would like to thank Uta Hinrichs for her valuable assistance creating the first version of the video. We are also grateful for comments and discussion from Lawrence Fyfe and many other members of the Interactions Lab and the Innovis group. This work was supported in part by Canada's Natural Science and Engineering Research Council (NSERC), Canadian Foundations for Innovation (CFI), Alberta's Informatics Circle of Research Excellence (iCORE) and by SMART Technologies Inc.

References

1. Akustich, `http://modin.yuri.at/tangibles/data/akustisch.mp4` (last accessed January 2011)
2. Bischof, M., Conradi, B., Lachenmaier, P., Linde, K., Meier, M., Pötzl, André, E.: Xenakis: combining tangible interaction with probability-based musical composition. In: Proceedings of TEI 2008, pp. 121–124. ACM, New York (2008)
3. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, Oxford (1999)
4. Blaine, T., Perkis, T.: The Jam-O-Drum interactive music system: a study in interaction design. In: Proceedings of DIS 2000, pp. 165–173. ACM, New York (2000)
5. Condio, `http://media.aau.dk/~gtal05/condioProject.html` (last accessed January 2011)
6. Cubase, `http://www.steinberg.net/en/products/cubase/cubase6_start.html` (last accessed January 2011)
7. Davidson, P.L., Han, J.L.: Synthesis and control on large scale multi-touch sensing displays. In: Proceedings of NIME 2006, IRCAM, pp. 216–219 (2006)
8. Fyfe, L., Lynch, S., Hull, C., Carpendale, S.: SurfaceMusic: Mapping Virtual Touch-based Instruments to Physical Models. In: Proceedings NIME 2010, pp. 360–363 (2010)
9. Garage Band, `http://www.apple.com/ilife/garageband/` (last accessed January 2011)
10. Gardner, M.: On cellular automata, self-replication, the Garden of Eden and the game life. Scientific American 224(4), 112–117 (1971)
11. Hochenbaum, J., Vallis, O.: Bricktable: A Musical Tangible Multi-Touch Interface. In: Proceedings of Berlin Open Conference 2009, Berlin, Germany (2009)
12. Hochenbaum, J., Vallis, O., Diakopolous, D., Murphy, J., Kapur, A.: Designing Expressive Musical Interfaces for Tabletop Surfaces. In: Proceedings of NIME 2010, pp. 315–318 (2010)
13. Hauert, S., Reichmuth, D.: Instant City, `http://www.instantcity.ch` (last accessed January 2011)
14. Jordá, S.: Sonigraphical instruments: from FMOL to the reacTable. In: Proceedings of NIME 2003, pp. 70–76 (2003)
15. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proceedings of TEI 2007, pp. 139–146. ACM, New York (2007)
16. Max/MSP, `http://cycling74.com/` (last accessedJanuary 2011)

17. Microsoft Corporation. Microsoft Surface, http://www.microsoft.com/surface/ (last accessedJanuary 2011)
18. Krueger, M.W., Gionfriddo, T., Hinrichsen, K.: VIDEOPLACE - An Artificial Reality. In: Proceedings ofCHI 1985, pp. 35–40. ACM, New York (1985)
19. Newton-Dunn, H., Nakano, H., Gibson, J.: Block jam: a tangible interface for interactive music. In: Proceedings of NIME 2003, pp. 170–177 (2003)
20. Noteput, http://www.jonasheuer.de/index.php/noteput/ (last accessed January 2009)
21. Patten, J., Recht, B., Ishii, H.: Audiopad: a tag-based interface for musical performance. In: Proceedings of NIME 2002, pp. 1–6 (2002)
22. Scott, S.D., Carpendale, M.S., Inkpen, K.M.: Territoriality in collaborative tabletop workspaces. In: Proceedings of CSCW 2004, pp. 294–303. ACM, New York (2004)
23. Sound Storm, http://subcycle.org/ (last accessed January 2011)
24. Stavness, I., Gluck, J., Vilhan, L., Fels, S.: The MUSICtable: a map-based ubiquitous system for social interaction with a digital music collection. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) ICEC 2005. LNCS, vol. 3711, pp. 291–302. Springer, Heidelberg (2005)
25. Stereotronic Multi-synth Orchestra, http://www.fashionbuddha.com/ (last accessed January 2011)
26. Taylor, S., Izadi, S., Kirk, D., Harper, R., Garcia-Mendoza, A.: Turning the tables: an interactive surface for vjing. In: Proceedings CHI 2009, pp. 1251–1254. ACM, New York (2009)
27. Villar, N., Lindsay, A.T., Gellersen, H.: Pin & Play & Perform: a rearrangeable interface for musical composition and performance. In: Proceedings of NIME 2005, pp. 188-191 (2005)

Presentation Technique of Scents Using Mobile Olfactory Display for Digital Signage

Sayumi Sugimoto[1], Ryo Segawa[1], Daisuke Noguchi[1], Yuichi Bannai[2], and Kenichi Okada[1]

[1] Faculty of Science and Technology, Keio University
Hiyoshi 3-14-1, Kohoku-ku, Yokohama, Kanagawa, 223-8522 Japan
{sugimoto,segawa,noguchi,okada}@mos.ics.keio.ac.jp
[2] Global Environment Promotion Headquarters, Canon Inc.
Shimomaruko 3-30-2, Ohta-ku, Tokyo, 146-8501 Japan
bannai.yuichi@canon.co.jp

Abstract. Understanding and attention value of the advertisement will be advanced by adding scents to the digital signage. However, it was difficult to have corresponding one-to-many relationships, movements of users, and precise chronological control of scents. In this study, using mobile olfactory display, we propose the digital signage with scent which takes account of users' movements. The concept of this study is constructing the system having scents, movements, and communication. This system was built by enabling to receive the scent ejection signal from the advertisement, achieve the distance by the image of web-camera, and eject the scents with the strength in accordance with distance, to have the control of scents which accedes to substance of advertisement and positional relationship. As a result of evaluations, the olfactory information was carried down to users with great accuracy. The scents production of the advertisements will be possible with the use of this system.

Keywords: Olfactory Information, Olfactory Display, Pulse Ejection, Olfactory Characteristics, Scented Digital Signage.

1 Introduction

Information and communication via computers tends to be limited to visual and audio information. However, transmission via all five senses (sight, hearing, touch, smell and taste) has lately attracting much attention [1]. Within those five senses, the olfactory information affects humans largely because the information achieved by the olfactory organs is transmitted to the cerebral limbic system, where governs emotions and memories. Therefore, olfactory information is thought that is has a large importance of providing high effect on humans [2] as well as 3D sound systems and 3D movies.

Trials on transmitting olfactory information together with audio/visual information are being conducted currently in the field of multimedia such as digital signage to

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 323–337, 2011.

have more lush expressions, diversified information, and higher sense of reality [3]. However, most of the researches on scents presentation using existing technique had limitations of olfactory devices and thus it was difficult to precisely control the scents chronologically and spatially. In the use of scents on digital signage, the accommodation toward the relationship of one-to-many and the movement of users were challenging because the existing technique could only present the scents to the limited space. Moreover, because only the control of ON and OFF was possible, the accordance of visual information and olfactory information was difficult and thus the setting of scented advertisements in close distance was almost impossible.

In this study, we aim at constructing the scented digital signage system which has been difficult to realize in the past. The concepts of this study are building the system with scents, movements, and communications. The system was built by enabling reception of the scent ejection signal from the advertisement, achievement of the distance by the image of web-camera, and ejection of the scents with the strength in accordance with distance, to allow the control of scents which accedes to contents of advertisement and positional relationship.

This paper is organized as follows: Section 2 introduces related work to our proposal, which is described in detail in section 3. Section 4 contains the descriptions of our devices, and section 5 shows our evaluation results. At last, section 6 concludes this paper.

2 Related Work

Trials on the transmission of olfactory information together with audio/visual information are ongoing. Work first started when Heilig developed Sensorama [4] in the 1950s. Sensaroma is the first virtual reality (VR) system that presented olfactory information together with audio/visual information. The effects of the scented movies towards the viewers were studied by Tomono et al [2, 5]. This study concluded that the understanding of the contents will be increased when the image of the movie and the scents are corresponded to each other. On the other hand, this study also concluded that if the image and the scents are not corresponded, it cause the disturbance and discomfort.

The another example of service that merged scents to movies, Nakamoto et al. [6] designed a smell synthesis device that presents the odor of a virtual object remotely. The system analyzes the smell to be transmitted and presents the analyzed data as the composition ratio of the odor elements. On the receiver side, a feedback control changes the ratio of the odor elements owned by the receiver to reproduce the target odor.

More, NTT Communications used the aroma-emitting device [7] which is able to save the recipes and delivery schedules to emit the scents in accordance with the image of movie from under the seats in movie theatre [8]. They also studied on the difference of passersby's action when the display of the shop has added the scents. As a result, it enhanced the impression of the store and achieved more attention on advertisement. The service for cell-phones which provide scents via internet is now constructing. Kaye's article [9] also describes some systems that add scent to web content, and computer controlled olfactory displays such as iSmell [10] and Osmooze [11] are utilized in these systems.

The wearable olfactory display with location sensor had been constructed by Yamada et al [12]. It controls the density of odor molecules; it can present the spatiality of olfaction in an outdoor environment. The olfactory information transmitting system consists of the aforementioned display, a sensing system using three gas sensors, and matching database. The user can experience a real sense of smell while they are walking around through the system by translating obtained olfactory information.

However, the scent presentation techniques of previous works were designed merely to create the sense of experiencing a scent, and they used higher densities and longer presentation times of scent than actually necessary, and thus were hindered by the problem of olfactory adaptation due to scent lingering in the air. To this end, we describe in the next section our proposed technique that makes it possible to experience multiple scents at the same time and which expresses the distance relation between objects as presented by the audio-visual information.

3 Characteristics of Olfaction

A fragrance substance is a compound that stimulates the olfactory cells in the nasal cavity. Fragrance substances can be inorganic substances such as hydrogen sulfide and ammonia, although most are organic compounds. It is said that of the approximately two million kinds of organic compounds in existence, about four hundred thousand of these have an odor [13]. However, humans perceive and recognize about five thousand scents routinely. The characteristics of human olfaction are now briefly described.

3.1 Olfactory Threshold

The olfactory threshold is the value used as a standard to express the strength and weakness of a scent. Three kinds of values are generally used for the olfactory threshold: the detection threshold, the recognition threshold, and the differential threshold [14], usually expressed in units of mol (concentration) and mass percentage. However, because the olfactory threshold is a measure of the lowest olfactory stimulus intensity at which an individual can perceive scent, this value does not reflect the intensity (strength and weakness) of the scent perceived.

Detection threshold: the smallest density at which scent can be detected and where the user does not need to recognize the kind of smell.

Recognition threshold: the smallest density at which the kind of scent can be recognized, and its value reflects the ability of the user to express quality and characteristics of the scent.

Differential threshold: the density at which the user can distinguish the strength of a scent, where its value reflects the ability of the user to detect changes in the stimulus and to quantify the change.

Generally, such changes are expressed as the % change of stimulation quantity of the original. In the case of olfaction, it differs with different kinds of scent, but is in the range of about 13-33%.

3.2 Adaptation

Adaptation is the phenomenon where sensory nerve activity is decreased by continuous stimulation by odor molecules. Adaptation itself and the speed of recovery from adaptation differ according to the kind of scent. Adaptation is gradually strengthened over time, but is restored for a short time (3-5 minutes) by eliminating the scent.

In addition, there are various patterns of adaptation, influenced by the kind of scent and recognition factors.

4 Proposal

Understanding and attention value of the advertisement are expected to be advanced by adding scents to the digital signage.

Researches on the effect of scents added to the digital signage have already been conducted, but it had no spatial controls over scent presentation. Thus, presentation of multiple scents to the space was difficult. Moreover, because the scents could only be presented to limited space, it was difficult to have corresponding one-to-many relationships and movements of users which can be seen in the system such as digital signage.

In this study, we propose the scented digital signage using the mobile olfactory display which is able to have precise control of scents. This proposal aims at the implementation of scent presentation corresponds to the advertising contents and positional relationship to users moving freely, which leads to the increase in users' understanding, interest and remembrance of contents. More, this proposal constructs the basement of scented media to become widely used.

The requirements to the proposal are described as follows.

- Minimizing olfactory adaptation and scent lingering in the air
- Scents presentation corresponds to contents and positional relationships
- Scents presentation corresponds to visual information

Olfactory adaptation and scents lingering in the air banefully affect the recognition of scents because of the characteristics of olfactory described above, and it causes the failure of scent presentation. Olfactory adaptation prevents the long-term presentation of scents, and scents lingering in the air cause the interfuse of multiple scents. The precise and complicated scents presentation will be possible when those effects were reduced.

The scents to be presented varies among the advertisements, thus the presentation requirements must be configured with respect to each advertisement. In other words, the structure, which enables to set kinds, strength, and ambit of the scents and eject them in accordance with users' movement, is needed. Also, the sense of perspective, whether the user is getting close to or moving away from advertisement, needs to be presented in order to tell the source of scents easily. These complicated scent presentation delivers the contents clearly and is expected to further the attention values and understanding of advertisement.

At last, olfactory information must be corresponded to visual information. When these information disaccord, users will sense the discomfort.

The methods to fulfill those requirements are described with details in the following sections.

4.1 Scents Ejection Policies

In this study we set the policies in order to realize the scent presenting corresponds to human's olfactory characteristics. The policies are:

1. Eject the 0.1sec scents pulse with the interval of 1.3 sec.
2. Eject one scent at a time
3. Amount of scents are increased by double series
4. Eject the scents in accordance with visual information

Policy 1 is to reduce the effects of olfactory adaptations and scents lingering in the air. Ejecting the scents by the pulse minimize the lingering scent diffuse in the air and diminish the olfactory adaptation by discrete stimulus [15].

Policy 2 is to handle multiple scents. Ejecting the multiple scents at the same time causes the interfusion and scents become unrecognizable. However, due to the results of the separable recognition threshold of scents, the kinds of scents could be recognized if the interval of pulses is more than 1.2 sec. [16]. Thus, by applying Policy 1 and Policy 2, we assumed that the multiple scents presentation become possible.

Policy 3 is for presenting the changes of scents' strength. Our previous research [17] showed that increasing the amount of scents in double series enables to present the sense of perspective. In this study, we divided the strength of scent into some levels and the amount of scents is doubled when the level of strength is increased. Here, we call the strength of the scents at the level 1 "base ejection amount". When the strength level is indicated as S and the base ejection amount as E, the total ejection amount is represented as follows.

$$B = E \cdot 2^{(S-1)} \qquad (1)$$

Policy 4 is for corresponding the visual information and olfactory information. Here, five senses of human are related to each other and the understanding and reality increases when corresponded information is given.

Ejecting the scents by following those 4 policies, it is able to present the scents based on olfactory characteristics of human because those policies are to reduce the effects of olfactory characteristics and correspond the changes of multiple scents' strength. Herewith, it is able to present the scents in accordance with users' movements and contents, which was difficult to realize with traditional scent presenting technique.

4.2 Scent Presentation Corresponds to Contents and Positions

In this section, we describe the presentation technique in accordance with contents and position in detail.

The concept of proposal is constructing the system with scents, movement, and communication. It receives the scent ejection signal from the advertisement and achieves the distance by the image of web-camera to present the scents which accedes to substance of advertisement and positional relationship. Traditional scented digital signage had not supposed the strengths and kinds of scents to change, which led them to present one kind of scent to the space. However, there is the need to transmit the information about the scent and the source of the scent to users in order to change the presenting scents in accordance with users' movement.

In this study, the advertisement communicates with users via the wireless device to send the information of scents. The scent information and scent ejection order are sent from the advertisement at regular intervals. The AR markers were used to discern the source of scent by using the web camera.

Figure 1 shows the implementation image of the system. As shown in the figure, the devices are set on the users and the contents and distance information is achieved. Figure 2 is the image of the experimental environment. The markers labeled A, B and C and were put on the wall as if it were the advertisement, and is sending the scent ejection order at interval of 1.3 sec. These orders include 4 kinds of information. In this study, according to 4 scent presenting policies, the width of pulse ejection is set at 0.1sec.

- Kind of scent represent the advertisement
- Kinds of scents that each marker represents
- Strength of each scent

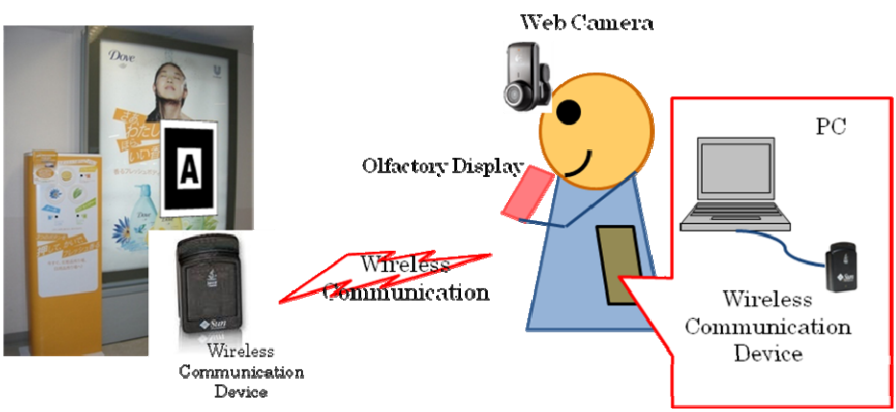


Fig. 1. Implementation Image

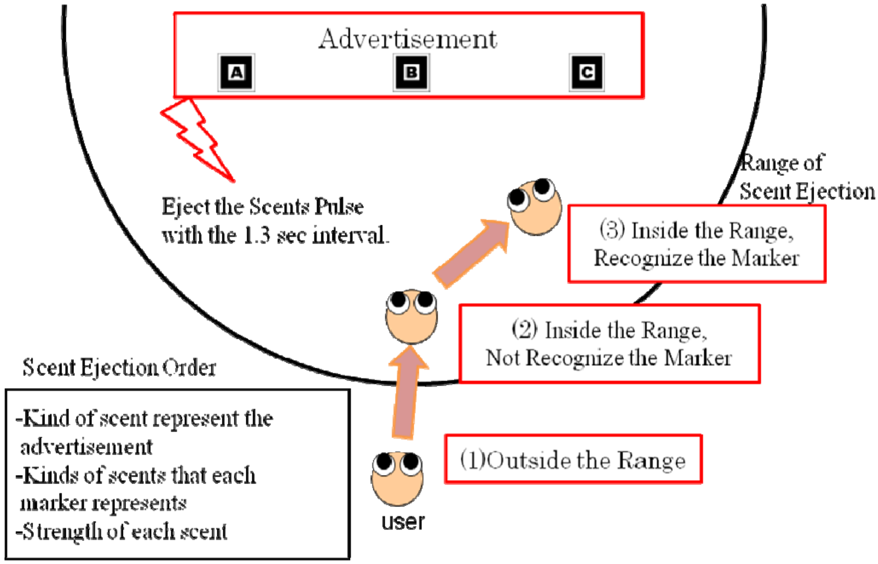


Fig. 2. Image of the experimental environment

The state (1) in figure 2 shows the case when user is outside the range of scent ejection, and the scent is not ejected.

The state (2) in figure 2 shows the case when user is inside the range of scent ejection but cannot recognize the marker, and the scent of advertisement is ejected. The purpose of this background scent is to increase the attention value of advertisement.

The state (3) in figure 2 shows the case when user is inside the range of scent ejection and able to recognize more than on marker. In this case, the scent is ejected by following the ejection priorities. Following the ejection priority and presenting the scent corresponds to the marker that user is looking at increase the understanding of contents because of the synergic effect of visual and olfactory information.

4.3 Ejection Priorities

In order to have corresponding between the visual information and olfactory information, we set the ejection priorities in this study. The olfactory priorities are the values range from 0 to 5 and contained by each marker and the background. The ejection priorities are update every time when it receives the scent ejection orders from the image of web camera. Then, the scent that holds highest ejection priority is ejected. When 2 or more scents have the same ejection priority, the scents of marker closest to the center is ejected.

The weight of the ejection priorities are as follows, and the weight and the percentage of the time that marker was detected are multiplied to calculate the updating ejection priority.

- center:+3
- in the range out of center：+2
- out of range:－ 1

Also, the ejection priority of background scent is set at 2.5, which is half of the maximum ejection priority. It is decrease by 2 when the scent is ejected.

Using those ejection priorities, it is able to present the scent of marker that the user is looking at.

4.4 Scent Strengths and Scent Range

We devised that the scent is ejected when the user looks at the marker. Here, to present the sense of perspective, the changes in range and strength of scents are required.

We supposed that the stronger scent has the larger range, and it is reflected on scent ejection order. The range is calculated from each of the markers. We indicate the base scent strength level as S, the scent strength decreased by the distance as S’, constant value as K, and distance as D in the formula. Here, S’ is in range of 1 to S.

$$D = K \cdot \sqrt[3]{2^{(S-S')}} \tag{2}$$

We took the cubic root because the scent chemicals diffuse three-dimensionally. Figure 3 shows the image of maximum ejection when base scent strength S is equal to 1 and 4.

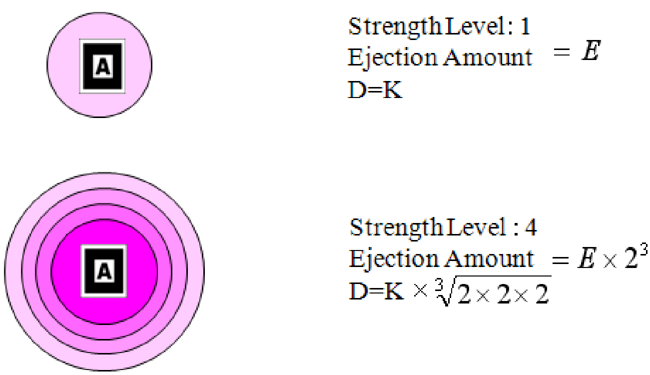


Fig. 3. Ranges and Strength of Scents

5 Devices

5.1 Mobile Olfactory Display

Figure 4 (a) shows the mobile olfactory display, and Figure 4 (b) shows the image of wearing the display.

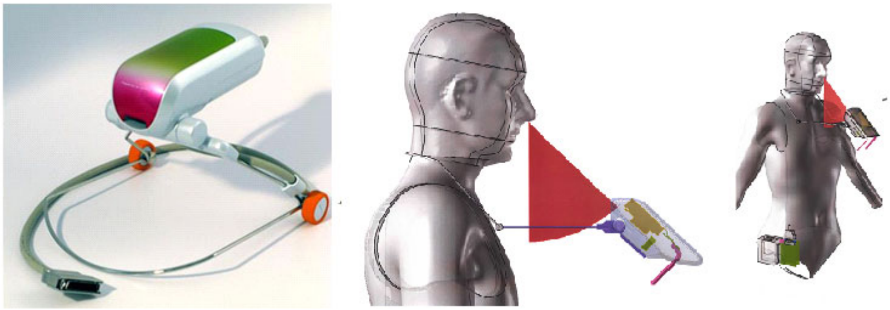


Fig. 4. (a) Mobile Olfactory Display (b) image of wearing display

This display uses the thermal method used in ink-jet printer to produce a jet which is broken into droplets from the small hole in the ink tank. The thermal method installs a heater in a part of the slight narrow tube filled with ink. In order to emit ink, bubbles are formed in the ink by instantaneous heating. This method is advantageous for miniaturization because the structure is simple.

An ejection heater is installed in the liquid route which continues to the exit hole. The ejection heater runs hot so that bubbles are generated in scent in the liquid route. Through the generation of pressure, the scent head emits scent from the exit hole. The scent fills the liquid route and liquid room behind. The scent in the liquid room is supplied along the supply route from the scent tank. By the micro fabrication technique, the production technology of the head which has more 100 exit holes was established. This head can emit scent of several pico-liters.

The unit average ejection quantity from one small tank is 4.7 pl and one large tank is 7.3 pl. It was confirmed to be approximately constant without depending on the residual quantity of ink on examination. In addition, the user can set the ejection number of times from one hole in 100 msec to 1-150 times, which we denote the "volume". By setting the volume, the ejection control is possible for a unit of 667 micro sec.

The display can set up 1 scent-ejection head. This head can store three small tanks and one large tank, thus this display can contain 4 kinds of scents maximum. There are 127 minute holes in the head connected to the small tank and 255 minute holes in the head connected to the large tank. Moreover, the display can emit scent from multiple holes at the same time, so the ejection quantity is adaptable to 0-127 (small tank), 0-255 (large tank).

Also, the display is equipped with a fan and the maximum wind velocity is 1.0m/sec. The scent presentation hole is a rectangle of 2 cm length and 6 cm width.

The battery is designed to be hanged around the waist, and the mobile olfactory display is designed to wear around the neck. The angle of the display is adaptable to 0-45degrees.

Because the quantity of scents emitted is small and diffuse quickly, this invented olfactory display is suited for personal use.

5.2 SUN Spot

We used SunSPOT from Sun Microsystems as a wireless communication device. SunSPOT uses Zigbee to transmit the information without any losses within range of 10meters. We first placed the SunSPOT remote sensor on advertisement and connected the SunSPOT base station to the laptop. This enables the communication between the advertisement and the user. Also, when the web camera could not recognize the AR marker, the distance between advertisement and the user is calculated by RSSI (Received Signal Strength Indication).

5.3 Web Camera

We used the web camera named 2-MP Portable WebcamC905m from Logicool. The frame rate of this camera is 30 fps at maximum, and the angle of view is 76 degree. This camera is used to recognize the AR marker to measure the direction and distance to advertisement of user. The distance was calculated from the number of the height pixels of the marker.

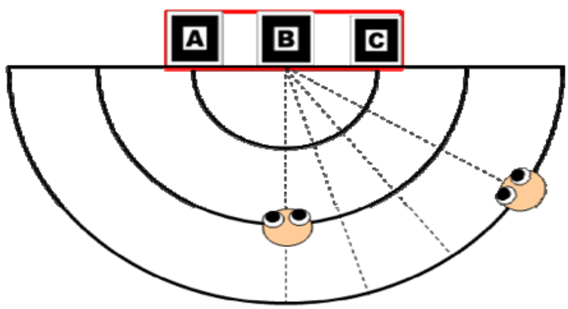


Fig. 5. Experimental environment for Experiment 1

The participants were asked to mount all hardware for two reasons. The first reason is our focuse on presenting the scents of the object that the user is looking at in the personal space. In order to find the object looked at by the user, the webcam must be worn by the user, and to present the scent personally, the olfactory display must be worn by the user to avoid the presentation of scents to non-users. The second reason is the mechanism of the devices. The olfactory display we used must be connected to the computers to be controlled, and so does the receiver of the wireless communication device to receive the signal from the device set on the advertisement.

6 Evaluation and Discussion

6.1 Evaluation on Achieving the AR Markers

The purpose of this experiment is to evaluate the acquisition accuracy of the marker and to consider the range of implementing proposed system. If the marker that user is looking at and the marker that web camera achieved had large difference, it is difficult to observe Policy 4. The experiment was conducted with 11 participants.

Figure 5 shows the overhead view of experimental environment. As shown in Figure 5, the AR makers labeled A, B, and C and were placed on the wall, as if it is an advertisement, with equal distance among them. Then, the participants were asked to look at the marker which was indicated, and evaluated the image from web camera to see whether the camera catches the indicated marker at the center. At the same time, the information about which marker is retrieved by the webcam is shown on the display of computer. By calculating the accuracy, the data is collected.

Here, the positions of the participant were fixed to 12 points, and the experiment was conducted by changing the distance from wall to the user, angle, and distance between the markers. Each of the conditions is as follows.

- Distance to the marker; 1m, 2m, 3.
- Angle related to the marker: 0, 20, 40, 60,
- Distance between markers; 1.0m, 0.5m

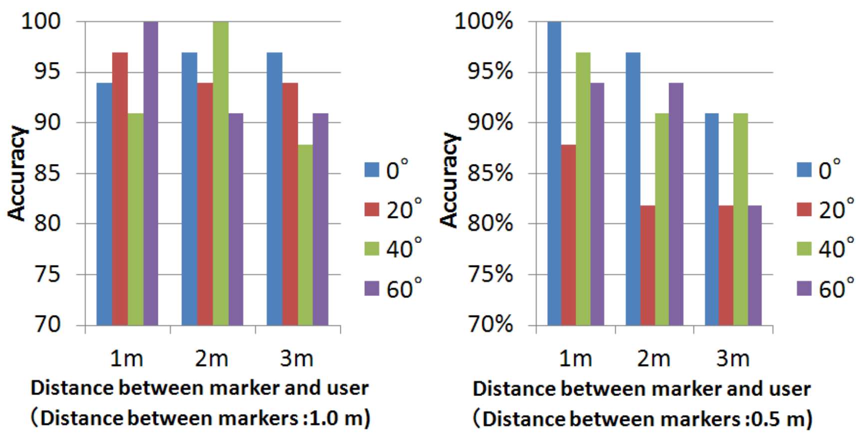


Fig. 6. Result of Experiment 1

The result is as shown in Figure 6. The x-axis is the distance between user and markers, and the y-axis is the accuracy of setting the marker in the center of image.

The result shows that the accuracy is more than 80% for all the conditions. Also, it indicates that the accuracy declines as the user move away from the wall.

Then, the data according to distances between markers are analyzed using two-way ANOVA (distance to the marker, angle related to the marker). The analysis showed that there were no significant differences when the distance of markers were 1.0 m. It means that the accuracy of setting the marker in the center of image is almost same at any position. On the other hand, there were significant difference when the distance of markers were 0.5m ($p<0.5$). However even, the worst accuracy of setting the image in center was 80%, thus the effect on implementation is small.

6.2 Evaluation on Transmitting Olfactory Information

The purpose of second experiment is to evaluate whether the olfactory information is transmitted to the user accurately and is contributed to find the source of the scent. The experiment was conducted to 14 participants.

10 AR markers were put on the wall with equal distance among them, and the level 1 scent of peppermint was presented all the time as a background scent. Under this environment, participants were asked to move around freely to find the marker with scent of lavender or lemon. Then the ratio of collect answers were calculated and analyzed. The experimental conditions were as follows.

1. 1 scent (distance among marker is 20 cm)
2. 2 scent (distance among marker is 20 cm)
3. 1 scent (distance among marker is 40 cm)
4. 2 scent (distance among marker is 40 cm)

Table 1. The strength level scents and their range

Strength Level	Range of Scents (m)
1	3.1
2	2.5
3	2.0

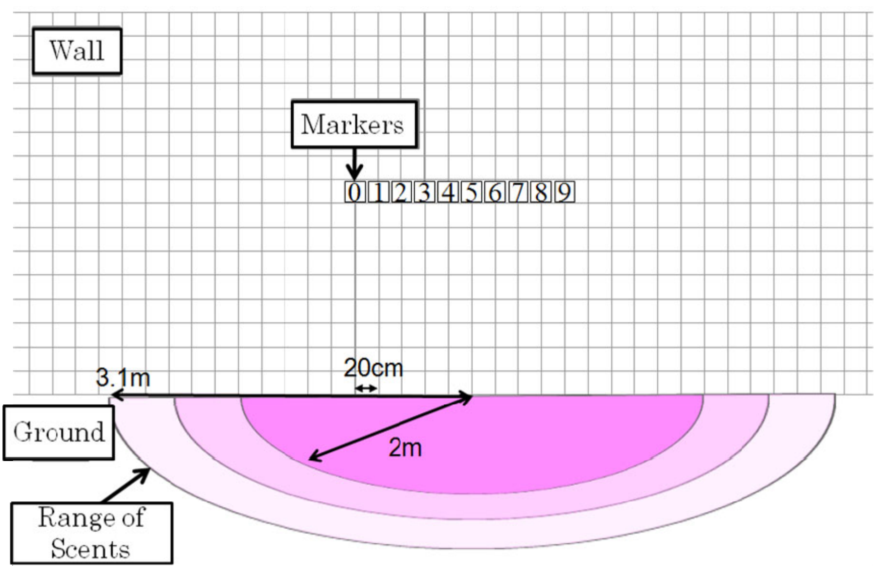


Fig. 7. Image of the range and perspective of scent from obliquely upward

The scent of lavender (strength level 3) was used when only one scent is presented, and the scents of lavender (strength level 3) and lemon (strength level 3) were used when two scents are presented. For all experiment, each of the scent of lavender or lemon was put on one out of 10 AR markers.

The levels of the scents were changed in accordance with the distance from the marker as shown in Table1. Figure 7 is the image of the range and perspective of scent from obliquely upward when experimental condition 1, when the scent of lavender was put on the AR marker labeled 5 is conducted.

The result is shown in Figure 8. The y-axis represents the percentage of correct answers of finding the scent source. From this graph, it is implied that larger the distance among markers, larger the percentage of correct answers.

Then, Then, the data are analyzed using two-way ANOVA (distances between markers, number of scents presented). The analysis resulted that the distance among markers had the significant difference ($p<0.01$). It indicates the difficulty of finding the source of scents by olfactory when the AR markers were placed in the distance of 20cm. On the other hand, it is accurate to find the scents source when the AR markers are placed with the distance of 40 cm.

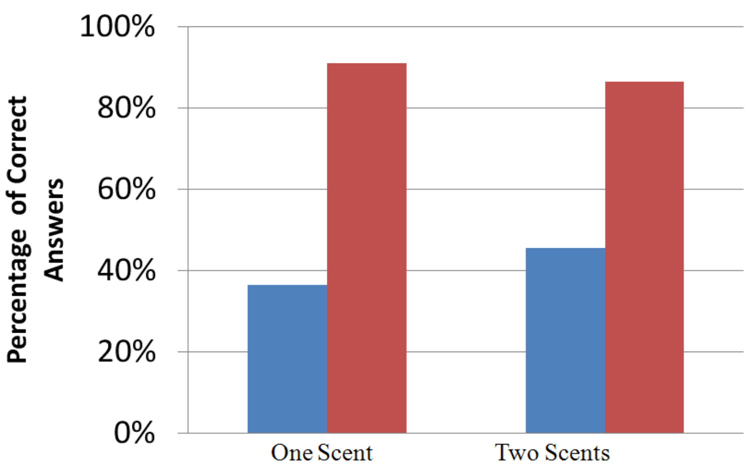


Fig. 8. Results of Experiment 2

Moreover, there was no significant difference based on the numbers of scents ($p>0.05$). This means that the presentation and transmission of multiple scents is possible with the same environment which single scent is presented.

From the overall result of this experiment, it can be concluded that transmission of olfactory information and searching of scent source are possible when the AR markers were placed with the distance of 40cm. Also, when the distances of markers were 20cm, it can be assumed that the accuracy will be increased when visual information is added.

7 Conclusion

Studies on transmitting information via five senses are currently conducted to present more realistic experiences in the field of multimedia. Presenting the olfactory information together with visual and audio information increases understanding and attention value of the contents, and thus the studies on adding the scents to digital signage has been conducted.

However, many of the researches have the limitation of olfactory displays, and it was difficult to control the scents chronologically and spatially. Moreover, because of the control difficulties, presenting more than one scents from the advertisement was almost impossible.

In this study, to overcome the scent control difficulty, we proposed the scented digital signage using mobile olfactory display. This system not only adds multiple scents on one advertisement but also present the scents in accordance with users' movement, advertisement contents, and spatial relationship.

We first proposed 4 policies of scent ejection based on our previous studies. Then we implemented the system with scents, movements, and communication observing the policies. This presents the scents according to the user's movement and advertisement's contents.

From the results of experiments, we verified the ability of this system. This system is able to present multiple kinds of scents and to find the source of scents only by using olfactory information when the AR markers were more than 40cm apart.

As the future work, the appliance of moving picture ads and the revision of policies in accordance with the moving picture ads as well as the experiments with multiple users.

Acknowledgement. This work is supported in part by a Grantin-Aid for Scientific Research(B) ,2011, 23300049 from the Ministry of Education, Culture, Sport, Science and Technology in Japan.

References

1. Kim, J.-D., Kim, D.-J., Han, D.-W., Byun, H.-G., Ham, Y.-K., Jung, W.-S., Park, J.-S., Oh, S.-K.: A Proposal Representation, Digital Coding and Clustering of Odor Information. Computational Intelligence and Security 1, 872–877 (2006)
2. Tomono, A., Yamamoto, S., Utsunomiya, M., Ikei, D., Yanagida, Y., Hosaka, K.: Effect that the Image Media with Scent Gives to Contents Understanding. In: Human Interface Symposium 2004, pp. 249–254 (2004)
3. Hirose, M., Tanigawa, T.: Information and Communication of Odor. Number 21 in Aroma Science Series. Fragrance Journal Ltd. (2007)
4. Retrofuture: Sensorama's pre-virtual reality, `http://www.retrofuture.com/sensorama.html`
5. Tomono, A., Kanda, K.: Approach of Improving Contents Quality by Smell Addition to Image. Journal of Institute of Electrical Engineers of Japan 128(12), 478–486 (2008) (in Japanese)
6. Nakamoto, T., Nakahira, Y., Hiramatsu, H., Oriizumi, T.: Odor Recorder Using Active Odor Sensing System. Sensors and Actuators B 76, 465–469 (2001)

7. NTT Communications to Test Aroma-Emitting Digital Signage, http://www.ntt.com/release_e/news07/0010/1017.html
8. NTT Com to Pilot Test Mobile Fragrance Communication Service, http://www.ntt.com/release_e/news08/0004/0407.html
9. Kaye, J.: Making Scents: aromatic output for HCI. Interactions 11(1), 48–61 (2004)
10. Edge Review: DigiScent Ismell, http://www.edgereview.com/ataglance.cfm?category=Edge&ID=136
11. http://www.osmooze.com/
12. Yamada, T., Tanikawa, T., Hirota, K., Hirose, M.: Wearable olfactory display: Using odor in outdoor environment. In: Proc. VR 2006, p. 199 (2006)
13. Kawasaki, T., Nakajima, M., Tonoike, M.: Aroma Science Series 21 Characteristics andAnalytical Estimation of Smell Material. Fragrance Journal Ltd (2003) (in Japanese)
14. Term and Commentary of Odor, Odor Control Association of Japan (2003) (in Japanese)
15. Sato, J., Ohtsu, K., Kadowaki, A., Bannai, Y., Okada, K.: Scent Presentation Technique by Pulse Ejection to avoid Olfactory Adaptation. Information Processing Society of Japan 49(8), 2922–2929 (2008)
16. Kadowaki, A., Sato, J., Ohtsu, K., Bannai, Y., Okada, K.: Measurement of Olfactory Responses to Odor Stimulation Switched in Single Breath. Journal of Japan Association on Odor Environment 40(1), 1–10 (2009) (in Japanese)
17. Ohtsu, K., Sato, J., Bannai, Y., Okada, K.: Scent Presentation Technique to Enable a Creating of Dynamic Perspective. Information Processing Society of Japan 50(4), 1435–1443 (2009)

“Oh Snap” – Helping Users Align Digital Objects on Touch Interfaces

Jennifer Fernquist, Garth Shoemaker, and Kellogg S. Booth

Department of Computer Science
201-2366 Main Mall, The University of British Columbia
Vancouver, BC Canada V6T 1Z4
{adara,garths,ksbooth}@cs.ubc.ca

Abstract. We introduce a new snapping technique, Oh Snap, designed specifically for users of direct touch interfaces. Oh Snap allows users to easily align digital objects with lines or other objects using 1-D or 2-D translation or rotation. Our technique addresses two major drawbacks of existing snapping techniques: they either cause objects to “jump” to snap locations, preventing placement very close to those locations, or they “expand” motor space so that on direct-touch interfaces objects lag behind the user’s finger. Oh Snap addresses both of these problems using an asymmetric velocity profile similar to a technique for filtering degrees of freedom in multi-touch gestures that was introduced by Nacenta et al. (2009). Oh Snap applies the velocity profile to multiple “snapping” constraints. A user study revealed a 40% performance improvement over no snapping for 1-D translation, 2-D translation, and rotation tasks when snap lines or angles were targeted. We found that Oh Snap performs no worse than traditional snapping, while retaining its important functional benefits. The study also investigated optimal parameter settings and Oh Snap’s accuracy in supporting the placement of objects near to, but not at, snap locations, which traditional snapping techniques do not support. Oh Snap was found to be competitive with non-snapping interfaces for these tasks.

1 Introduction

Touch interfaces, such as multi-touch tabletops, interactive wall displays and mobile devices, are growing in popularity. As a result, many researchers are investigating their usefulness for completing an increasingly diverse collection of tasks, including: the control of robots [19], the control of systems [8, 21], managing artifacts [2, 3, 13], and software engineering [9]. Most of these systems support the selection and manipulation of digital objects on the screen using direct touch, exploiting the naturalness of physical direct interaction. For example, users may touch and drag a digital robot to position its real-world counterpart, move and group digital documents, drag a 1-D slider, or rotate a dial.

Unfortunately, touch interfaces are sometimes not well suited to precise manipulation. The “fat finger problem” [27] makes selection of specific targets or placement of digital objects at precise locations or orientations difficult. Shortcomings

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 338–355, 2011.

in current sensing technology and the difficulty inherent in resolving touch contacts also contribute to the problem.

There have been several techniques developed that attempt to facilitate precise touch interactions on both large and small touch interfaces [2, 3, 5, 22, 23, 29, 34]. Work has also been done to develop better methods for manipulating digital objects [15, 17, 25]. However, the fundamental issue surrounding the fat-finger problem remains. Very little has been done to improve the alignment and precise positioning of digital objects in a touch environment.

Alignment tasks in a computer interface are often assisted with "snapping" techniques [6, 7, 10, 24]. This dates back at least as far as Ivan Sutherland's groundbreaking Sketchpad system [28], which included snapping constraints. Snapping techniques are one of the most common object alignment methods, and are widely used in computer-aided design (CAD) and other drawing programs. Traditional snapping causes digital objects to instantaneously "jump" and then "stick" to a line or grid point that has snapping capabilities once the object is within some threshold distance from the snap location.

While this technique is sometimes sufficient for use with relative input devices, it is less suited to direct touch interfaces. Additionally, locations within the threshold area near snap locations may be inaccessible: if a user wishes to position a digital object within that area, the snapping functionality must be turned off. Toggling snapping functionality is onerous at best for single users, but it is even more of a problem on large collaborative displays, where the management of multiple widget sets and user states is a perennial problem [26].

More subtle snapping techniques have been developed, such as snap-and-go [4]. These permit digital object positioning near snap locations. Snap-and-go works by expanding motor space at a snap location, resulting in objects that stop, rather than jump. However, this method is not suitable for direct touch interfaces, because it can break the correspondence between finger and object.

In this paper we describe a new snapping technique, Oh Snap, designed specifically for direct touch interfaces. We employ a technique first introduced by Nacenta et al. [20]. Our technique is designed to support quick snapping to any one of a set of lines, angular orientations, or other constraints, while still allowing objects to be positioned in close proximity to one another and while maintaining a close correspondence of the user's finger with the dragged object throughout. This set of benefits is unique to our technique. Oh Snap provides a subtle snapping effect that needn't be explicitly enabled or disabled. It avoids limitations in existing alternatives and thus facilitates placements that other techniques do not. We compare our work to the earlier work by Nacenta et al. [20] in more detail after describing the new technique and a two-phase user study that was conducted to assess our technique.

2 Related Work

Many researchers have investigated approaches for supporting direct manipulation with objects on touch surfaces. Wobbrock et al. [33] investigated user-defined gestures for general interactions on multi-touch tabletops and found that touching-and-dragging is the most natural method for translating digital objects, and that dragging by the corner is the most natural way to rotate objects. Similarly, Micire et al. [19] conducted an analysis of user-defined gestures for robot manipulation on a multi-touch tabletop. They too found that touch dragging was the most used gesture for positioning robots. Kruger et al. [15] and Liu et al. [17] developed additional methods for performing fluid rotation and translation of objects using direct touch.

Many tasks designed for a touch interface can benefit from precise positioning of digital objects. Nóbrega et al. [21] created LIFE-SAVER, an interactive visualization system for a touch interface to analyze emergency flood situations. Studies by Bjørneseth et al. [8] used a touch table for dynamic positioning of maritime equipment. This safety-critical task requires careful translation and rotation for specifying vessel positioning and heading, respectively.

2.1 The "Fat Finger" Problem

Many touch devices capture a touch contact 'point' that is actually a relatively large 2-D region [27]. This is often converted into a single (x,y) pixel coordinate for compatibility with traditional pointing models that assume a single point of interaction. However, there is no guarantee that this single point is the true contact point intended by the user. Despite many advances in technology, this problem persists in part due to a lack of sophisticated sensing techniques, but also because the intended point of interaction is inherently ambiguous. Techniques such as "focus+context" lenses have been designed to help mitigate this problem within information visualization applications [30], but no general solution exists for all types of applications.

Sensing limitations may give rise to a variety of issues when manipulating objects, such as unintentional movement of the object. For example, when a finger lifts up from the screen its contact area changes shape, which may result in a change to the calculated pixel coordinates. If a user were attempting to precisely place a digital object, this might cause the object to shift from the desired target.

There has been work to resolve the fat finger problem by obtaining a more accurate touch contact point [12, 31], providing feedback to the user about the success/failure of the touches [32], and incorporating selective zooming [29]. Although these techniques can provide a more accurate touch contact location, they do not assist substantially in object alignment tasks.

2.2 Existing Snapping Techniques

Traditional snapping techniques, such as snap-dragging [7], cause objects to automatically jump to snap locations once they are within a predefined distance from the snap location. Basic snapping is highly effective in assisting alignment tasks; however, while it is sometimes sufficient for use with relative input devices, such as computer mice, it is less well suited to direct touch interfaces. It is highly unintuitive if an object a user is touching suddenly jumps around underneath the user's finger.

Worse yet, traditional snapping does not support the placement of objects near to, but not exactly at, snap locations. Disabling snapping can address this problem, but explicitly toggling snapping is highly undesirable for a touch interface, especially in a collaborative environment. The snap toggling function would either have to be a global function, affecting all users, or a local function, which would require additional information to determine the identity of the activator. It might also have to be placed in one or more menus, accessible to all users, or activated with a possibly complex gesture, requiring at least some additional training of users [11].

Snap-and-go, introduced by Baudisch et al. [4], is a snapping technique that does not require toggling on or off. It functions by expanding motor space at snap lines, resulting in objects stopping at the desired location as opposed to automatically jumping there once they are within some threshold distance. This allows objects to be placed near snap lines as well as directly on them, unlike traditional snapping.

Snap-and-go works well for relative input devices, such as mice. With relative devices, snap-and-go stops a dragged object at the snap line as the user keeps moving the mouse beyond the snap line. After a short distance, the object begins moving with the mouse again. Unfortunately, snap-and-go is not suitable for direct touch interfaces where the correspondence between a user's finger and an object under manipulation should ideally be maintained at all times. If a user were to drag an object across a snap-and-go line, the finger would permanently move out ahead of the object. Indeed, the more snap lines an object crosses, the farther the object would lag behind the finger that was dragging it, in effect "losing" any direct object-finger correspondence.

There have been other attempts to solve these problems. Pseudo-haptic feedback has been used to improve interactions with graphical user interfaces, causing screen widgets to "feel" sticky, magnetic, or repulsive [16]. Researchers have developed sticky widgets [18] and "force fields" [1] to help with window alignment in mouse-based environments. These ideas could be adapted to the translation and alignment of objects on touch tables; however, they would suffer from the same drawbacks as snap-and-go, because the finger could "lose" the object. The problem remains of maintaining a close correspondence between a user's finger and the object that is being manipulated.

3 The "Oh Snap" Technique

Oh Snap is a snapping technique designed specifically for touch interfaces. It possesses several benefits, including: it eases alignment of objects with snap lines, it doesn't require toggling modes, it maintains the correspondence of finger to object, and objects can be placed close to snap lines or other objects without snapping interfering.

The basic idea behind the Oh Snap technique is shown in Fig. 1. To begin, a user drags an object as she normally would until the object first touches a snap line, at which point the object stops moving even if the user does continue to drag her finger. The object remains stationary unless the user's finger travels a small distance (the *snap-width*) beyond where snapping has occurred. Once the finger travels beyond the *snap-width*, the object starts moving at a rate faster than the finger is moving. Once the finger travels further, beyond the *catch-up width*, the object will have caught up to the finger, and dragging continues as usual. Of course, if the user lifts her finger while the object is snapped to the line, the object remains in its aligned position.

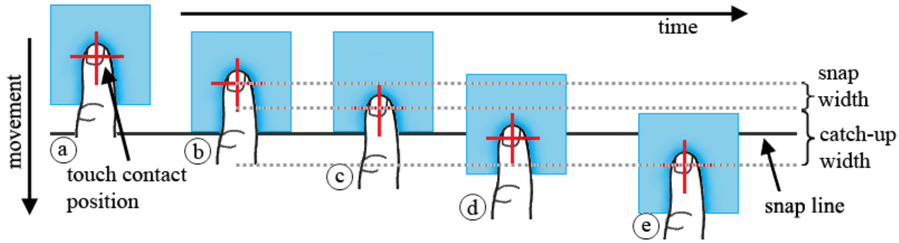


Fig. 1. As a finger moves an object downward to a snap line (a) the object "snaps" when its leading edge touches the snap line (b). As the finger continues downward, the object remains snapped to the line (c). When the finger is beyond the snap width, the object un-snaps and starts catching up to the finger (d). When the finger reaches *snap width* + *catch-up width* the object has returned to its original position relative to the finger (e).

During the catch-up phase an object travels at a rate faster than the finger. The motion of an object in the catch-up phase is defined by a linear interpolation that calculates the object's position proportional to the finger's position within the catch-up region. Pseudo code for the algorithm is shown in Fig. 2. The number of pixels the object travels for each pixel the finger travels is determined by the ratio calculated in Eqn. 1, which is shown on the left in Fig. 3. This is the rate at which an "un-snapped" object catches up to the finger. The ratio can also be considered to be the size of a *super pixel* relative to a real pixel, the distance moved by the object each time the user's finger moves one real pixel.

```
linearInterpolation(fingerX, fingerOriginX, snapWidth,
    catchUpWidth)
{
    return fingerOriginX + (fingerX-fingerOriginX-snapWidth) *
        (snapWidth + catchupWidth) / catchUpWidth;
}
```

Fig. 2. Code fragment for the linear interpolation function that returns the position of the object moving in the x direction if the object is snapped and the finger position is in the 'catch-up' area. In this code `fingerX` is the current finger position, and `fingerOriginX` is the position the finger was in when snapping occured.

A temporal diagram of the Oh Snap technique is shown in Fig. 3. The graph shows how an object first travels normally, how it then snaps to a line while the user's finger is in the snap region, and how it eventually catches back up to the finger because it travels faster than the finger in the catch-up region.

3.1 Snap Width and Catch-Up Width

The ratio in Eqn. 1, and thus the size of the super pixels, must be carefully chosen. When the *catch-up width* is very large, the ratio approaches one so objects effectively never catch up. Conversely, when the *catch-up width* is close to zero, the ratio approaches infinity and objects jump to their original position underneath the user's finger as soon as they un-snap. This can make it difficult to position an object at a location closer than the *snap width* to a snap line.

ratio = ((snap width) + (catch-up width)) / (catch-up width) (1)

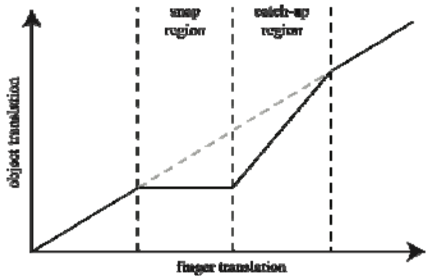


Fig. 3. The relationship of object translation relative to finger translation as an object moves, first normally, then snapped (stationary) in the snap region, and eventually catching up to the finger again as it leaves the catch-up region

Ideally, the magnitude of *catch-up width* + *snap width* should be less than the width of an average touch contact. Wang and Ren [31] found this to be 36 pixels (0.4mm/pixel) for an oblique touch from an index finger. The *snap width* should be large enough to accommodate users overshooting a target, otherwise objects un-snap before the user's finger has a chance to stop. A balance must be struck so that catching-up is imperceptible but still occurs quickly enough to be useful. If the ratio is too close to one, some positions near a snap line can be unreachable, due to quantization effects, unless the touch sensors provide sub-pixel resolution. For example, if the ratio (and super pixel size) is 2, a position 3 pixels away from a snap line is unreachable, because the object will be moving in steps of 2 pixels as the finger moves in steps of 1 pixel. This can be mitigated either by lifting the finger right after crossing the snap line, which flags the object as un-snapped, and then putting it down again and resuming with normal dragging, or dragging the object far enough away from, and then back towards, the snap line. Neither seems like a very good solution, which emphasizes the need for proper selection of *catch-up* and *snap widths*. We discuss the selection of these parameters in section 5.2 and evaluate three different parameter sets in the second phase of our user study.

3.2 Benefits of the Oh Snap Technique

The benefits of Oh Snap are summarized in Table 1. First, Oh Snap preserves the position of the user's touch point on a digital object relative to that object. This feature is especially useful if users drag objects across snap lines when they have no intention of aligning the object to those lines. Although the object will temporarily snap to those lines as it crosses them, the object will eventually catch up with the user's finger and return to its original relative position underneath it. This is crucial for touch tables or other direct touch interfaces. Second, Oh Snap allows users to place digital objects near snap lines as well as align them with snap lines without having to toggle the snap capabilities on and off. This is important in collaborative environments where toolbars that may hold the snap toggle might not be accessible (reachable) by some users. Lastly, because Oh Snap supports all object positioning tasks, there is no need to incorporate mode switching functionality into the interface.

Table 1. Comparison of snapping techniques

Technique	Fast snapping	Mapping maintained	Close placement
Oh Snap	Yes	Yes	Yes
snap-and-go	Yes	No	Yes
traditional	Yes	Yes	No
no snapping	N/A	Yes	Yes

4 Implementation

We implemented Oh Snap in C#, running on a SMART Table. We implemented three object movement types that are frequently performed on touch tables: 1-D translation, 2-D translation, and rotation. Rotation has been shown to be very useful on touch tables, especially in collaborative situations [14]. For translation, the position of each edge of an object is checked against the position of all environment lines it is parallel to. Orthogonal edges in the 2-D environment can be snapped independently. For rotation, the snap width is measured as an angle rather than as a pixel count.

```
OhSnap(objectBorder, snapLineX, snapWidth, catchUp){
   if (objectBorder.rightX == snapLineX && !isSnapped){
      isSnapped = true;
      fingerPositionSnapped.x = currentFingerPosition.x;
   }
   if (isSnapped){
      fingerDiff = currentFingerPosition.x -
                   fingerPositionSnapped.x;
      if (fingerDiff <= snapWidth)
         return snapLineX;
      else if (fingerDiff > snapWidth &&
               fingerDiff <= snapWidth + catchUp)
         return linearInterpolation(
            fingerPositionSnapped.x, currentFingerPosition.x,
            snapWidth, catchUp);
      else
         isSnapped = false;
   }
   return objectBorder.rightX;
}
```

Fig. 4. Code fragment for the 1-D Oh Snap function for the case when the object is moved from left to right (lower to higher X coordinate). In this code `objectBorder` is a variable that references the boundary of the object and `rightX` is the x position of its rightmost edge. Additional cases are implemented similarly.

If a user snaps an object to a line and then lifts her finger from the screen, the object is no longer flagged as being snapped, whether it is aligned with the line or is within the (*snap width*) + (*catch-up width*) region. This is primarily useful if a user wishes to place an object near a snap line but has snapped the object and is having difficulty reaching the destination due to the object moving along super pixels. A code snippet for the 1-D Oh Snap function that handles an object moving from left to right is presented in Fig. 4. This uses the linear interpolation function in Fig. 2.

5 User Study

In order to objectively evaluate the performance of the Oh Snap technique, we performed a user study. The first phase focused on comparing the performance of Oh Snap to traditional snapping and to no-snapping for alignment tasks. The second phase investigated participants' ability to drag an object close to but not onto a line (i.e. to not snap the object to the line) when the Oh Snap feature is turned on.

5.1 Comparison of Snapping Techniques – Phase 1 of the User Study

The purpose of this phase of the user study was to evaluate the Oh Snap technique and compare its performance to traditional snapping as well as to no snapping. We investigated performance with three object movement types: 1-D translation, 2-D translation, and rotation.

Task. A participant's task was to use a single finger on the right hand to move a digital blue square, as quickly as possible, so that its edge(s) were aligned with the desired target line(s). The target lines were indicated with a black arrow, which turned green when the object was aligned with the indicated line. When all edges were aligned correctly (1 edge for the 1-D and rotation tasks, 2 edges for the 2-D task), the square also turned green, indicating success. Screenshots of the start and end of successfully completed 1-D, 2-D, and rotation tasks are shown in Fig. 5.

At the start of a trial, a button at the bottom of the screen labeled 'GO' became active after one second. When the active button was touched, the button disappeared, the square became active (indicated by its changing color from grey to blue), and the participant could then begin the trial. The timer began when the participant first touched the square and it stopped when the participant lifted the finger from the screen with the appropriate edge(s) of the square aligned to the target line(s). Each trial required successful alignment. If a participant lifted the finger when the square was not fully aligned, the participant would have to touch the square and move it again to complete the trial.

The layouts of the three tasks were designed so that their components occupied the lower portion of the screen. This allowed participants to sit at the table and comfortably reach the digital objects. For the 1-D translation task, participants moved the square so that its right edge was aligned with a single target line. The right edge of the square began at a distance 305 pixels (17cm) to the left of the target line. In the 2-D translation task, the square was to be aligned so that its top edge and left edge were aligned to horizontal and vertical target lines, respectively. At the start of a trial, the top left corner of the square was 490 pixels (27cm) from the intersection corner of the target lines. In the rotation task, participants rotated a rectangle, anchored at one end, so that the central line protruding from it was parallel to the target line. The target line was horizontal and the anchored rectangle began vertical at the start of a trial, rotated 90 degrees from the target line. A pilot study showed that crossing multiple distracter snap lines did not affect performance time. To simplify instructions to participants we did not use distracter lines in the final version of the study.

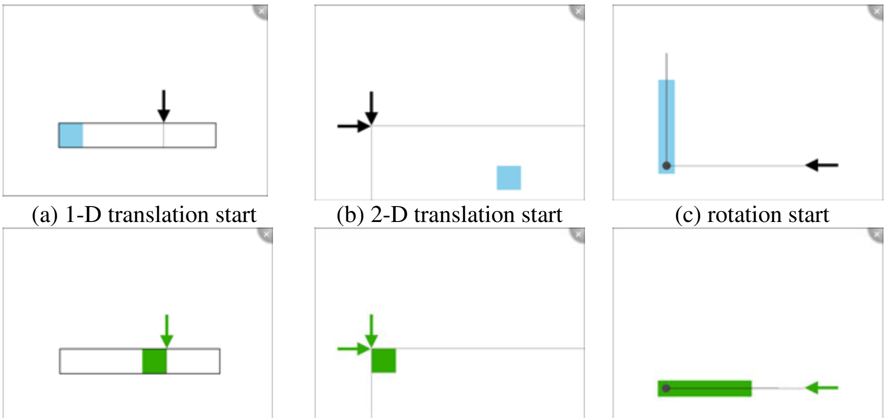

(a) 1-D translation start (b) 2-D translation start (c) rotation start
(d) 1-D translation end (e) 2-D translation end (f) rotation end

Fig. 5. Screenshots of the three Phase 1 tasks at the start and end of a trial

Interface Techniques. There were three snapping interface conditions: *no snapping*, *Oh Snap*, and *traditional snapping*. Trial tasks were identical across conditions, but the snapping behaviour of the target lines differed. In the no snapping condition, the target line did not cause objects to snap, and the edge of the square had to be placed within ± 2 pixels of the appropriate target line(s). Pilot testing revealed that the no snapping tasks were nearly impossible to complete without a small tolerance due to the touch sensing limitations of the SMART Table and the fat finger problem.

In the traditional snapping condition, the target lines had the traditional snapping behaviour with a snapping threshold of 10 pixels so that if the appropriate square edge was within that threshold of the target line, it would automatically be translated to the target line position. In the Oh Snap condition, the target lines had the Oh Snap behaviour with a snap width set to 10 pixels and the catch-up width set to 10 pixels. To conduct a fair comparison, we gave Oh Snap and traditional snapping the same threshold to make them as equivalent as possible.

Experimental Design. The study was a fully counter-balanced, within subjects 3×3 (Snapping Technique × Movement Task) design with 20 trials for each treatment. For every trial, we recorded the task completion time and the number of times participants lifted a finger from the table (referred to as a *touch up*).

Participants were given training and the opportunity to practice each movement task (with no snapping) at the start of a session. For data analysis, we discarded the first 5 trials of each treatment of 20 trials to reduce learning effects.

Participants. Eighteen right-handed participants (2 female) between the ages of 21 and 40 (μ = 26.7) with normal color vision were recruited from our institution. Six participants had previously used a tabletop display, but only briefly during demos or to play games. Each participant received $10 for participating. Prior ethical approval was obtained from our university's behavioral research ethics board.

Apparatus. The user study was conducted on a SMART Table from SMART Technologies. The table had a 57.2cm × 42.9cm screen, with a resolution of 1024 × 768 pixels, and a 70Hz refresh rate. The application used for this user study was written in C# using the SMART Table SDK.

Hypotheses. We had two hypotheses: (1) Participants would perform faster with Oh Snap than with no snapping, and (2) participants would not perform slower with Oh Snap compared to traditional snapping. We were also interested in seeing how snapping technique would impact *touch ups*, the number of times users lift a finger from the table during a task.

Results for Phase 1. We performed a repeated measures ANOVA on the dependent variable trial time. A Bonferroni adjustment was applied to all pair-wise comparisons. Mean trial times for all snapping techniques across all movement types are shown in Table 2.

Table 2. Mean trial time (in seconds) for each of the snapping techniques across all movement types

Snapping Technique	Movement Type			
	All	1-D	2-D	Rotation
No snapping	2.507	2.301	3.517	1.703
Oh Snap	1.513	1.582	1.667	1.287
Traditional snapping	1.387	1.362	1.651	1.147

Performance. As shown in Table 2, the average trial time in seconds was 2.507 for no snapping, 1.513 for Oh Snap, and 1.387 for traditional snapping. There was a significant main effect of snapping technique ($F(2,34)=30.062$, $p<.0005$, $\eta^2=.639$) and movement type ($F(2,24)=27.529$, $p<.0005$, $\eta^2=.618$). There was a significant interaction effect of snapping technique × movement type ($F(4,68)=9.493$, $p<.007$).

We ran pair-wise comparisons between techniques and found that both Oh Snap ($p<.0005$) and traditional snapping ($p<.0005$) were faster than no snapping, but traditional snapping was not significantly faster than Oh Snap. We also ran pair-wise comparisons on movement types and found significant differences between all pairs: 1-D and 2-D ($p<.005$), 1-D and rotation ($p<.0005$), and 2-D and rotation ($p<.0005$).

To understand how snapping technique impacted performance for each movement type, we ran a repeated measures ANOVA for each of the three movement types.

There was a significant effect of snapping technique for rotation ($F(2,16)=17.634$, $p<.0005$, $\eta^2=.688$), 1-D translation ($F(2,16)=14.411$, $p<.0005$, $\eta^2=.643$), and 2-D translation ($F(2,16)=10.128$, $p<.001$, $\eta^2=.559$). Both Oh Snap and traditional snapping were significantly faster than no snapping for all movement types ($p<.0005$). Traditional snapping was significantly faster than Oh Snap only in the 1-D task ($p<.01$); however, it was only 14% faster whereas Oh Snap was 31% faster than no snapping in the 1-D task. A chart showing the mean trial times for each movement type grouped by snapping technique is in Fig. 6.

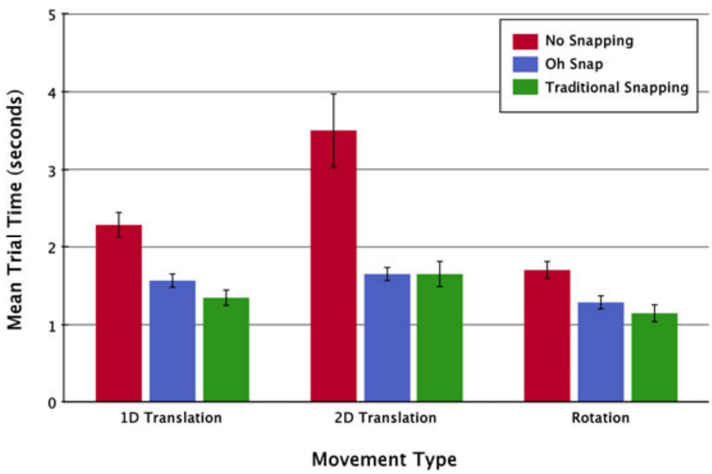


Fig. 6. Mean trial time by movement type grouped by snapping technique. Error bars represent standard error.

Number of touch-ups. We also conducted a 3×3 (Snapping Technique × Movement Task) repeated measures ANOVA for the average number of touch ups in a set of trials. There was a significant main effect of both snapping technique ($F(2,34)=21.427$, $p<.0005$, $\eta^2=.558$) and movement task ($F(2,34)=9.45$, $p<.001$, $\eta^2=.357$). Pair-wise comparisons showed that participants lifted their finger up significantly more in the no snapping condition ($\mu=1.438$) compared to either Oh Snap ($\mu=1.19$) or traditional snapping ($\mu=1.185$) with $p<.0005$ for both. There was no significant difference in the number of touch ups between Oh Snap and traditional snapping. There was a significant interaction effect for snapping technique × movement type ($F(4,68)=6.274$, $p<.023$).

Questionnaire data. Questionnaires were administered after each snapping technique to collect subjective ratings and receive comments using TLX-based Likert-style questions. Pair-wise comparisons using Wilcoxon Signed Ranks Tests for each of the questionnaire questions are presented in Table 3.

Participants ranked the three snapping techniques from best to worse in terms of preference. A Friedman test showed that technique significantly impacted rankings

$\chi^2_{(2,N=18)}$=10.371, p<.006). Both Oh Snap and traditional snapping had a mean rank of 1.67 while no snapping had a mean rank of 2.56 (lower is better).

Discussion of Phase 1. Both the Oh Snap and the traditional snapping techniques were significantly faster than no snapping, supporting hypothesis (1). On average, Oh Snap was 40% faster than no snapping. Because traditional snapping was not significantly faster than Oh Snap on average, hypothesis (2) is also supported. As reported in the results, although traditional snapping was significantly faster than Oh Snap in the 1-D tasks, it was only 14% faster, but Oh Snap was 31% faster than no snapping for such tasks. This small degradation suggests that Oh Snap is a reasonable alternative to traditional snapping for alignment tasks on touch interfaces, and it has the very important added benefit that users can place objects in close proximity to snap locations.

Table 3. Summary of significance differences between snapping techniques for each post-condition questionnaire question. **A**=Oh Snap, **B**=No snapping, and **C**=Traditional snapping. Non-significant comparisons are labeled *n.s.*

Significant differences between snapping techniques in subjective answers			
Question	**Snapping Technique**		
	A better than B	**C better than A**	**C better than B**
Mental demand	.013	*n.s.*	.003
Physical demand	.033	*n.s.*	*n.s.*
Task difficulty	.005	.021	.001
Success	.03	*n.s.*	.007
Hard work	.003	*n.s.*	.003
Fulfilled	.001	*n.s.*	.0005

Although this phase of the user study showed that Oh Snap performs well for alignment tasks, it did not investigate how Oh Snap might benefit proximity placement. To properly assess this feature, we conducted a second phase of the user study to compare several variations of the snap width and catch-up width parameters to find the best values for each.

5.2 Parameters for Snap Line Proximity – Phase 2 of the User Study

This phase investigated the variation of the Oh Snap parameters *snap width* and *catch-up width*, and measured their effect on user performance for positioning digital objects in close proximity with (but not aligned to) a snap line. We compared Oh Snap variants to the no snapping technique. We hoped to find a balance between a small ratio (Eq. 1) to reduce super pixel size, and minimizing the sum *catch-up width* + *snap width* so that it was not much larger than the average touch contact width.

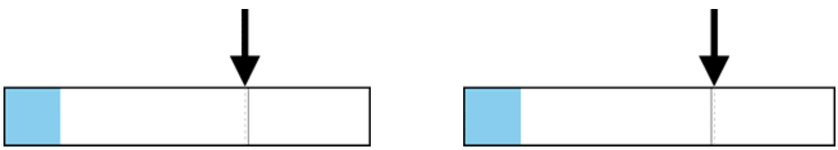
Apparatus and Participants. The apparatus and the participants were the same as in Phase 1. Participants started Phase 2 shortly after they completed Phase 1.

Task. The task was very similar to the 1-D alignment task from Phase 1. Participants were asked to translate a square so that it was close to, but not quite aligned with, a snap line. A dotted target line, with an arrow pointing to it, was placed 5 pixels before or 5 pixels after the Oh Snap line. Participants were asked to drag the square so that its right edge was aligned with the dotted target line. As in Phase 1, once the square was aligned with the target line, both the arrow and the square turned green to indicate successful alignment.

A screenshot of the start of each of the tasks (after the 'GO' button had been pressed) is shown in Fig. 7.

Interfaces. There were four sets of snapping parameters.

- no snapping
- a 4/3 ratio with 15px catch-up width and 5px snap width
- a 3/2 ratio with 20px catch-up width and 10px snap width
- a 4/3 ratio with 30px catch-up width and 10px snap width



(a) Target line is *before* snapping line. (b) Target line is *after* snapping line.

Fig. 7. Screenshots of each of the two tasks at the start of a trial in Phase 2

Traditional snapping was not used because the target lines were within the traditional snapping threshold area and thus impossible to reach with that technique.

We chose these parameter sets based primarily on our findings during pilot testing, when we found that ratios of less than 3/2 made it difficult to accomplish proximity tasks. We chose 40 pixels, or 2.24cm on the SMART Table, as the largest *catch-up width* + *snap width* to investigate if a width larger than the average touch contact width of 1.44cm [31] would yield positive results. A large value for *catch-up width* + *snap width* allows for tenacious snapping of an object's position while also maintaining imperceptibility, both of which are desirable. If that size gave good results, future Oh Snap implementations would have to take great care not to overlap the snap and catch-up regions of different snappable lines.

Experimental Design. The study was a partially counter-balanced, within subjects 4×2 (Snapping Parameter Set × Proximity Task) design with 20 trials for each treatment. For each trial, we recorded the task completion time. As in Phase 1, participants were given training and the opportunity to practice each movement task (with no snapping) at the start of a session. For the purposes of data analysis, we discarded the first 5 trials of each treatment of 20 trials to reduce learning effects.

Hypotheses. We had two hypotheses: (3) The Oh Snap conditions would be no slower than the no snapping condition, and (4) participants would perform fastest with the largest catch-up-size-to-snap-width-ratio and the largest snap-width. When the ratio is large (i.e. the catch-up size is 3 times larger than the snap width) the size of super pixels is decreased so it should be easier to perform fine-grained movement of a digital object that has been snapped. We expected that participants would perform better placing objects just before the snap line than just after it. Because objects could avoid being snapped when placing them before the snap line, this proximity task would be less difficult.

Results for Phase 2. We ran a repeated measures ANOVA on the dependent variable trial time. A Bonferroni adjustment was applied to all pair-wise comparisons. Mean trial times for all Oh Snap parameter sets across both proximity tasks are presented in Table 4.

Performance. There was a significant main effect of Oh Snap parameter set ($F(3,51)=12.702$, $p<.0005$, $\eta^2=.428$) and proximity task ($F(1,17)=6.626$, $p<.02$, $\eta^2=.28$). Mean trial times for all parameter sets are presented in Fig. 8.

Table 4. Mean trial time (seconds) for each of the Oh Snap parameter sets across both proximity types (just before and just after a snap line)

snap width	catch-up width	Ratio	Proximity Type		
			Both	Before	After
No snap	No snap	N/A	2.309	2.236	2.382
5px	15px	4/3	3.350	3.689	3.012
10px	20px	3/2	3.559	3.919	3.198
10px	30px	4/3	3.662	4.128	3.196

Pair-wise comparisons revealed that the no snapping condition was significantly faster than all parameter sets ($p<.0005$ for the first, $p<.002$ for the second, $p<.0005$ for the third). Participants did not complete trials significantly faster with any Oh Snap parameter set compared to any other. There was a significant interaction effect of parameter set × proximity task ($F(3,15)=3.884$, $p<.031$). Mean trial times for each proximity task type grouped by Oh Snap parameter set are shown in Fig. 8.

Surprisingly, the proximity task that required participants to position the square just before the snap line took significantly longer than positioning the square just after the line ($p<.02$). We anticipated that because the objects could avoid being snapped when placing them before the snap line, this proximity task would be less difficult and faster as a result. In sharp contrast, squares positioned after the snap line must be snapped and therefore we thought they would take longer.

Perhaps participants knew that snapping would occur in the 'after' task and learned how to adjust their movements to accommodate it. Conversely, in the 'before' task, participants may have worked slower so that they did not snap to the line. Participants may have occasionally overshot the target position, though not always, resulting in inconsistent behaviour preventing them from mastering the task.

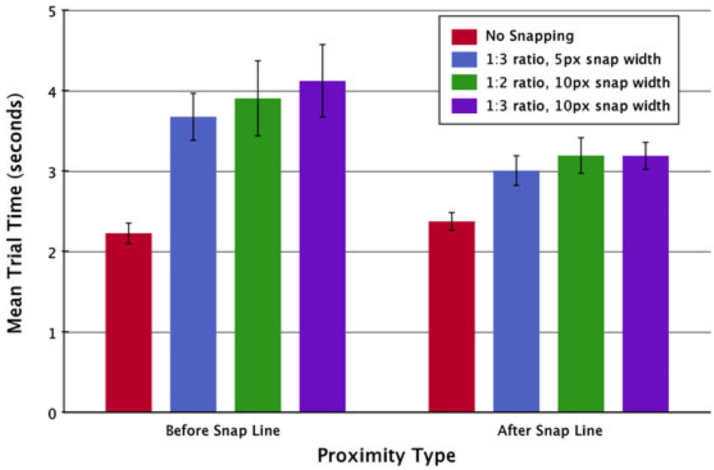


Fig. 9. Mean trial time by proximity type grouped by Oh Snap parameter set. Error bars represent standard error.

Number of touch ups. We conducted a 4 × 2 (Parameter Set × Proximity Task) repeated measures ANOVA for the number of touch ups in each treatment. There was a significant main effect of both parameter set ($F(3,51) = 6.79$, $p<.001$, $\eta^2=.285$) and proximity task ($F(1,17)=5.236$, $p<.035$, $\eta^2=.235$). Pair-wise comparisons of Oh Snap parameters revealed the no-snapping condition had significantly fewer touch ups than the second ($p<.002$) and fourth ($p<.008$) parameter sets, but not the third.

Discussion of Phase 2. Hypotheses (3) and (4) were not supported: the no snapping condition was faster than the three Oh Snap conditions and there were no significant differences in mean trial time between the parameter sets. While this means that there is a price, in terms of raw speed, for using Oh Snap, if snapping is desired the no-snapping condition is not a viable option. Oh Snap's real competitors are other snapping techniques. The advantage it has is that it permits placement of objects at any location, including right near a snap line. There is very little penalty for doing this compared to no snapping, and it cannot be done at all with other techniques.

There was no difference between the number of touch ups in the no snapping condition and in the condition using the third Oh Snap parameter set (3/2 ratio, 10px snap width). We recommend those parameters for future Oh Snap implementations.

6 Conclusions and Future Work

We introduced and evaluated a novel snapping technique supporting the alignment of digital objects on a touch interface. Oh Snap maintains the position of objects underneath the user's finger. It allows placement of objects close to snap lines, without having to toggle the snap capabilities off and on. This feature is especially important for collaborative environments. Oh Snap offers a mean 40% performance improvement over no snapping for 1-D, 2-D and rotation tasks, and it does not perform significantly worse than traditional snapping for those tasks on a touch table.

We investigated three parameter sets for Oh Snap in two proximity placement tasks. The no snapping condition performed better than the parameter sets, but the differences were not very large. Placing objects near a snap line is an infrequent task compared to snapping objects or placing objects in open space, both of which are tasks at which Oh Snap excels. We argue that, for direct touch surfaces, Oh Snap is superior to no snapping or traditional techniques because of its unique combination of attributes (Table 1). Furthermore, the lack of a significant difference between parameter sets for Oh Snap indicates that there is probably a wide range of suitable values, meaning that developers can choose values that best suit their users' needs.

6.1 Comparison to Nacenta et al.'s Prior Work

The Oh Snap velocity profile is similar to the "magnitude filtering" used by Nacenta et al. [20] to block selected degrees of freedom. In their approach velocity profiles are used to determine whether motion is the beginning of a constrained gesture or if it is "noise" in the current gesture. In contrast to this, Oh Snap employs the profile not just when movement is *initiated*, but *whenever* a snap line is encountered. Oh Snap uses the velocity profile *within* a chosen degree of freedom (1-D or 2-D translation, or rotation) to decide when to *stop*, whereas Nacenta et al. use it to *change* or *start* a chosen degree of freedom.

Nacenta et al. use symmetric constraint regions. Our regions are asymmetric; no snapping occurs approaching a snap line, only after it is passed. This makes it less likely a target just ahead of a snap line will be "missed." If a target just after a snap line is missed, "backing up" will not be handicapped because the asymmetry means that no snapping is applied in the reverse direction once the object becomes un-snapped.

6.2 Extensions to Oh Snap

Oh Snap works equally well on a traditional single-touch table or on a multi-touch table. If multi-touch is available, the number of touches could be used as a mode switch, which would allow seamless transitions between snapping and no-snapping. Future work for these and other snapping techniques will involve a new bimanual Oh Snap interaction that does not require explicit implementation of system-defined snap lines. An Oh Snap user will touch the interface with two fingers of the same hand and which define a temporary snap line to which they can align an object.

Acknowledgements. This research was funded in part by the Networks of Centres of Excellence of Canada through GRAND, the Graphics, Animation and New Media NCE, by the Natural Sciences and Engineering Council of Canada under the postgraduate scholarship program and the discovery grants program, by the University of British Columbia through a University Graduate Fellowship, and by an equipment donation from SMART Technologies. Defense Research and Development Canada supplemented the NSERC postgraduate scholarship. Additional equipment and other infrastructure necessary for the research were provided by funding from the

Canadian Foundation for Innovation and the British Columbia Knowledge Development Fund.

References

1. Ahlström, D., Hitz, M., Leitner, G.: An evaluation of sticky and force enhanced targets in multi target situations. In: Proc. NordiCHI 2006, pp. 58–67 (2006)
2. Albinsson, P.-A., Zhai, S.: High precision touch screen interaction. In: Proc. CHI 2003, pp. 105–112 (2003)
3. Aliakseyeu, D., Subramanian, S., Lucero, A., Gutwin, C.: Interacting with piles of artifacts on digital tables. In: Proc. AVI 2006, pp. 159–162 (2006)
4. Baudisch, P., Cutrell, E., Hinckley, K., Eversole, A.: Snap-and-go: helping users align objects without the modality of traditional snapping. In: Proc. CHI 2005, pp. 301–310 (2005)
5. Benko, H., Wilson, A.D., Baudisch, P.: Precise selection techniques for multi-touch screens. In: Proc. CHI 2006, pp. 1263–1272 (2006)
6. Bier, E.A.: Snap-dragging in three dimensions. In: Proc. I3D 1990, pp. 193–204 (1990)
7. Bier, E.A., Stone, M.C.: Snap-dragging. In: Proc. SIGGRAPH 1986, pp. 233–240 (1986)
8. Bjørneseth, F.B., Dunlop, M.D., Strand, J.P.: Dynamic positioning systems: usability and interaction styles. In: Proc. NordiCHI 2008, pp. 43–52 (2008)
9. Frisch, M., Dachselt, R.: Benefits of interactive display environments in the software development process. In: Proc. CHASE 2008, pp. 53–56 (2008)
10. Gleicher, M.: Image snapping. In: Proc. SIGGRAPH 1995, pp. 183–190 (1995)
11. Gutwin, C., Greenberg, S.: Design for individuals, design for groups: tradeoffs between power and workspace awareness. In: Proc. CSCW 1998, pp. 207–216 (1998)
12. Holz, C., Baudisch, P.: The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In: Proc. CHI 2010, pp. 581–590 (2010)
13. Khalilbeigi, M., Steimle, J., Muhlhauser, M.: Interaction techniques for hybrid piles of documents on interactive tabletops. In: Proc CHI EA 2010, pp. 3943–3948 (2010)
14. Kruger, R., Carpendale, S., Scott, S.D., Greenberg, S.: How people use orientation on tables: comprehension, coordination and communication. In: Proc. GROUP 2003, pp. 369–378 (2003)
15. Kruger, R., Carpendale, S., Scott, S.D., Tang, A.: Fluid integration of rotation and translation. In: Proc. CHI 2005, pp. 601–610 (2005)
16. Lecuyer, A.: Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback. Presence: Teleoper. Virtual Environ. 18(1), 39–53 (2009)
17. Liu, J., Pinelle, D., Sallam, S., Subramanian, S., Gutwin, C.: TNT: improved rotation and translation on digital tables. In: Proc. GI 2006, pp. 25–32 (2006)
18. Mandryk, R.L., Rodgers, M.E., Inkpen, K.M.: Sticky widgets: pseudo-haptic widget enhancements for multi-monitor displays. In: Proc. CHI 2005, pp. 1621–1624 (2005)
19. Micire, M., Desai, M., Courtemanche, A., Tsui, K.M., Yanco, H.A.: Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In: Proc. ITS 2009, pp. 41–48 (2009)
20. 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)

21. Nóbrega, R., Sabino, A., Rodrigues, A., Correia, N.: Flood emergency interaction and visualization system. In: Sebillo, M., Vitiello, G., Schaefer, G. (eds.) VISUAL 2008. LNCS, vol. 5188, pp. 68–79. Springer, Heidelberg (2008)
22. Olwal, A., Feiner, S.: Spatially aware handhelds for high-precision tangible interaction with large displays. In: Proc. TEI 2009, pp. 181–188 (2009)
23. Olwal, A., Feiner, S., Heyman, S.: Rubbing and tapping for precise and rapid selection on touch-screen displays. In: Proc. CHI 2008, pp. 295–304 (2008)
24. Raisamo, R., Raiha, K.-J.: A new direct manipulation technique for aligning objects in drawing programs. In: Proc. UIST 1996, pp. 157–164 (1996)
25. Reetz, A., Gutwin, C., Stach, T., Nacenta, M., Subramanian, S.: Superflick: a natural and efficient technique for long-distance object placement on digital tables. In: Proc. GI 2006, pp. 163–170 (2006)
26. Ringel Morris, M., Paepcke, A., Winograd, T., Stamberger, J.A.: TeamTag: exploring centralized versus replicated controls for co-located tabletop groupware. In: Proc. CHI 2006 pp. 1273–1282 (2006)
27. Siek, K.A., Rogers, Y., Connelly, K.H.: Fat finger worries: How older and younger users physically interact with PDAs. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 267–280. Springer, Heidelberg (2005)
28. Sutherland, I.E.: Sketchpad: a man-machine graphical communication system. In: Proc. AFIPS 1963, pp. 329–346 (1963)
29. Vogel, D., Baudisch, P.: Shift: A technique for operating pen-based interfaces using touch. In: Proc. CHI 2007, pp. 1063–1072 (2007)
30. Voida, S., Tobiasz, M., Stromer, J., Isenberg, P., Carpendale, S.: Getting practical with interactive tabletop displays: designing for dense data, "fat fingers," diverse interactions, and face-to-face collaboration. In: Proc. ITS 2009, pp. 109–116 (2009)
31. Wang, F., Ren, X.: Empirical evaluation for finger input properties in multi-touch interaction. In: Proc. CHI 2009, pp. 1063–1072 (2009)
32. Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., Benko, H.: Ripples: utilizing per-contact visualizations to improve user interaction with touch displays. In: Proc. UIST 2009, pp. 3–12 (2009)
33. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092 (2009)
34. Yang, L., Wang, F., Deng, H.: A survey on target selection technique for touch sensing devices. In: Proc. MVHI 2010, pp. 534–537 (2010)

The Link-Offset-Scale Mechanism for Improving the Usability of Touch Screen Displays on the Web

Willian Massami Watanabe, Renata Pontin de Mattos Fortes, and Maria da Graça Campos Pimentel

Computer Science Department, Institute of Mathematics and Computer Science at University of Sao Paulo, Sao Carlos, Brazil
{watinha,renata,mgp}@icmc.usp.br

Abstract. Touch-screen interfaces have become a widespread-input-device tendency for computer systems. In this context, many studies investigate how to improve general usability for touch-screen devices. These studies consider different interaction design features that improve the usability for touch sensitive surfaces, considering the low accuracy it presents, given obstacles such as the "fat finger problem", low-perception of pointing mechanisms, difficulties in the selection of small objects, among others. This work aims at presenting the link-offset-scale touch interaction mechanism for improving the usability for touch-screen devices. The link-offset-scale mechanism makes use of web-application-structure meta-data (identifying links) to provide feedback information about the selection of links in touch interfaces, while the surface is touched by the user. The link-offset-scale mechanism's primary goal is to reduce the number of errors that users commit while interacting with touch-screen devices in the Web.

Keywords: Touch-screen displays, pervasive computing, ubiquitous computing, usability in touch-screen devices, web usability.

1 Introduction

Touch-screen devices are becoming more and more frequent in personal computers, airports, banks and other environments [11,20]. Abowd et al. [1] describe touch screen devices as direct input mechanisms, that fit Pervasive Computing environments by providing natural interaction between the user and the computer. Touch screen interfaces are classified as "direct" input devices because of the integration between interface control and other information in a single surface [17]. The unification of input/output, control/feedback, hand gestures and eye movements suggests to researchers the idea that touch screen displays are intuition-oriented, particularly for non-frequent computer users [14].

There have been many studies which approach how to improve general usability for touch screen devices [8,9,11,12,18,20]. These studies consider different interaction design features that improve the usability for touch screen displays, considering the low accuracy it presents given the "fat finger problem", low-perception of pointing

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 356–372, 2011.

mechanisms, difficulties in the selection of small objects, among others [17]. It is worth noticing that these studies present design principles that should be addressed by interface designers.

On the other hand, the large capacity of Web for providing information accounts for multiple possibilities and opportunities for users. The development of high performance networks and ubiquitous devices allow users to retrieve information content from any location and in different scenarios or situations they might come across in their lives. Further to retrieving information, current technologies also enable users to contribute to the authoring of content on the Web by means of forums, blogs and wikis, in the so called “Web 2.0” [19]. However, a great amount of effort would be required to redesign all web applications now available in order to make them conformant with these design principles, and thus make the web more usable for touch screen displays.

In this context, this work aims at elaborating an automatic adaptation strategy for web application presentations in order to improve their usability when users are interacting with touch screen devices. Currently, the Web presents content in HTML, XML, JSON, among other formats. All these data are presented through web browsers that abstract the different platforms [4]. With the establishment of Ajax [6] and the evolution in the HTML5 profile specification [7], there has been a significant improvement in the interactivity of web applications. And, in this context, this work has the objective of making use of these newly evolved technologies (Ajax and HTML5) to maximize usability for touch screen displays, by implementing functionalities that automatically adapt the presentation of web applications to improve the general interactivity of websites for touch screen inputs.

The key functionality implemented in this paper is named link-offset-scale. The functionality provides feedback as soon as users have their finger over link elements in a webpage. The feedback presents a larger version of the link content (scaled) and is positioned in a different location to the link itself so as to be visible to the user, but not at the same position where his/her fingers are on the touch device. As the user stops touching the link, the link-offset-scale mechanism redirects him/her to the webpage the link referred to. This mechanism’s goal is to improve the usability of touch screen devices by increasing the accuracy and usability of touch interactions for web applications already available. In order to validate the described method, we conducted usability experiments on web applications, comparing accuracy and task realization timing metrics between users interacting with and without the resource.

This work is structured as follows: next section (Section 2) presents the state-of-art studies on touch interaction, which aim at improving accuracy of touch devices; Section 3 describes our design solution for improving usability of touch displays in the Web; Section 4 details a usability experiment carried out to measure the performance of our proposal; Section 5 presents a discussion about the results of the experiment; and finally, Section 6 presents some final remarks and future works.

2 Related Work

Even though touch screen displays are classified as intuition-oriented and easy to learn input devices, they present several limitations to be overcome by the user [17], which include:

1. Difficulty in the selection of small objects;
2. Problems related to the visual occlusion of elements and to imprecision and lack of calibration of the device;
3. And the non-existence of an analogue touch interaction for every possibility of mouse interaction.

These limitations are related to the accuracy of touch interactions with the display. Holz and Baudisch [8] explain that touching a small target represents an error prone activity that can be observed by having users acquire a small target repeatedly. The more inaccurate the device, the wider the spread of distribution of contact points with the surface. When considering the clicking activity of touch devices, the wider spread of contact points results in a higher risk of having users miss the target. Iwase and Murata's [10] experiments also support the conclusion that touch interfaces present a higher rate of errors than mouse pointing interactions.

A common explanation for the inaccuracy of touch devices relates to the fat finger problem [18]. The fat finger problem model considers that the width of the finger has a negative impact on the interaction. As users touch the surface, it is not clear which exact point should be considered as input by the computer, since the finger contact area is wider than that of a pixel-level pointing device such as the mouse. At the same time, the finger or arm used during the touch interaction occlude the target and prevents the target click point from providing visual feedback.

Holz and Baudisch [8] provide another model to describe the inaccuracy of touch devices. Their model of perceived input pointer considers that when a user tries to acquire a target, the actual pointing location considered by the computer is misaligned by a couple of millimeters from the target location where the finger is placed. Moreover, the fact that touch devices report the touch location at an offset increases the risk of missing the target.

In order to overcome the inaccuracy in touch displays, many studies focus on defining ways to design user interfaces which prevent users from missing their targets. Following this line of research, Huang et al. [9] investigated the factors that influenced the usability in LCD touch screens. The study reports that the factor which mostly affects usability according to their experiments is the touch field (size, location, space and density) of icons. Other studies that provide guidance on how to design an application to make it more usable for touch interfaces involve contact area widgets [12] that recommends the use of sliding widgets and contact area activation rather than using a single[1] pixel activation UI; and Apple's iOS Human Interfaces Guidelines[1] that recommend that targets be at least 44x44 pixels in area. It is worth noticing that these guidelines require applications to be designed by developers according to their recommendations. However, when specifically considering the Web, it would require a great amount of effort to re-build and re-deploy all websites and web applications, considering these guidelines and recommendations.

In this context, it becomes necessary to develop new strategies to reduce the number of errors when users are interacting with interfaces that do not follow usability guidelines for touch interfaces. These strategies adapt the interaction and presentation mechanisms of web applications in order to overcome the limitations of

---

[1] http://developer.apple.com/library/ios/#documentation/UserExperience/Conceptual/MobileHIG/UEBestPractices/UEBestPractices.html

low accuracy for touch displays. One clear example of an interaction adaptation strategy is the zooming based technique [3]. In order to acquire a target, the user can zoom in to enlarge the target until s/he can comfortably touch it, a similar approach to the pinch-to-zoom functionality present in Apple touch screen devices.

Albinsson and Zhai [2] argue that when zoomed to a subarea of the interface, the user loses the global context view of the application. Among techniques that improve accuracy for touch displays while not loosing the global context view of the application, it can be highlighted: Take-Off [16], Cross-lever, Precision- handle [2] and finger prints and position extraction [8].

It is worth noticing that these techniques make touch less direct, what might reduce the intuitiveness of the input (Take-Off, Cross-level and Precision-handle) [8] or require improvements in the hardware of touch devices (finger prints and position extraction).

Our approach, the link-offset-scale technique, tries to keep the direct-input behavior, while making no additional requirements towards the hardware of conventional touch screen devices and improving the accuracy of the interaction. The link-offset-scale functionality works specifically in the web context, using standard web technologies (Ajax and HTML5) to identify the parts of a webpage that require pointing assistance and adapt the webpage presentation in order to minimize the number of link acquirements during the interaction.

3 Adapting Link Presentation - the Link-Offset-Scale Functionality

Clicking in hyperlinks is the most frequent cause of navigation actions inside a web browser [5]. However, considering the growth in popularity of touch screen devices, there is an eager need to develop web solutions that fit the touch interaction scenario, improving the error rate for web application usage in touch devices.

In this context, we present the link-offset-scale technique, to improve the usability of touch interactions in the Web. The goal of the link-offset-scale functionality is to reduce the number of click errors for touch screen devices, by providing feedback information about the links that have been targeted by the user. The mechanism tries to solve the fat-finger problem [18] and the perceived input pointer [8] models by presenting feedback information about links available on webpages, before the user actually activates them.

The feedback information consists of the same HTML content that is placed inside each link of the webpage, but the information is scaled to 3 times the link's actual size (300% of the link's original size). As users touch an specific link (analogue to the mouse down event interaction in web applications) the feedback information about that specific link is presented. The feedback information presentation in the link-offset-scale technique is illustrated in Figure 1. As the feedback information is presented, an animation effect is also applied to the feedback information HTML element, in order to get the user's attention (fade in and movement effect).

Fig. 1. Outline of the link-offset-scale functionality: as the user positions his finger over a specific link (touching the display), the browser presents the content of that link in a separated magnified element next to the user's finger, so as to be in his field of view

The feedback information is placed on top of the target link's original position, with a right offset, if the link is positioned within 50% of the left side of the webpage, or a left offset, if the link is positioned within 50% of the right side of the webpage. The feedback information is positioned on top of the link in order not to have its view blocked by the user's fingers, hands or arms, during the touch interaction. The left or right offset are applied in order to try to avoid having feedback information blocking other links as users move their fingers through links while touching the screen, and considering that many interaction design patterns established for web applications[2] place different links on top of one another (Headerless Menu design pattern, for example) or besides one another (Main Navigation Menu design pattern, for example).

As the users take their fingers off the link (mouse up event), the feedback information is hidden. If the users stop touching the surface while they have their fingers over a link element, the link is activated and the webpage is redirected to the link's URL reference. This behavior is applied to all link elements presented in the webpage. The link-offset-scale mechanisms allow users to check the feedback information before actually acquiring the target link. If the feedback information does not match the users' target clicking link, they can adjust their finger positions in order to activate the correct link. We expect that this possibility of adjustment will reduce the number of errors made by users during the touch interaction.

The click event in a conventional touch device is interpreted as a link activation activity inside a web browser. In the link-offset-scale functionality, we break the click event into three other events: mouse down, mouse over and mouse up events. As the user dispatches a mouse down event, the browser starts listening to mouse over events. If a mouse over event to a link element is dispatched, the feedback information for that link is presented to the user. From there, the user can decide if s/he has targeted the correct link or if finger position adjustments are required. If the user has touched the link s/he meant to, s/he can stop touching the display in order to dispatch a mouse up event and be redirected to the selected URL link. If the user has

[2] http://www.welie.com/patterns/

targeted a wrong link s/he can try readjusting his/her finger to reach the expected link. The comparison between the conventional touch device experience and the link-offset-scale experience is illustrated in Figure 2. In case the users decide to interrupt the activity whilst their fingers are still touching the screen, they can put their fingers over a non-link HTML element (an element that does not provide feedback information), and stop touching the screen. This action will prevent any link activation and page redirection from taking place.

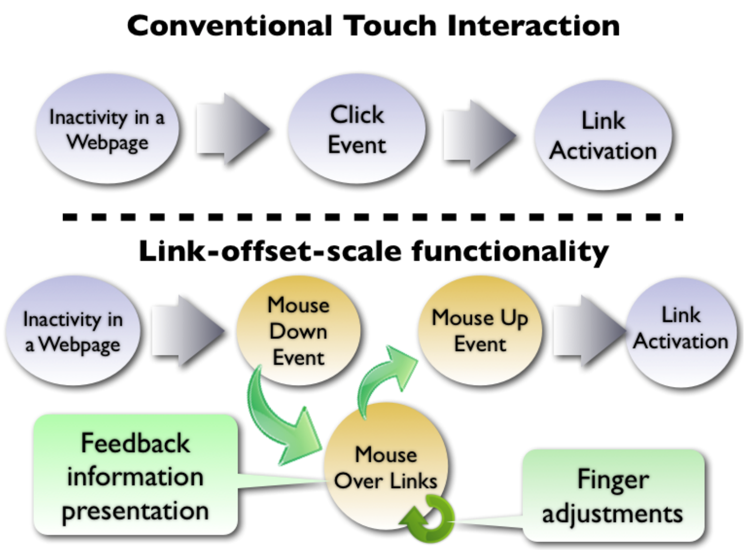


Fig. 2. These state charts illustrate the differences in behavior between the conventional touch experience and the link-offset-scale functionality touch experience

Our mechanism can be interpreted as a hybrid between the Take-Off approach [16] (since we provide targeting aids and break the acquisition task in targeting and then acquiring) and the zooming based technique [3] (as we provide a magnified version of the links information for users). However, it is worth noticing that our approach does not provide an offset pointer to the touch interaction, and thus, the link-offset-scale functionality does not make touching less direct than the Take-Off approach. And the zooming effect does not imply losing contextual information from the general view of the interface. Our mechanism can also be seen as an extension of the *Previewing Information Operation* [13], which proposes generalizing the presentation of a preview of the destination content [15] in order to facilitate the user's decision about when to follow the link, while keeping the original context – in our case, however, the same information presented in the source page is presented to help the user to make this decision.

As a drawback for the proposed approach, we expect the link-offset-scale functionality to increase the average pointing time required to acquire a link element. The possibility of increasing the average pointing time results in the link selection activity's increase in complexity in the link-offset-scale functionality, when compared to the conventional touch device behavior, as illustrated in Figure 2. However, in the Web context, a click error event does not have any impact in pointing time, but it increases the general task completion time, since the user has to wait for the new webpage to load in order to verify if the pressed link is the meant one, and, in case it is not, press the back button and try to acquire the right link again. These facts require empirical validation to assess how the general task time is affected by the use of the link-offset-scale functionality in order to verify if the increase in pointing time is actually compensated by the decrease in number of errors and loading time of the webpages.

Another action frequently carried out in web applications is scrolling. Web pages usually rely on scrolling to present all their content to users. In conventional touch interaction software, the use of swipe touch actions to scroll the webpage is presented as a well-accepted affordance option in interface design. The use of this type of affordance can be observed in Apple touch devices, Windows 7 drivers for touch screen devices, Google Android smartphones, among others. However, the swipe-to-scroll action prevents the link-offset-scale functionality from being used. If users try to move their fingers over the links in order to get the feedback information, the device will recognize a swipe action in the surface and scroll the screen, instead. So the swipe-to-scroll affordance cannot be applied concurrently to the link-offset-scale mechanism, at least when considering the swipe-to-scroll affordance conventional implementation (that recognizes the swipe gesture anywhere on the surface).

Taking that into account, in order to allow the proposed link-offset-scale functionality to work properly with the swipe-to-scroll affordance, we re-implemented the swipe-to-scroll gesture recognition feature in a way that it works solely on the extreme right side of the screen. More specifically, the swipe-to-scroll affordance will only be activated if the gesture is realized within 10% of the rightmost side of the surface. Any other gesture realized in the other 90% of the surface will work according to the link-offset-scale mechanism. Besides that particular change in the swipe gesture recognition, we also increased the scrolling sensitivity towards the swipe action in order to improve the scrolling time in the web application.

We have implemented a functional prototype for the link-offset-scale mechanism together with our custom implementation of the swipe-to-scroll affordance as a Mozilla Firefox addon, using the Addon-SDK[3] with other standard web technologies (HTML, JavaScript and CSS).

4 Evaluation

Our primary hypothesis is that the link-offset-scale functionality actually reduces the number of errors in acquiring links in a web application. In order to test our hypothesis, we conducted a usability task-oriented experiment that evaluated the number of errors users acquired when interacting with the system. We also monitored the general task time, so that we could analyze how the number of errors and the link-offset-scale mechanism pointing time impacted the task completion time.

The experiment consisted of comparing the number of errors and completion time for a variety of acquiring link tasks, with users interacting with conventional touch

[3] https://jetpack.mozillalabs.com/

screen interaction software and with the implementation of our link-offset-scale functionality.

Next sections describe experiment details about the participants (Section 4.1), the methodology (Section 4.2), and the results (Sections 4.3 and 4.4).

4.1 Participants

For our experiment we had a group of 14 participants. The participants' age ranged from 22 to 36 years old (12 men and 2 women). Test participants consisted of Computer Science graduate students from the ICMC-USP (Institute of Mathematical Sciences and Computer Science at University of Sao Paulo). All participants reported having used touch interfaces at least once a month in bank cash machines, where three of them used it up to twice a week and six used it more than twice a week.

4.2 Experiment Setup and Methodology

We used a Positivo UNION TOUCH 2500[4] as the touch device for the experiment. Even though the Positivo UNION TOUCH 2500 recognizes multi touch gestures, the tasks we carried out consisted only of single touch gestures. The touch-screen devices had Windows 7 installed. When executing a test with the conventional touch software, the computer ran with its standard configurations. When executing a test with the link-offset-scale mechanism, we disabled the standard gesture recognition feature of the operational system in order not to have the swipe-to-scroll operational system interfering with our own implementation of that affordance.

The experiment consisted of comparing the number of errors and completion time for a variety of acquiring link tasks, with users interacting with conventional touch screen interaction software and with our implementation of the link-offset-scale functionality.

Tasks consisted of simple acquiring link activities. The links were placed inside the main content of the webpage or inside UI components detailed as Interaction Design Patterns[5]. The links are also part of real-world websites available on the Web. Some tasks also required the use of scrolling functions in order to evaluate whether the modification of the swipe-to-scroll affordance affected the experiment results. Next, we describe the list of tasks, and the websites and Interaction Design Patterns associated to them:

- **Task 1:** textual element (paragraph HTML element) in http://gc.blog.br.
- **Task 2:** Tag Cloud Interaction design pattern in http://gc.blog.br.
- **Task 3:** Tag Cloud Interaction design pattern in http://gc.blog.br.
- **Task 4:** Headerless Menu Interaction design pattern in http://gc.blog.br.
- **Task 5:** Doormat Menu Interaction design pattern in http://www.wikipedia.org.
- **Task 6:** Accordion Menu Interaction design pattern in http://www.wikipedia.org.

---

[4] http://www.positivoinformatica.com.br/www/pessoal/tudo-em-um/tudo-em- um/positivo-union-touch-2500/visao-geral

[5] http://www.welie.com/patterns/

- **Task 7:** Directory Navigation Interaction design pattern in http://www.wikipedia.org.
- **Task 8:** Footer Site Navigation Interaction design pattern in http://www.wikipedia.org.
- **Task 9:** Faceted Navigation Interaction design pattern in http://www.icmc.usp.br.
- **Task 10:** Fly-out Menu Interaction design pattern in http://www.icmc.usp.br.
- **Task 11:** Fly-out Menu Interaction design pattern in http://www.icmc.usp.br.

It is worth noticing that previous knowledge about one of the websites used in the study (http://gc.blog.br, http://www.wikipedia.org or http://www.icmc.usp.br) could lead a participant to complete the task faster and thus generate a biased result of the experiment. Bearing that in mind, we pointed out and highlighted the target link for the participants before each task. Thus, the participants' task completion time was not affected by their familiarity with the task's web application and they were not required to look for the link within the webpage.

If a participant missed acquiring a link of the task, the system presented an error message and the participant was asked whether s/he would like to continue with the experiment or give up the task and go to the next one. It is worth noticing that the error message was instantly presented to users after any misplaced acquiring interaction. In real world tasks, however, there would be no error message and users would need to wait the next page to load to, then, identify whether their activity led to an unexpected result. Therefore, errors in the experiment had less impact on the task time than there would be in real world clicking tasks.

It is also worth noticing that errors were interpreted differently according to the system being tested. While testing the conventional touch interaction, all misplaced click events were classified as errors in the experiment. While testing the link-offset-scale functionality, all mouse up events that happened over an HTML link element were classified as errors, however mouse up events dispatched over non-link HTML elements were not classified as errors (standard action for canceling a link acquirement in the link-offset-scale mechanism).

The general procedure undergone by all participants consisted of the following steps:

1. Base explanation about the test goals of measuring the number of errors and task completion time.
2. Conventional touch mechanisms test session:
   a. Base explanation about how to use the conventional touch interaction mechanisms in webpage navigation.
   b. Start of the test session where the participant was asked to complete the 11 tasks using conventional touch mechanisms.
3. Link-offset-scale mechanisms test session:
   a. Base explanation about how to use the link-offset-scale interaction mechanisms in webpage navigation.
   b. Start of the test session where the participant was asked to complete the 11 tasks using the link-offset-scale touch mechanisms.

As it is possible to notice from these steps, all participants interacted with both touch interaction mechanisms (conventional touch functionality and the link-offset-scale functionality). However, the order they used each of the touch interaction mechanisms could generate biased experiment results. Using one mechanism could lead users to misinterpret the other. Therefore, we randomly changed the order of step 2 (Conventional touch mechanisms test session) and step 3 (Link-offset-scale mechanisms test session), so that half the participants executed step 2 before step 3, and vice versa.

In the experiment, the participants were not given time to become familiarized with the link-offset-scale functionality before starting the evaluation session. The main goal of the test was to verify if our approach reduced the number of errors in touch screen displays in web applications. If participants were allowed to familiarize themselves with the link-offset-scale mechanisms, the evaluation would be impacted in favor of the link-offset-scale approach, since most users were familiar with conventional touch interaction mechanisms and had never used the link-offset-scale mechanisms before.

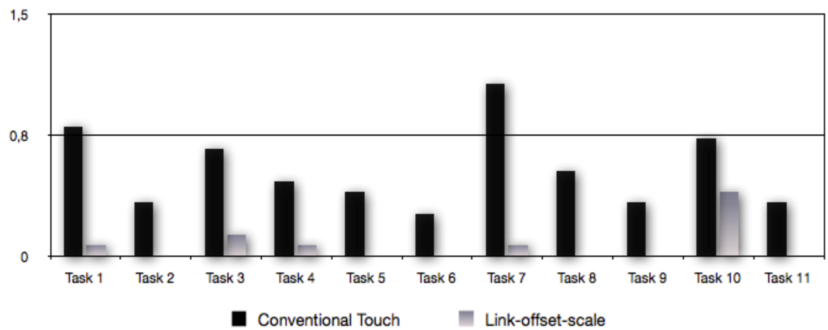


Fig. 3. Average number of errors for each task, according with the use of conventional touch screen presentation mechanisms and the enhanced presentation mechanisms with the link-offset-scale functionality

After having completed both test sessions, we asked the participants for comments and feedback on their experience with both touch interaction mechanisms.

4.3 Error Rate Analysis

In the experiment, it was verified that the number of errors was lower for all tasks when users were interacting with the link-offset-scale functionality. Figure 3 illustrates the average number of errors per task. In this figure, it can be observed that the average number of errors was lower for all tasks in our sample.

Tasks 1, 2, 4, 7, 8, 9 and 11 presented significantly less errors with the link-offset-scale mechanisms when compared to the conventional touch mechanisms, using a paired t-test analysis (with p-value inferior to 0.05). The average standard deviances and p-values data for all tasks are showed on Table 1.

Table 1. Average standard deviation for the number of errors for the conventional touch mechanism (CM as conventional touch average and CD as conventional touch standard deviation) and link-offset-scale interaction mechanisms (LM as link-offset-scale average and LD as link-offset-scale standard deviation) and the t-test p-value for analyzing whether the link-offset-scale mechanism presents less errors than the conventional one

Tasks	CM	CD	LM	LD	p-value
Task 1	0.8	1.16	0.06	0.26	0.01764
Task 2	0.33	0.74	0	0	0.04806
Task 3	0.66	1.13	0.13	0.53	0.06762
Task 4	0.46	0.65	0.06	0.26	0.02685
Task 5	0.4	1.15	0	0	0.0947
Task 6	0.26	0.61	0	0	0.05193
Task 7	1.06	1.16	0.06	0.26	0.002753
Task 8	0.53	0.93	0	0	0.02005
Task 9	0.33	0.49	0	0	0.009318
Task 10	0.73	1.12	0.4	0.85	0.1678
Task 11	0.33	1.08	0	0	0.1193

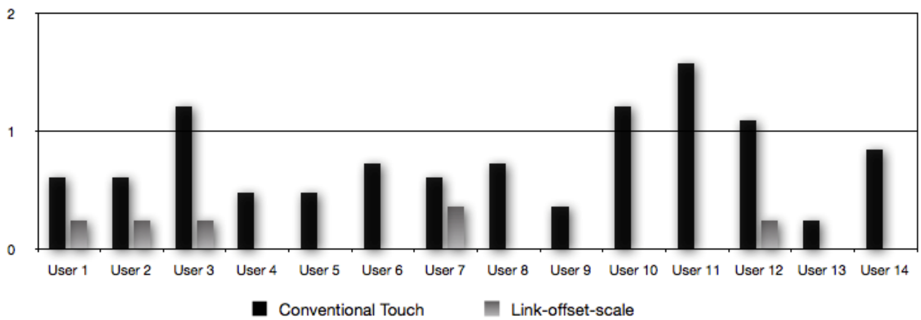


Fig. 4. Average number of errors per user, according to the use of conventional touch screen presentation mechanisms and the enhanced presentation mechanisms with the link-offset-scale functionality

Considering the average number of errors per user, we could also verify some positive results. All users performed better with the link-offset-scale mechanisms than with the conventional touch ones. Figure 4 presents the average number of errors per user.

4.4 Task Time Analysis

The experiment's results showed that, in the sample, the link-offset-scale functionality had a longer average task completion time for each task in comparison to the conventional touch interaction mechanisms. The only exception was for Task 7, which presented an inferior average completion time when users were interacting with the link-offset-scale mechanisms. The explanation for Task 7's different behavior is that it presented the highest difference in average number of errors when comparing the two approaches (1.066 average number of errors when interacting with the conventional touch interaction mechanisms against 0.066 when interacting with the link-offset-scale functionality). Errors during the test session increased the task completion time, since users were required to identify the error messages on the screen and try again. In Task 7, the increase in task completion time due to the number of errors during the interaction with the conventional touch interaction mechanisms compensated for the longer pointing time required by the link-offset-scale functionality.

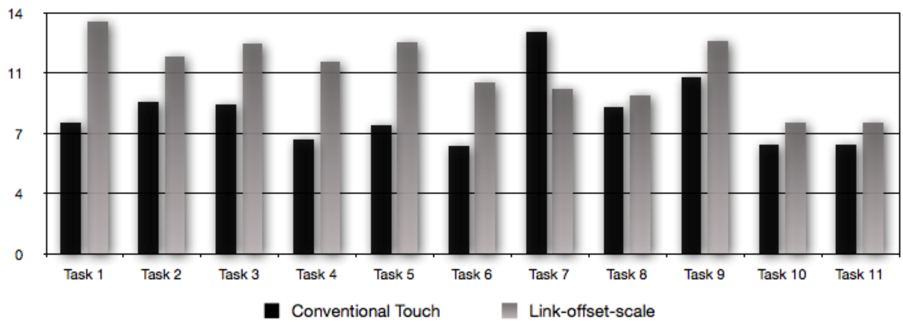


Fig. 5. Average time required for completing each task according to the use of conventional touch screen presentation mechanisms and the enhanced presentation mechanisms with the link-offset-scale functionality

It is important to notice that only two tasks presented significant differences between the task completion time for the conventional touch interaction mechanisms and the link-offset-scale mechanisms (Tasks 1 and 4), in a paired t-test with p-value under 0.05. Details about the t-test performed for each task are presented in Table 2.

Table 2. Average standard deviation for the task completing time for the conventional touch mechanisms (CM as conventional touch average and CD as conventional touch standard deviation) and link-offset-scale interaction mechanisms (LM as link-offset-scale average and LD as link-offset-scale standard deviation) and the t-test p-value for analyzing whether the link-offset-scale mechanism presents a greater task completing time than the conventional touch mechanism

Tasks	CM	CD	LM	LD	p-value
Task 1	7.6	4.12	13.53	11.40	0.03690
Task 2	8.86	6.94	11.46	6.79	0.09581
Task 3	8.86	6.13	12.2	15.21	0.2162
Task 4	6.66	4.81	11.13	5.22	0.0006582
Task 5	7.46	11.66	12.33	17.95	0.2009
Task 6	6.26	5.52	9.93	9.30	0.09794
Task 7	12.93	7.55	9.6	7.79	0.8727
Task 8	8.53	6.76	9.2	4.53	0.3786
Task 9	4.06	2.23	6.66	8.46	0.1087
Task 10	10.26	7.32	12.4	9.36	0.2588
Task 11	6.33	5.38	7.66	5.40	0.2458

Considering the average task completion time for each user, it can be verified that 9 users performed better with the conventional touch interaction mechanisms, while 5 users performed faster with the link-offset-scale mechanism. These data are illustrated in Figure 6.

5 Discussion

All subjects presented greater security (fewer errors) when interacting with the link-offset-scale functionality, as detailed in Figure 4. It was also verified that the average number of errors, depending on the task, support our main hypothesis that the link-offset-functionality reduces the number of misplaced link acquirements during the interaction, considering the test results in the sample. The feedback information presentation before the actual link selection action results in a better performance in terms of number of errors, whilst still keeping the direct input nature of touch devices.

The experiment also verified that the average task completion time is increased when users interact with the link-offset-scale functionality. That is a result of the increased pointing time required when interacting with the link-offset-scale mechanisms caused by the greater interaction complexity illustrated in Figure 2. However, the task completion time increment due to errors in the experiment is inferior to the task completion time increment observed in real world examples. Since in real-world examples errors require users to wait for the loading time of the website, recognizing that the website was wrongly selected, go back to the previous website and then restart the interaction. Further evaluation needs to be carried out in order to completely validate whether the general task completion time is longer when interacting with the link-offset-scale mechanism in the Web.

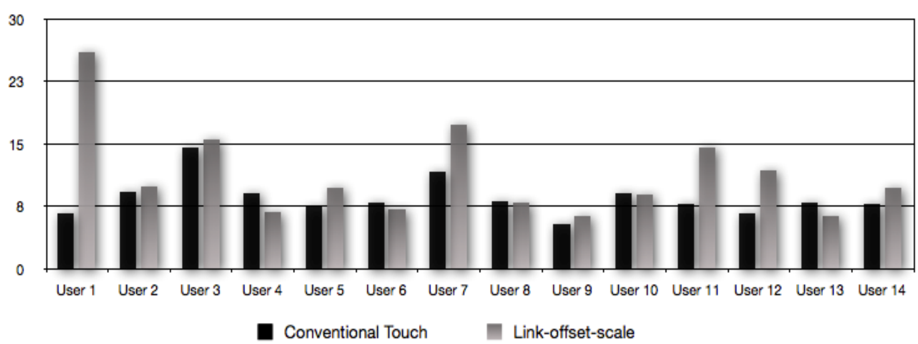


Fig. 6. Average time required for completing tasks for each user, according to the use of conventional touch screen presentation mechanisms and the enhanced presentation mechanisms with the link-offset-scale functionality

The link-offset-scale mechanism addresses the problem of low accuracy in touch screen displays, and its results are similar to other approaches (Take-Off, Cross-lever and Precision Handle) presented in the Related Work section (Section 2): there is a lower number of errors in the approach - however, a higher pointing time can be observed when interacting with the system. Nevertheless, the link-offset-scale approach maintains the direct input behavior of touch screen displays (different from the Take-Off, Cross-lever and Precision Handle strategies) and requires no additional hardware improvement (finger prints and position extraction) in order to reduce the number of errors in the touch interaction.

At the end of each testing session, we asked the participants to give us feedback on their interaction experience with the conventional touch interaction mechanisms and the link-offset-scale functionality. From the participants comments we have identified the following advantages and disadvantages that were not formally measured nor defined in the quantitative results section:

- **Differences in performance among tasks:** The high error rate in Task 1 was probably a result of it being the first task in the session - participants were more likely to make errors. Task 7 presented the highest error rate in the conventional touch mechanisms evaluation sessions as a result of the size of the target link of the task (located in the footer of the webpage). As the link-offset-scale mechanism provides a zoomed version of the text information placed inside the acquired link, the size of the link did not severely impact the rate of number of errors. Finally, task 10 and 11 consisted of a fly-out menu interaction design pattern and, therefore, required a more complex interaction in order to acquire the links placed in them. The fly-out menu requires the user to place the cursor (touch) in some parts of the menu. As the user places the cursor in these parts, additional parts of the menu are shown, and the user is able to click/acquire the target links in the task. We believe that the more complex nature of the interaction in fly-out menus resulted in a more severe impact on task 10, since it was the first task realized with this type of User Interface component.
- **Slow processing for the feedback information presentations:** some users stated that the feedback information for the links were not synchronous to the movement of their fingers over links. The feedback information was presented with an animation effect, in order to catch the users attention. However, the general functionality was implemented using JavaScript in the browser, therefore the processor was possibly overwhelmed by the number of animation effect executed with an interpreted programming language. In this context, we will consider removing the animation effect or reduce the animation time, so that feedback information is presented synchronously to the users fingers movement in the screen. It is also worth noticing that we run our experiments using a Mozilla Firefox Browser version 3.6.12, that does not contain hardware acceleration for visual effects in web applications. And hardware acceleration enabled browsers might reduce the slow observed by users while processing for the links feedback information presentation routines.
- **Advantages in interacting with smaller links:** most users reported that they were greatly benefited by the link-offset-scale functionality when trying to click on smaller links. These users reported that the link-offset-scale functionality is not required for all links, but should be applied for smaller links and widgets with links close to one another. Another advantage observed was that the magnifying functionality of the link-offset-scale mechanisms at a scale of 300% helped users to read and identify link feedback information.

- **Link-offset-scale selection strategy:** during the experiments, it was observed that users normally touched a non-link element of the screen and then dragged their finger to the target link they were aiming at, when interacting with the link-offset-scale mechanisms. This link selection strategy might have been used because, despite the feedback information presentation, the users were not aware of how to adjust their fingers when they were positioned over the wrong link. This was due to the fingers or hand blocking the links close to the one the user had his/her fingers on. Therefore, users started the mouse down event from a non-link element in order to draw a linear route that intercepted the target link and then they dragged their fingers over this linear route until the feedback information presented matched the one that described the target link. This behavior was probably one of the causes for the superior pointing time required by the link-offset-scale functionality, and solving it might improve the task completion time for the mechanism.
- **Lateral scroll difficulties:** it was observed that some users presented difficulties when interacting with the custom implementation of the swipe-to-scroll affordance, such as when clicking links that appeared under the swipe-to-scroll activation area. In order to activate links under the scroll area, the user would have to start his/her drag movement from outside the scroll area, then enter the scroll area whilst still touching the screen, and then remove his/her fingers off the surface when the target link feedback information was presented. This interaction pattern characterizes a non-intuitive activity and should be improved in next versions of the link-offset-scale functionality. Some users also reported that they lacked feedback information about when the custom implementation of the swipe-to-scroll functionality was active.
- **Scroll speed:** our custom implementation of the swipe-to-scroll affordance had an increased sensitivity, when compared to the standard implementation of the affordance. A few users commented that the increased sensitivity made scrolling activities faster, which might generate biased results for the task completion time data analysis. However, the scrolling did not affect all tasks and the scroll operation was only executed once during the task, regardless of the number of errors captured during the interaction. Therefore, we believe that the scrolling increased sensitivity did not generate biased results for the task completion time analysis.

It is worth noticing that all participants were computer experts (all graduate Computer Science students), where half of them can be classified as touch interaction experts, given the high frequency with which they use touch devices. In this context, it becomes difficult to evaluate the learning complexity of the proposed approach. As a future work, it would be interesting to run the experiment with a different user group, considering specifically computer inexperienced users, to verify whether the results and conclusions present different tendencies and evaluate whether the interaction complexity generated by the link-offset-scale functionality also impacts the usability of the touch interaction.

6 Final Remarks

This work aimed at presenting the link-offset-scale touch interaction mechanism for improving the usability for touch screen devices. The link-offset-scale mechanism makes use of web application structure meta-data (identifying links) to provide feedback information about the selection of links in touch interfaces. This feedback information is presented as the users move their fingers over the surface, while touching HTML link elements available on a webpage. As the feedback information is presented prior to the actual link activation, users are able to adjust their link selection in order to avoid errors during the interaction. The link-offset-scale mechanism's primary goal is to reduce the number of errors that users have while interacting with touch screen devices in the Web.

We conducted a usability experiment in order to evaluate whether the proposed approach reduces the number of errors during an interaction. The experiment results show that the number of errors was significantly lower when interacting with the link-offset-scale functionality. However, it was observed that the general task completion time was higher for the interactions that used the functionality.

As future works, we consider re-running the experiment with different user groups (computer inexperienced groups, more specifically), improving the implemented features in order to address the users' concerns about the slow processing feedback information presentation and scroll difficulties. We also plan on deploying multi-platform versions of the functionality in order to compare the users' usability performance when interacting with different touch devices.

Acknowledgments. The authors would like to thank CAPES, CNPq and FAPESP for the financial support.

References

1. Abowd, G.D., Mynatt, E.D., Rodden, T.: The human experience. IEEE Pervasive Computing 1, 48–57 (2002), IEEE Educational Activities Department, http://portal.acm.org/citation.cfm?id=612822.612832
2. Albinsson, P.A., Zhai, S.: High precision touch screen interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2003, pp. 105–112. ACM, New York (2003), http://doi.acm.org/10.1145/642611.642631
3. Bederson, B.B., Hollun, J.D.: Pad++: A zooming graphical interface for exploring alternate interface physics. In: Proceedings of User Interface and Software Technology, pp. 17–26. ACM, New York (1994), http://doi.acm.org/10.1145/192426.192435
4. Berners-Lee, T.: Information management: A proposal. W3C archive (1990), http://www.nic.funet.fi/index/FUNET/history/internet/w3c/proposal.html
5. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the World-Wide Web. Comput. Netw. ISDN Syst. 27, 1065–1073 (1995), http://portal.acm.org/citation.cfm?id=206477.206540
6. Garret, J.J.: Ajax: A new approach to web applications (2005), http://www.adaptivepath.com/ideas/essays/archives/000385.php
7. Hickson, I.: A vocabulary and associated APIs for HTML and XHTML. W3C Working Draft (October 2010), http://www.w3.org/TR/html5/

8. Holz, C., Baudisch, P.: The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 581–590. ACM, New York (2010), http://doi.acm.org/10.1145/1753326.1753413
9. Huang, H., Tsai, W.C., Lai, H.H.: Factors influencing the usability of icons in the LCD touch screens. In: Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction: Ambient Interaction, UAHCI 2007, pp. 878–887. Springer, Heidelberg (2007), http://portal.acm.org/citation.cfm?id=1763296.1763396
10. Iwase, H., Murata, A.: Empirical study on improvement of usability for touchpanel for elderly - comparison of usability between touch-panel and mouse. In: IEEE International Conference on Systems, Man and Cybernetics 2002, pp. 252–257. IEEE Computer Society, Los Alamitos (2002)
11. Malacria, S., Lecolinet, E., Guiard, Y.: Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 2615–2624. ACM, New York (2010), http://doi.acm.org/10.1145/1753326.1753724
12. Moscovich, T.: Contact area interaction with sliding widgets. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, UIST 2009, pp. 13–22. ACM, New York (2009), http://doi.acm.org/10.1145/1622176.1622181
13. Pimentel, M.d.G.C.: Evaluation of alternative operations for browsing hypertext. In: Conference on People and Computers IX, pp. 145–162. Cambridge University Press, Cambridge (1994), http://portal.acm.org/citation.cfm?id=211382.211395
14. Shneiderman, B.: Designing the user interface: strategies for effective human-computer interaction, p. 466. Addison-Wesley Longman Publishing Co., Inc., Boston (1986)
15. Shneiderman, B.: User interface design for the hyperties electronic encyclopedia (panel session). In: Proceedings of the ACM Conference on Hypertext 1987, pp. 189–194. ACM, New York (1987), http://doi.acm.org/10.1145/317426.317441
16. Shneidermnan, B.: Touch screens now offer compelling uses. IEEE Softw. 8, 93–94 107 (1991), http://dx.doi.org/10.1109/52.73754
17. Tsai, W.C., Lee, C.F.: A study on the icon feedback types of small touch screen for the elderly. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 422–431. Springer, Heidelberg (2009), http://dx.doi.org/10.1007/978-3-642-02710-9_46
18. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 657–666. ACM, New York (2007), http://doi.acm.org/10.1145/1240624.1240727
19. Watanabe, W.M., Neto, D.F., Bittar, T.J., Fortes, R.P.M.: WCAG conformance approach based on model-driven development and WebML. In: Proceedings of the 28th ACM International Conference on Design of Communication, SIGDOC 2010, pp. 167–174. ACM, New York (2010), http://doi.acm.org/10.1145/1878450.1878479
20. Wu, M., Shen, C., Ryall, K., Forlines, C., Balakrishnan, R.: Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In: Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 185–192. IEEE Computer Society, Washington, DC, USA (2006), http://portal.acm.org/citation.cfm?id=1109723.1110635

The Effects of Personal Displays and Transfer Techniques on Collaboration Strategies in Multi-touch Based Multi-Display Environments

Stefan Bachl[1], Martin Tomitsch[2], Karin Kappel[1], and Thomas Grechenig[1]

[1] Research Group for Industrial Software (INSO), Vienna University of Technology, Wiedner Hauptstr. 76/2/2, 1040 Wien, Austria
{stefan.bachl,karin.kappel,thomas.grechenig}@inso.tuwien.ac.at
[2] Design Lab – Faculty of Architecture, Design and Planning, The University of Sydney, 148 City Road, NSW 2006, Sydney, Australia
martin.tomitsch@sydney.edu.au

Abstract. Multi-touch tabletop systems promise to enhance collaboration in multi-display (MDE) environments. However, little is known about the effects of combining shared multi-touch tabletops with multi-touch tablet computers (tablets) as the collaborators' personal displays. In this paper we present the implementation of a MDE with multi-touch input on both shared and personal displays and its evaluation regarding task performance, user preferences and collaboration strategies compared to a standard multi-touch tabletop setting. Eight participant pairs had to solve a collaborative sorting task using three different transfer techniques. Based on an analysis of video recordings, log files and user feedback we identified task solving and collaboration strategies. The use of tablets enabled participants to follow diverse strategies and participants preferred the collaboration using tablets, while overall task performance and the amount of close collaboration were higher without the tablets.

Keywords: Co-located collaboration, interactive surfaces, multi-display environment, multi-touch, tabletop interaction.

1 Introduction

In the last few years multi-touch surfaces have gained a lot of attention, not only because of the increasing presence in consumer products and advances in finger tracking technologies but also because of benefits that come with this form of interaction. Existing challenges regarding the interaction design of multi-touch interfaces are balanced by advantages that come with this technology such as intuitive, direct interaction [3, 12], gestures [33] and the possibility of concurrent co-located collaboration. Multi-touch surfaces promise to enhance collaboration and to allow creating more engaging, flexible and easy to use interfaces to support the activities and workflows of teams working together [6, 9].

Additionally, portable devices are becoming more flexible and easy to use. Mobile phones and tablet computers continue to increase in functionality. Paired with instant

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 373–390, 2011.

on ability, portable devices are slowly finding their way into collaborative multi-display environments [2, 19]. For instance, a collaborative environment can support the integration of mobile devices into a shared network [25], the transfer and sharing of documents between devices to facilitate collaboration, or other more complex collaborative workspace scenarios [4, 20, 26, 27]. Consequently, synchronization and transfer techniques for sharing documents between devices become more important, whether data is exchanged locally or via the cloud. Emerging technologies even encourage approaches for transferring documents through digital representations projected onto users' hands while moving through a room [32].

Most collaborative environments have one thing in common: They are based on a shared space where collaborators can interact and communicate and coordinate their tasks. In digitally supported collaborative environments, this shared space is often represented by multi-touch tabletop systems. There is a large body of research available that investigated interactions with and around multi-touch tabletops and their effects on collaboration and task performance [8, 10, 15, 23, 24]. However, little information is available regarding techniques for sharing and transferring digital content between several multi-touch surfaces and the role of personal devices in digitally supported collaborative environments.

In this study we therefore investigate a multi-display environment consisting of a shared tabletop display and personal devices with multi-touch support on all devices. More specifically, we investigated the effects of personal displays and transfer technique on performance, awareness and collaboration strategies. For this purpose we propose three unidirectional transfer techniques for sending documents from the shared surface to personal displays, represented by tablet computers.

2 Related Work

Related work in the area of collaborative tabletop environments can be separated into single- and multi-display environments. Single-display environments (SDE) include all setups with a shared display for collaboration and either no additional input or one additional input device for each participant. Examples for input devices are computer mice or pen-based graphical tablet computers. Multi-display environments (MDE) consist of one or multiple shared displays for collaboration and include additional input devices with displays for each collaborator. Examples for input devices with displays are mobile phones, laptop computers, pen-based tablet computers (tablets), or as in our study (multi-)touch tablets.

2.1 Single-Display Environments

Hornecker et al. [9] compare a multi-touch tabletop setup where collaborators can interact with their fingers to a setup where instead each participant controls the interface with a computer mouse. Measuring the awareness of participants about collaborators' activities and their overall shared understanding of behavior, they conclude that the multi-touch tabletop system results in higher awareness and consequently to better task performance and results. Müller-Tomfelde and Schremmer [14] compare multi-touch with mouse interaction on a shared surface by allowing individual participants to choose their preferred input method. Although the input does not have an impact on performance, the authors state that participants are likely to stick to their initial input preferences. Isenberg et al. [10] identify different collaboration styles of participants working on a collaborative problem-solving task, analyze their strategies and conclude with design implications. Rogers and Lindley [22] compare a horizontal to a vertical display with pen-based input on both. They conclude that the vertical setup is better for communal and audience-based viewing and annotating, whereas the horizontal setup is superior for collaborative activities.

2.2 Multi-Display Environments

Wallace et al. [31] use a collaborative setup with a vertical display (acting as output only) to compare input with multiple mice to laptop computers. They conclude that collaborators using computer mice have a higher awareness of their partners' activities. Collaborators break off to individual tasks rather than monitor the shared activities when working with a laptop computer. Hawkey et al. [8] use a pen-based digital whiteboard system in combination with pen-based tablet computers to investigate the effects of distance between collaborators and the digital whiteboard on interaction and awareness. They conclude that the collaboration intensifies when users are near to the whiteboard system and also close to each other. Shen et al. [25] examine an interactive multi-touch tabletop for face-to-face collaboration where users can easily share their content from personal devices such as laptop computers. They identify three collaboration areas: The private area is not visible or accessible for others, the personal area is visible, but not accessible for others and the public (or shared) area is visible and accessible for collaborators. Nacenta et al. [17] use a pen-based horizontal surface with stationary pen-based tablets to investigate different transfer techniques between both devices together. The transfer techniques are tested with different distances between the stationary tablet and the horizontal surface. The authors describe design directions and recommendations for the use of individual transfer techniques. Tan et al. [28] present a collaborative job-shop scheduling task for MDEs with mouse-based interaction only. Nacenta et al. [16] present factors in MDEs and cross-display object movement techniques.

3 Supporting Transfer Techniques in Multi-Display Environments

We propose three one-way transfer techniques for sharing digital documents between a shared and personal multi-touch displays in a collaborative MDE. We developed the three techniques together with the interfaces for both shared and personal displays.

3.1 Shared Display

Each document is represented by a virtual representation on the shared display showing the entire content of the document. Users can resize documents within a predefined scale range and move them across the surface. The system allows users to group documents regardless of size (Figure 1) by moving two documents' side edges close to each other and releasing the documents. Grouped documents can be moved as a unit and individual documents can still be resized. To separate grouped documents users have to touch two adjacent documents, start dragging them apart from each other and release them. For instance, this grouping interaction can support sorting of a large number of individual documents.

Fig. 1. Documents on the shared display are represented as single-page representations with color-coding indicating documents that are currently opened by any of the collaborators on their tablets (1-4). Documents can be attached to each other along their vertical edges and single documents remain resizable (5).

3.2 Personal Display

The personal display features a document tray at the bottom of the screen where small representations of documents that are currently opened on the tablet are displayed. Touching a document's representation opens the respective contents on the document view area and highlights the representation in the tray. Each document's representation contains a small circular button at the top right corner to close the document. The brightness of the document view area is reduced to provide comfortable reading conditions. A document opened on a tablet also keeps its representation visible on the *shared* surface. We use color-coding for identifying users in the shared environment. When a user opens a document on their tablet, a border with the corresponding color is displayed around the document's representation on the shared display. The same color-coding is used on the tablets to support the mapping between tablet and user (Figure 1). If two or more users open a document at the same time, multiple borders with the respective colors are added to the document's representation on the shared display (Figure 1). This visualization helps users to easily identify and track opened documents. For simplicity reasons users can open only up to five documents at the same time on each tablet. When users try to open more than five documents at a time, an according message is displayed in form of a pop-up on the tablet. Users can resize a document's representation on the shared screen and scroll longer documents through simple touch interaction on the tablet. Participants are not able to change the order of documents or attach documents to each other on the tablets.

3.3 Transfer Techniques

Based on our review of related work we identified five techniques for supporting document transfer in MDEs through continuous touch interaction. Mouse and pen input were disregarded to avoid any negative effects of changing the mode of interaction between devices and since their differences to multi-touch were already examined by others [3, 9, 14, 18]. For further selection we evaluated the five identified techniques in a preliminary user study. Two techniques, a variation of pick-and-drop [21] and a three-finger multi-touch swipe gesture, were dropped due to limitations identified in this preliminary study. More specifically, we observed that in our context multi-touch gestures did not have any benefits for document transfer and were less accepted by users.

The final transfer techniques described below represent three distinct investigations into the design space allowing users to transfer documents from the shared display onto their personal displays through directly selecting the document on the shared display (*buttons*), moving the document onto a virtual representation of the personal display (*containers*), or selecting the document on the personal display (*lenses*). The proposed system only supports unidirectional transfer of documents from the shared display to the users' personal displays but not vice versa (Figure 2).

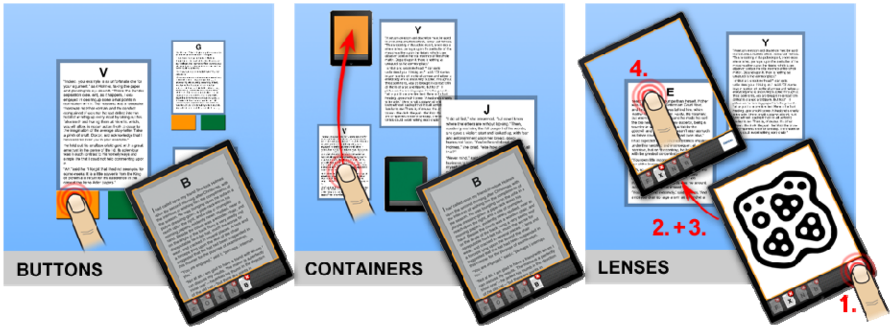


Fig. 2. Schematic representations showing the three different transfer techniques. Users can touch buttons to transfer an individual document (left), drag-and-drop documents or groups of documents into containers (middle) or use the tablets as lenses by activating the lens mode (1), selecting an area showing the desired document(s) (2), transferring the area onto the tablet (3), and selecting document(s) to open on the tablet (4).

Buttons. To support the *buttons* transfer technique, each document on the shared display features one button per user attached to an edge of the document's representation. The buttons are associated with users through color mapping. Touching a button (Figure 2) transfers the document onto the corresponding tablet. The *buttons* technique enables users to send documents to their collaborators.

Containers. In the *containers* transfer technique a visual element on the shared surface represents a user's tablet. The number of visual elements corresponds to the number of tablets connected to the system. We use an icon representing a tablet as visual element to convey its meaning as container mapped to the user's tablet. Again, we use color for mapping containers with tablets. Users can move the containers across the screen to allow positioning for convenient access. Containers are not resizable to avoid confusion with the document elements and always remain on top to avoid occlusion. Employing the drag-and-drop metaphor, users can drag a single document or groups of attached documents and drop them into one of the containers

(Figure 2), which consequently opens them on the corresponding tablet. Documents remain on the shared display and jump back to their initial position of the drag gesture after the transfer has been completed. Like the *buttons* technique, the *containers* technique also enables users to send documents to their collaborators.

Lenses. In the *lenses* transfer technique tablets act as lenses and can capture parts of the content of the shared display. We use fiduciary markers for detecting the positions of the tablets (see section 4.5). The interface of the tablets features an additional button in the bottom right corner to activate the lens mode (Figure 3). While holding the button users can move the tablet above the shared surface. A translucent representation of the tablet is displayed on the shared surface to support this interaction and provide feedback. When releasing the button, the content of the tablet's viewing area changes to the exact view of the shared display that is displayed beneath the current position of the tablet (Figure 2 and Figure 3). Participants can open documents on their tablet by simply touching their representations captured through the lens. When touching the button in the bottom right corner again or tilting the tablet while in lens mode, users return to their previously opened document.

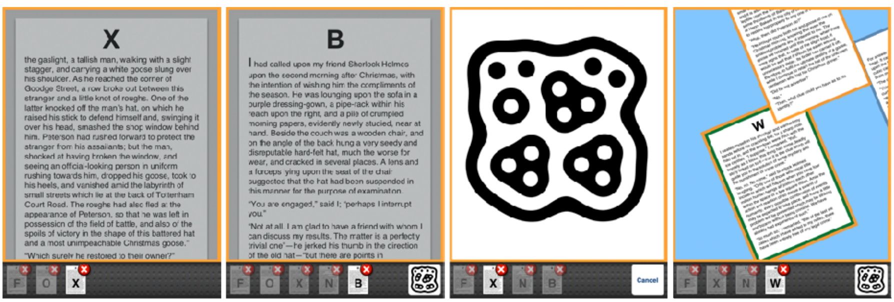


Fig. 3. Screenshots of the software running on the personal displays. Participants could read documents (first image) with reduced brightness. For the *lenses* transfer technique, an additional button was displayed in the bottom right corner (second image). When pressing the button a fiduciary marker was shown (third image). After releasing the contents of the shared surface were displayed (last image).

4 Evaluation Study

To investigate the proposed transfer techniques we conducted a task-based study with eight participant pairs. We used a within-subject design with four conditions in total, three being the different transfer techniques, and one allowing participants to only use the shared display to solve the given task. Conditions were counterbalanced to eliminate learning effects.

4.1 Task

We prepared a task, which required participant pairs to collaboratively sort ten pages of a book chapter. Ten single-page documents were randomly placed (all reduced to minimum size) on the shared display at the start of each condition. The goal was to bring the documents into the correct order by grouping pages together. Four chapters of "The Adventures of Sherlock Holmes" by Sir Arthur Conan Doyle were retrieved from Project Gutenberg [5]. The length of each chapter was reduced to about 2,100 words and separated into ten single pages. Some pages ended with full stops, others were separated in the middle of a sentence. The text also contained dialogs that helped participants to identify pages that belonged together. The first paragraph on the first page contained a large-sized capital letter to provide a starting point for solving the task. The text on the last page did not fill the whole page to help identifying the last page. These clues were given intentionally to enable participants to solve the task in a reasonable amount of time, as they had to complete the task for each condition. Participant pairs were presented with a different chapter from the book for each condition. At the top of each document, a randomized letter between A and Z in a large font was placed for identification. The letter helped participants to identify documents, but also to support communication (e.g. "*I think 'E' goes before 'F'*"). The text size in the PDF-documents was set to 25 points to make it readable on both shared and personal displays, although on the shared display legibility depended on the scale factor applied to a document.

4.2 Participants

For our study we recruited 16 participants (9 male, 7 female), aged between 19 and 47 years (mean: 25.7; SD: 7.07), and grouped them into eight pairs to examine the "collaborative coupling style" between participant pairs [29, 30]. 13 of the 16 participants were students from the University of Sydney. In six of the eight pairs participants already knew their partner before taking part in the experiment. Participants were partners, co-workers or friends. Participants were screened before the experiment in order to make sure that their English skills were sufficient to solve the given task. Six participants were native English speakers. On a scale between "fluent", "good" and "fair" eight participants rated their English skills as "fluent" and only two participants as "fair". None of the participants had read "Sherlock Holmes" within the last 5 years. Two of the 16 participants were left-handed. Participants were also tested for color-blindness. Three of the 16 participants did not have experiences with touch interfaces, five participants had used large multi-touch displays before.

4.3 Procedure

Before beginning with the experiment participants gave their informed consent to the study. Participants were asked to sit or stand in front of the tabletop next to each other. A within-subject design was chosen: Each participant pair was asked to complete the task consecutively in all four conditions (with different book chapters). The first condition (*without*) only involved the shared display, the other three included the use of the tablet computers as personal displays. In the conditions with personal displays participants were able to transfer documents from the tabletop onto their tablets. Each condition supported one of the three specific techniques for transferring documents (Figure 4).

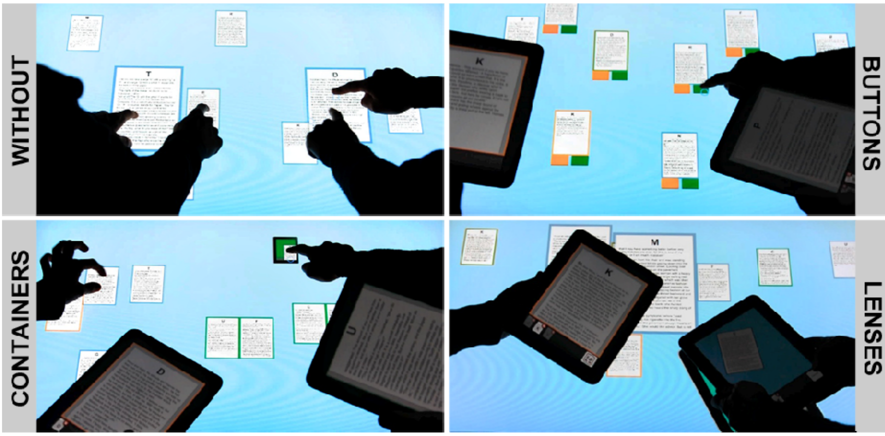


Fig. 4. Frames of a video showing the four conditions of our study: Working without tablets (top left), using buttons to transfer documents to the tablets (top right), drag-and-dropping documents into containers (bottom left) and using the tablets as lenses displaying parts of the contents of the shared surface (bottom right)

To eliminate potential learning effects within the groups, the order of the conditions was counterbalanced using Latin Squares. A short tutorial was given to the participants prior to each condition. During the tutorial participants had the opportunity to briefly interact with the system to make sure they understood the interface and transfer technique (in the conditions that involved a tablet). After the introduction the system was restarted and set up for the experiment. This procedure ensured that all participants understood the task and the transfer techniques. Except for a few questions during the experiments, additional help or assistance was not required. The task completion time per condition was about 11.5 minutes on average. Participants had the option to take a break after the first two conditions. After completing all four conditions, participants were asked to complete a questionnaire. Additionally, a short interview consisting of open questions with both participants together was conducted to gain deeper insights into their personal preferences and observations.

4.4 System Implementation

The system used for the evaluation consisted of two parts, a software running on the main computer controlling the shared surface and a separate piece of software for controlling the personal tablet computers. The system allowed documents displayed on the shared display to be transferred onto the tablets using the transfer techniques described above (Figure 4).

The application running on the main computer controlled the shared display and communicated with the tablet computers. The Java-based multi-touch framework MT4J [13] was used for implementation. The main application processed messages received from the tracking software. The communication between the tracking software and our application was necessary to identify the positions of the users' fingers touching the surface and also the positions of the fiduciary markers in order to identify the positions of the tablets used in the *lenses* transfer technique. The rotation of documents on the shared display was disabled, as it would have added complexity for the action of grouping documents.

The application running on the tablet computer was implemented with the Cocoa Touch framework (native iOS development based on Objective-C) since first generation Apple iPads are used for the underlying hardware platform. To enable fast switching of variables, experiment settings (e.g. *condition* or *participant identification*) were included in the tablet's global settings outside of the application. The tablet computers were connected with the main application via a wireless network connection to communicate with each other. The PDF documents were redundantly stored on the main computer and both tablets. To minimize latency just the corresponding identification numbers to open the documents were synchronized instead of transferring the document files themselves.

4.5 Experimental Setup

The study was conducted in one of our faculty's lab spaces. A custom-made horizontal multi-touch display built into a table was used as the shared display. The 42-inch display with a resolution of 1600x900 pixels was augmented with Laser Light Plane Illumination (LLP) to track the touches of participants' fingers. Two cameras were used to track interactions: One infrared camera was connected to a computer running Community Core Vision (CCV) [1], tracking the touches of the participants' fingers. A second visible light camera was connected to another computer running ReacTIVision [11] to track the fiduciary markers in the *lenses* condition. Even though ReacTIVision is capable of tracking both fingers and fiduciary markers, we chose to use CCV for tracking finger interactions because of its higher accuracy. The main application for the shared display was running on the second computer. All computers including the two tablets were connected via a wireless network connection. No performance issues emerged during the experiments.

Two first generation Apple iPads were used as the personal displays. Their 9.7-inch capacitive multi-touch display featured a resolution of 1024x768 pixels. Compared to the shared display, the iPads were slightly superior in regards to accuracy and display resolution.

A video camera was mounted at the side of the tabletop to record interactions of the participants on the shared display. For privacy reasons the faces of the participants were not captured. In all four conditions, chairs were provided for sitting at the table, but participants were explicitly told that they could decide on their own whether they preferred to stand or sit while completing the task. The cameras for tracking the interactions were mounted at one side of the table. Therefore, participants could only work on the opposite side. The tabletop featured additional horizontal space on its shorter sides, which participants could use to put down the tablets.

4.6 Data Collection and Analysis

To ensure the quality of the data collected during the experiment, we used several sources for recording data. The video camera recorded interactions on the shared display. The main application controlling the shared display and the two applications running on the tablet computers logged all user interactions and stored them in a database. Additionally, we used data retrieved from the questionnaires, interviews and logging sheets for our analysis.

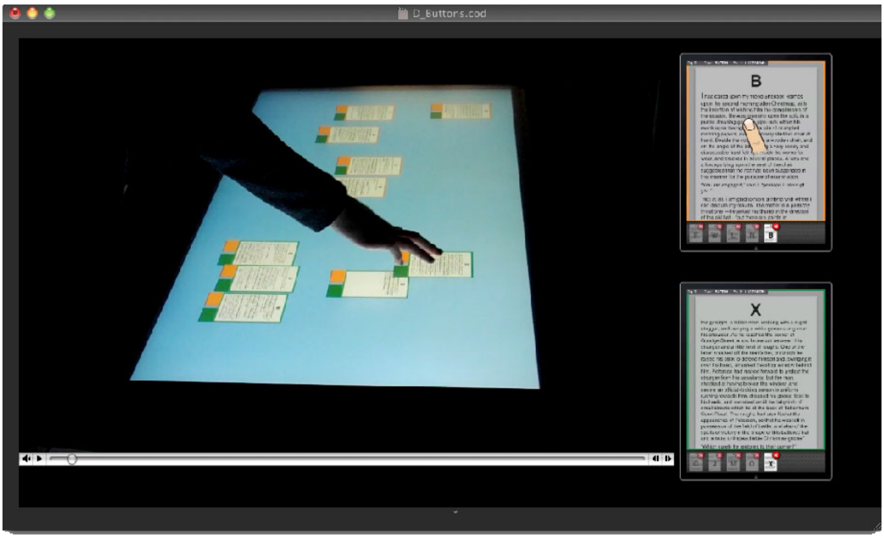


Fig. 5. A screenshot showing a reduced view of the video coding tool VCode [7]. Video recordings of interactions during the experiment on the shared surface (left) were combined with log file-based visual simulations of interactions on the two tablets (right).

For the video analysis we used VCode [7] and adapted the controls of the video playback with custom keyboard shortcuts for better control. Our video footage did not contain interactions on the tablets. Therefore we modified the original application running on the tablets to simulate (play back) the interactions from the logs in our database. We created screen recordings of the simulations of both participants' tablets and included them into VCode (Figure 5). This helped us to fully understand what was going on during the experiments, as VCode allows operating multiple videos simultaneously with only one playback controller.

To analyze participants' awareness of each other's activities while working on the task, we initially attempted to use the awareness indices identified by Hornecker et al. [9]. However, these indices were not designed for a multi-display environment. Therefore interactions with the personal displays between participants were not incorporated. For instance, they did not comprise situations where participants used their tablet to show a document to their partner without communicating. We therefore analyzed the video recordings regarding interaction patterns and task solving strategies used by the participants to solve the given task. For each participant group and condition we rated various factors including the closeness of collaboration, the amount of interactions on the tablets, the amount of parallel vs. complementary work and the amount of reading and rechecking grouped elements, to identify collaborative strategies in a setup with personal tablets as additional tools.

One-way repeated measures analysis of variance (RM-ANOVA) tests were conducted to analyze statistical differences between the base condition and the condition with the tablets regarding task performance (task solution time) and efficiency (e.g. number of interactions). Bonferroni correction was applied when comparing individual conditions. Further analysis of variance tests were conducted to find differences between the different transfer techniques. Friedman tests were used for analyzing 5-point Likert-scale ratings from post-experiment questionnaires and Wilcoxon Signed-Rank Tests with Bonferroni correction (significance level of 0.017) as post-hoc tests for comparing differences between pairs of conditions. For all other tests, a significance level of 0.05 was used.

5 Results and Discussion

5.1 Task Solving Strategies

During the study we observed how participant pairs solved the sorting task and their use of the shared and personal displays. As there was no best way to find the solution, strategies varied between groups but also between conditions with different transfer techniques. All pairs collaborated close together in the conditions without tablets, except one team that worked parallel for short periods. In our interview one participant said that the condition without tablets "*forced them to collaborate together*". Therefore the tabletop-only condition is not discussed in this section. Adding the personal displays allowed participants to choose diverse strategies. A common approach was to start with finding the first and last page of the chapter followed by looking for subsequent and/or precedent pages depending on the approach. Whereas the first page was placed either in the top or bottom left side of the shared surface, the last page was usually placed on the right side and represented the frame for the other documents.

The handling of the documents on the shared surface differed as well and was dependent on the used tactic. When a lot of documents were enlarged, more time was used for the arrangement of documents on the shared surface. Some teams did not enlarge documents at all and just used the tablets for reading. We calculated a scale factor of documents to determine their size over time. In the conditions with tablets, the documents' sizes over time (between 0.173 and 0.177) were even less than half of the sizes in conditions without tablets (0.371).

Within the conditions with the tablets, we identified four different strategies: The first strategy (*team-up*) is characterized by close collaboration throughout the task. The second strategy (*split-up*) is the exact opposite resulting in parallel work throughout the task. The third strategy (*prepare-and-conquer*) is separated into an initial phase of parallel work followed by close collaboration towards the end of the task. The fourth strategy (*divide-and-combine*) is characterized by recurring changes between close and loose collaboration. The four strategies are described in detail below, including examples of participants' tactics.

Team-Up. The first strategy was picked in twelve out of the 24 trials with tablets. It was characterized by very close collaboration between participants throughout most parts of the task. Participants did not spend time working on parallel activities (e.g. trying to find subsequent pages on their own) but teamed up to find relations between pages or even a specific page. Within this strategy, various tactics emerged. For instance, in the beginning participants opened five different documents on each tablet and started to search for relations between pages using the personal displays only. Other groups enlarged one document on the shared surface and tried to find subsequent pages on their tablets. However, this tactic of opening five documents on each tablet did not necessarily entail close collaboration. Other groups used this tactic to split up and work on their own.

Within the *team-up* strategy another tactic consisted of opening the same documents on both tablets to discuss relations between pages. Another approach was to work on the shared surface using the personal displays just occasionally. In this tactic the tablets were often positioned on the horizontal space on the shorter sides of the table and picked up for use when needed.

Split-Up. The second strategy involved splitting up of the task and working on separate activities in parallel resulting in loose collaboration. When splitting up, participants focused on reading on their tablets on their own, or browsed for pages on the shared surface without communicating or interacting with their partner. Only in five out of the 24 trials this strategy was chosen by three of eight teams. We expected pairs with participants who did not know each other from before to work parallel, but one of these two teams worked surprisingly close together. On the contrary, a couple decided to split up and both participants tried to find matching pages on their own.

Prepare-and-Conquer. In the third strategy partners began to work on parallel activities (*split-up*) during the initial phase of the task. In this period, participants tried to gain an overview of the chapter or just of a number of documents. Communication was reduced to be able to concentrate on reading. After participants found the first pages that belonged together, they switched to close collaboration (*team-up*) and worked together to complete the task as a team. The name of this strategy was determined by the initial "preparation" phase that helped "conquer" the task afterwards. This strategy was only chosen in four out of the 24 trials by three of the eight teams.

Divide-and-Combine. The fourth strategy was characterized by recurring changes in the amount of close collaboration. Teams regularly evaluated and changed their strategies based on the current situation. The name of this strategy in based on its pattern: Participants "divided" the task into parallel activities when necessary and "combined" their results repeatedly. For instance, after both participants arranged some documents on the shared surface and found some matching pages on the tabletop, they changed to parallel work with both participants searching for one specific document. When working in parallel, participants often spent time reading on the tablets. The *divide-and-combine* strategy was also found when participants were working on the shared surface only. However, only three out of the 24 trials with tablets (by three of the eight teams) were explicitly rated as *divide-and-combine* strategy. When we could not find separation of tasks but only close collaboration, trials were rated as *team-up* strategy.

We compared task completion times of all *team-up* strategies with the conjunction of times of all others strategies. On average, teams that used the *team-up* strategy with close collaboration were faster than teams choosing one of the other three strategies with parallel work included. Task completion times of team-up strategies were only slightly longer compared to the experiments without the use of tablets.

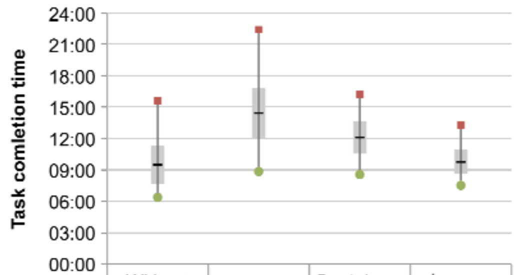
5.2 Transfer Techniques and Task Performance

Errors. All eight participant pairs were successful in completing the sorting task. Only in eight of the 32 trials (eight groups, four conditions) six groups made mistakes in the final sorting of the pages, resulting in 32 wrong (out of 288) pairings overall. We could not find any significant differences between conditions regarding the error points. We further could not find any significant differences between the four book chapters regarding errors.

Time. The mean task completion time per trial was 11 minutes and 28 seconds (SD: 3 minutes 48 seconds). We could not find any significant differences between the four book chapters regarding solution times.

We assumed that the more complicated the transfer technique was, the longer participants would need to solve the task. However, our data showed the opposite behavior (Figure 6). Task performance measurements showed that the *buttons* technique was slower compared to the others. Nevertheless, participants rated the *buttons* technique as the easiest. The *lenses* technique was reported to be more complicated and required too many steps, but was the fastest. The times differed significantly across the conditions ($F(3,21) = 4.488$, $p = 0.014$), but only the in *buttons* condition times were significantly higher ($M = 14.429$, 95% CI [10.62, 18.24]) compared to the *lenses* condition ($M = 9.783$, 95% CI [7.96, 11.61]) with $p = 0.02$.

We investigated the results and searched for explanations with the amount of use of the tablets. Some groups used the tablets less in the *lenses* condition compared to the other two conditions. Yet conditions with less use of the tablets did not necessarily finish faster than other groups. The analysis of interactions per minute on the tablets showed significant differences ($F(2,14) = 4.610$, $p = 0.029$) between conditions. However, we could not find significant differences when comparing the conditions pairwise. Comparing the interactions per minute on the shared surface resulted in significant differences ($F(3,21) = 6.980$, $p = 0.002$) between the condition without tablets ($M = 62.506$, 95% CI [43.68, 81,33]) and the *buttons* conditions ($M = 35.301$, 95% CI [22.69, 47.92]) with $p = 0.025$.



	Without	Buttons	Containers	Lenses
Min	06:24	08:51	08:33	07:32
Max	15:35	22:23	16:10	13:18
Mean	09:31	14:26	12:06	09:47
StDev	03:24	04:34	02:56	02:11

Condition

Fig. 6. Minimum, maximum and mean task completion times for the four conditions including standard deviation

With the help of the video analysis we discovered that participants showed confused and uncontrolled behavior in the *buttons* and *containers* conditions. The analysis of the number of document transfers per minute from the shared to the personal displays revealed significant differences ($F(2,14) = 11.966$, $p = 0.001$) between the *buttons* condition ($M = 3.641$, 95% CI [1.99, 5.3]) and the containers condition ($M = 1.89$, 95% CI [0.77, 3.01]) with $p = 0.043$ and between the *buttons* condition and the *lenses* condition ($M = 1.135$, 95% CI [0.52, 1.75]) with $p = 0.011$. Participants often closed and reopened documents repetitively to check their contents. The costs (in terms of number of touches and completion time) for transferring a document to the tablet is higher the more complicated a transfer technique is. Therefore, in the *lenses* technique participants seemed to reflect more about which documents they wanted to examine and which documents had already been reviewed or currently had been excluded from investigation.

Qualitative Findings. The subjective rating of participants after the experiment showed clear preference of the *buttons* technique, followed by *containers* and *lenses* techniques. Reasons retrieved from interviews included the fast and easy use of the buttons beneath the documents. Additionally, participants could easily send documents to their partners, which was especially helpful when one participant wanted to send documents to both tablets. Advantages of the *containers* technique included the behavior that reminded users of the use of desktop computers and that the user feedback that was better compared to the other techniques. One participant noted that dragging the documents into the containers was time consuming. The *lenses* technique was rated as most complicated and not efficient. A reason was that a series of steps had to be followed to transfer documents. In contrast, one participant said that after he figured out how to use it, the transfer was surprisingly easy. Two participants rated the *lenses* technique as their favorite as it was fun and engaging to use. One participant claimed that having to move the tablet across the surface even increased the awareness of the participant's activity in the *lenses* technique.

The complexity of the *lenses* technique can be traced back to our workaround with the fiduciary marker that required additional steps for transferring the documents. Compared to the *buttons* and *containers* techniques that only required one interaction to transfer a document, the *lenses* technique required four: Pressing the button, moving the tablet across the surface, releasing the button and finally selecting the document(s) to open. Improving the detection of the tablet's position could potentially reduce the transfer to two steps.

Another difference between the techniques is the metaphor of transferring documents resulting in the separation of interactions. Whereas in the *buttons* and *containers* techniques users have to transfer documents *from* the shared display by touching the corresponding buttons or dragging them into containers, the *lenses* technique allows *grabbing* documents from the shared surface without interactions on the shared surface. The metaphor of "*taking a photo*", like one participant noted, also impacts the task itself. The personal displays can be used for examining the documents. The shared display acts as the link between participants, representing their progress, for arrangement of documents and for building groups of pages that belong together in order to solve the task. Saving time for scaling and arrangement of documents on the shared surface combined with considerate opening of specific documents and reading them on the tablets was one efficient strategy to solve the task.

5.3 Awareness

For identifying differences between conditions relating to awareness we analyzed qualitative self-assessments from participants from the post-experiment questionnaire. Although the awareness in the condition without tablets was rated slightly higher and in the *lenses* condition slightly lower, we could not find any significant differences ($\chi^2(3) = 1.966$, $p = 0.58$). Results from our interviews were overall balanced despite contradictory opinions about the perceived awareness.

One participant said: *"When using the tablets it is harder to communicate and you do not know what the other one is reading"*. Other participants stated that using the tablets resulted in higher awareness. When comparing the results to the subjective ratings of participants to our assigned strategies, teams with close collaboration (*team-up*) rated high awareness in the conditions with the tablets. Another participant noted that *"the tabletop gives a feeling of connectedness and what is going on"*. Since documents had to be grouped on the shared surface, it acted like a progress view for the task. The color-coding we used for highlighting opened documents on the shared surface helped participants to maintain an awareness of collaborators' activities. For instance, one participant stated: *"When I looked at the screen [shared surface] I knew what he was doing"*. He added that he could see what documents his partner had opened but not which exact document he was currently reading. Interestingly, one participant claimed that she was not aware what her collaborator was exactly doing in the condition without the tablets. She explained this with saying that the shared surface was messier because of all the enlarged documents on it, thus making it more difficult to maintain an overview.

6 Conclusion

Our study is the first to analyze a collaborative MDE with multi-touch input on both shared and personal displays. To investigate the effects of multi-touch personal devices on task performance, collaboration strategies and awareness in multi-display environments we proposed three different techniques for one-way transfer of documents from a shared surface to tablets serving as personal devices. In the first technique (*buttons*), each document featured a button for transferring the documents. In the second technique (*containers*), the shared surface featured two containers representing the participants' tablets. The third technique (*lenses*) allowed participants to pick up documents by moving the tablet above the tabletop. The three main contributions of this paper are:

First, the identification of four different strategies teams used to coordinate and collaborate, ranging from close collaboration to parallel work throughout the task. The identified collaboration strategies are specific to MDEs. Teams varied strategies between the conditions with different transfer techniques. These findings extend the importance of transient design [10] to allow users to follow specific strategies for multi-display environments. Adding personal displays to collaborative environments further allows for completing tasks on either display, therefore adding to the flexibility of the environment. In the condition without the additional tablets, some groups felt forced to close collaboration, which did not result in significant faster task completion times.

Second, the indicative assessment of users' performance with respect to three one-way transfer techniques in a multi-touch based MDE. The analysis of task completion times showed that the *lenses* technique was significantly faster compared to the *buttons* technique. This result could be attributed to the fact that the simplicity of the *buttons* technique led to an increased document transfer between devices and less coordinated task-solving strategies. Ratings and qualitative feedback from participants indicated that the sequence of interactions required to transfer documents with the lenses technique were too complex, which can be attributed to technical limitations of our implementation. Using an improved mechanism for recognizing and tracking tablets would eliminate the use of a fiduciary marker, simplifying the interaction. The costs for transferring documents from the shared surface onto the tablets (in terms of interactions) seemed to not only have an impact on task-solving strategies and coordination, but was also an important factor for participants when comparing transfer techniques. Further research could investigate the effect of costs on task-solving strategies. More research is also needed to analyze the suitability of the techniques for different scenarios by comparing task types.

Third, the assessment of users' awareness of their collaborator's activities in a multi-touch based MDE, which showed no significant differences between the conditions, although we expected the condition with the shared surface only to receive higher rating in awareness. Qualitative participant feedback showed that using the shared surfaces as task progress view and for providing feedback about activities taking place on the personal displays is important for maintaining awareness amongst users.

References

1. Community Core Vision, http://ccv.nuigroup.com
2. Echtler, F., Nestler, S., Dippon, A., Klinker, G.: Supporting casual interactions between board games on public tabletop displays and mobile devices. Pers. Ubiquit. Comput. 13, 609–617 (2009)
3. Forlines, C., Wigdor, D., Shen, C., Balakrishnan, R.: Direct-touch vs. mouse input for tabletop displays. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 647–656. ACM, San Jose (2007)
4. Fox, A., Johanson, B., Hanrahan, P., Winograd, T.: Integrating information appliances into an interactive workspace. In: Computer Graphics and Applications, pp. 54–65. IEEE, Los Alamitos (2000)
5. Free eBooks by Project Gutenberg, http://www.gutenberg.org
6. Ha, V., Inkpen, K.M., Whalen, T., Mandryk, R.L.: Direct Intentions: The Effects of Input Devices on Collaboration around a Tabletop Display. In: 1st IEEE International Workshop on Horizontal Interactive Human-Computer Systems, pp. 177–184. IEEE Computer Society, Washington (2006)
7. Hagedorn, J., Hailpern, J., Karahalios, K.G.: VCode and VData: illustrating a new framework for supporting the video annotation workflow. In: Working Conference on Advanced Visual Interfaces, pp. 317–321. ACM, New York (2008)
8. Hawkey, K., Kellar, M., Reilly, D., Whalen, T., Inkpen, K.M.: The proximity factor: impact of distance on co-located collaboration. In: ACM SIGGROUP Conference on Supporting Group Work, pp. 31–40. ACM, Sanibel Island (2005)
9. Hornecker, E., Marshall, P., Dalton, N.S., Rogers, Y.: Collaboration and interference: awareness with mice or touch input. In: ACM Conference on Computer Supported Cooperative Work, pp. 167–176. ACM, San Diego (2008)
10. Isenberg, P., Fisher, D., Morris, M., Inkpen, K., Czerwinski, M.: An exploratory study of co-located collaborative visual analytics around a tabletop display. In: IEEE Symposium on Visual Analytics Science and Technology (VAST), pp. 179–186 (2010)
11. Kaltenbrunner, M., Bencina, R.: reacTIVision: a computer-vision framework for table-based tangible interaction. In: 1st International Conference on Tangible and Embedded Interaction, pp. 69–74. ACM, New York (2007)
12. Kin, K., Agrawala, M., DeRose, T.: Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In: Graphics Interface 2009, pp. 119–124. Canadian Information Processing Society, Kelowna (2009)
13. MT4J - Multitouch for Java, http://www.mt4j.org
14. Müller-Tomfelde, C., Schremmer, C.: Touchers and mousers: commonalities and differences in co-located collaboration with multiple input devices. In: 26th Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1149–1152. ACM, Florence (2008)
15. Müller-Tomfelde, C., Wessels, A., Schremmer, C.: Tilted tabletops: In between horizontal and vertical workspaces. In: 3rd IEEE International Workshop on, Horizontal Interactive Human Computer Systems, TABLETOP 2008, pp. 49–56 (2008)
16. Nacenta, M.A., Gutwin, C., Aliakseyeu, D., Subramanian, S.: There and back again: cross-display object movement in multi-display environments. Human-Computer Interaction 24, 170–229 (2009)
17. Nacenta, M.A., Aliakseyeu, D., Subramanian, S., Gutwin, C.: A comparison of techniques for multi-display reaching. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 371–380. ACM, Portland (2005)

18. 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)
19. Olwal, A., Feiner, S.: Spatially aware handhelds for high-precision tangible interaction with large displays. In: 3rd International Conference on Tangible and Embedded Interaction, pp. 181–188. ACM, Cambridge (2009)
20. Plaue, C., Stasko, J., Baloga, M.: The conference room as a toolbox: technological and social routines in corporate meeting spaces. In: 4th International Conference on Communities and Technologies, pp. 95–104. ACM, University Park (2009)
21. Rekimoto, J.: Pick-and-drop: a direct manipulation technique for multiple computer environments. In: 10th Annual ACM Symposium on User Interface Software and Technology, pp. 31–39. ACM, New York (1997)
22. Rogers, Y., Lindley, S.: Collaborating around vertical and horizontal large interactive displays: which way is best? Interacting with Computers 16, 1133–1152 (2004)
23. Scott, S.D., Grant, K.D., Mandryk, R.L.: System guidelines for co-located, collaborative work on a tabletop display. In: 8th conference on European Conference on Computer Supported Cooperative Work, pp. 159–178. Kluwer Academic Publishers, Helsinki (2003)
24. Scott, S.D., Sheelagh, M., Carpendale, T., Inkpen, K.M.: Territoriality in collaborative tabletop workspaces. In: ACM Conference on Computer Supported Cooperative Work, pp. 294–303. ACM, Chicago (2004)
25. Shen, C., Everitt, K., Ryall, K.: UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 281–288. Springer, Heidelberg (2003)
26. Streitz, N.A., Geißler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W., Rexroth, P., Seitz, P., Steinmetz, R.: i-LAND: an interactive landscape for creativity and innovation. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 120–127. ACM, New York (1999)
27. Streitz, N., Prante, T., Müller-Tomfelde, C., Tandler, P., Magerkurth, C.: Roomware©: the second generation. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems, pp. 506–507. ACM, Minneapolis (2002)
28. Tan, D., Gergle, D., Mandryk, R., Inkpen, K., Kellar, M., Hawkey, K., Czerwinski, M.: Using job-shop scheduling tasks for evaluating collocated collaboration. Pers. Ubiquit. Comput. 12, 255–267 (2008)
29. Tang, A., Tory, M., Po, B., Neumann, P., Carpendale, S.: Collaborative coupling over tabletop displays. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1181–1190. ACM, Montréal (2006)
30. Tse, E., Shen, C., Greenberg, S., Forlines, C.: How pairs interact over a multimodal digital table. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 215–218. ACM, San Jose (2007)
31. Wallace, J.R., Scott, S.D., Stutz, T., Enns, T., Inkpen, K.: Investigating teamwork and taskwork in single- and multi-display groupware systems. Pers. Ubiquit. Comput. 13, 569–581 (2009)
32. Wilson, A.D., Benko, H.: Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In: 23nd Annual ACM Symposium on User Interface Software and Technology, pp. 273–282. ACM, New York (2010)
33. Wu, M., Balakrishnan, R.: Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In: 16th Annual ACM Symposium on User Interface Software and Technology, pp. 193–202. ACM, New York (2003)

Evaluating Physical/Virtual Occlusion Management Techniques for Horizontal Displays

Waqas Javed, KyungTae Kim, Sohaib Ghani, and Niklas Elmqvist

School of Electrical & Computer Engineering,
Purdue University, West Lafayette, IN 47907-2035, USA
{wjaved,kimk,sghani,elm}@purdue.edu

Abstract. We evaluate unguided and guided visual search performance for a set of techniques that mitigate occlusion between physical and virtual objects on a tabletop display. The techniques are derived from a general model of hybrid physical/virtual occlusion, and take increasingly drastic measures to make the user aware of, identify, and access hidden objects—but with increasingly space-consuming and disruptive impact on the display. Performance is different depending on the visual display, suggesting a tradeoff between management strength and visual space deformation.

1 Introduction

As digital tabletop displays are becoming increasingly inexpensive, common, and available to the mass market, they will also start to be integrated into everyday work environments. In fact, such devices may even totally replace standard desktop computers in the future; as a case in point, Wigdor et al. [1] describe the use of a tabletop display for a single user over the period of a year. However, tabletop displays are horizontal surfaces, and such surfaces invite placing physical objects on them [2], such as paper, books, and even coffee mugs. This will inevitably give rise to problems where these physical objects occlude virtual objects on the tabletop display (Figure 1).

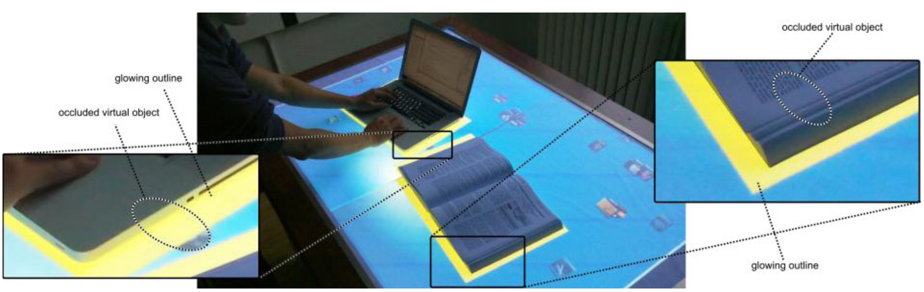


Fig. 1. Tabletop environment with physical objects (a laptop computer and a book) resting on its surface. The physical objects are occluding virtual objects on the display, so their outlines have been made to glow using one of our occlusion management techniques to indicate this.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 391–408, 2011.

Mitigating occlusion is not a new problem and has been addressed previously: for 2D desktop environments, hidden overlapped windows can be retrieved using the taskbar in Microsoft Windows or the dock in Mac OS X. More advanced techniques make windows semi-transparent [3], cut holes in them to expose hidden items [4], or spread them out to make hidden content accessible [5]. Other work considers situations when the user's hand and arm occlude the display on touch and pen-based systems [6–8]. However, with the exception of contextual display bubbles [9] and relations between virtual and digital media [2], little work has been performed on managing occlusion between virtual and physical objects beyond the user's own body.

We propose a framework for occlusion management in such hybrid physical/virtual environments based on three levels of occlusion management: awareness (knowledge that a particular target exists), identification (target is recognizable but cannot be interacted with), and access (the ability to fully interact with the target). These levels, together with additional characteristic parameters, define the design space of tabletop occlusion management. Using this design space, we derive six techniques that represent the spectrum of occlusion management: adding glow to the outlines of physical objects (GLOW, see Figure 1), displaying miniature icons of occluded objects (ICON), creating hybrid piles of virtual and physical objects (PILE), introducing a small overview map (MMAP), displacing virtual objects to empty space (MOVE), and, finally, replicating hidden area in unoccluded space (REPL).

Our evaluation of these techniques begins with a formative study intended to elicit requirements and constraints for a more in-depth evaluation. We then present results from two quantitative user studies designed to determine which technique was most efficient in managing occlusion. Both studies involved 12 (different) participants who performed unguided (Experiment 1) and guided (Experiment 2) visual search in the presence of both physical and virtual distracting objects. For Experiment 1, MOVE was the fastest technique, but the situation is reversed in Experiment 2, where displacing objects has a negative impact. We speculate that a combination is optimal.

2 Related Work

Existing research has studied occlusion management for virtual objects, for touch and pen-based interaction, and for tabletop displays. Below we review these in detail.

2.1 Virtual Occlusion

Windows often overlap in 2D desktop systems. To interact with an overlapped window, users must move, resize, or change the stacking order of windows. To avoid this problem, window managers typically provide ways to cascade or tile windows [10]. A window index, like the Windows taskbar or the MacOS Dock, allows for directly accessing windows even if they are hidden behind other windows. Another approach is to mitigate occlusion by placing windows in regions with the least overlap [11].

Alpha blending [12] can be used to show the content of occluded windows using transparency. The problem with this technique is that it is difficult to visually determine which content belongs to which window. Multi-blending [3] addresses this issue through a framework of transparency filters to allow simultaneous display of transparent windows. Free-space transparency [4] selectively turns unimportant window regions transparent and important regions opaque to optimize visibility.

2.2 Physical Body Occlusion

Occlusion in both direct touch and pen-based interaction arises when the user's own hand or arm covers some portion of the display. Ample research has shown that this occlusion is problematic and can severely impede performance [13, 14]. Forlines et al. [15] found that occlusion causes decay in performance in one-dimensional tapping and crossing tasks. Supporting this finding, Leithinger and Haller [16], while investigating menu placement for tabletops, discovered that menus suffering from occlusion had significant disadvantages compared to other menus. In fact, Vogel et al. [8] found that the user's hand and forearm can occlude up to half of a 12-inch display.

Various interaction techniques have been designed with occlusion in mind [17–19]. Hancock et al. [13] improve awareness in menu placement for tabletop display by designing menus that detects user handedness. Brandl et al. [20] also designed occlusion-aware menus that are positioned around the table so that they are not occluded.

Some techniques actively mitigate occlusion for touch displays. Ramos and Balakrishnan [18] use a sinusoidal-shaped slider to reduce occlusion from the user's hand in pen-based displays. CrossY [17] uses right-to-left movement to reduce occlusion with right-handed users. Shen et al. [21] utilize a popup widget to overcome occlusion caused by the user's finger for touch displays. This is similar to Shift [7], where a copy of the occluded region is moved to a non-occluded region. Building on this, more recent work on occlusion-aware interfaces [6] models which area is occluded and uses this knowledge to display hidden objects in a non-occluded screen area.

2.3 Physical/Virtual Occlusion

The environment-aware display bubbles presented by Cotting and Gross [9] are perhaps most relevant to our work. Their technique projects structured light onto the table to acquire inadmissible areas, which are then used to deform the display bubbles.

Recent work by Steimle et al. [2] studied usage patterns of physical and virtual media on tabletop surfaces. They found the expected drawbacks of physical occlusion, such as information hiding and interaction costs, but also—interestingly—some desirable features for organizing and partitioning information (particularly for hybrid piles [22]). Their results indicate the need for tabletop interfaces to explicitly support physical interaction on, above, and around the surface, further motivating our work.

3 Motivation: Everyday Use of Tabletops

Already in 1991, Shneiderman [23] listed the advantages of touch displays and predicted that, as more research will overcome their disadvantages, they will become more and more popular. In 1993, Wellner [24] first described the idea of an office desk as a digital display. True to form, tabletops are nowadays becoming increasingly available to the mass market, and are starting to appear in showrooms, lobbies, and museums. With the decreasing cost of large display technology and increasing demand from a growing potential consumer base, the step to adoption in the living room—and certainly the office—cannot be far away. In a pioneering effort, Wigdor et al. [1] describe the use of a tabletop for long-term office use and its pros and cons.

Following this reasoning, it is plausible that in the future, digital tabletops will be commonly used as office desks—or rather, that many of our horizontal surfaces in offices (such as desks) will be appropriated as displays [25]. However, the physical affordance of horizontal surfaces also invite placing various physical items on them [2], such as coffee mugs, papers, books, etc. When the desk is also a digital display, these items will start to hide virtual objects on the display.

4 Occlusion Management on Tabletop Displays

We define *tabletop occlusion management* as methods that give awareness, identification, or access to virtual objects that are occluded by physical objects on tabletops.

4.1 Detecting Physical Occlusion

The first step towards *managing* physical occlusion is to *detect* it. For tabletop displays built using diffuse illumination (DI) [26], this is relatively easy. DI tabletops have active infrared emitters that cause reflections on a physical object resting on the tabletop surface, allowing the camera tracking system to detect its actual outline. Taken together, these outlines represent occluded space on the display, and become the input for all types of occlusion management techniques.

The single remaining difficulty is for tabletop systems that already rely on DI technology for the touch interaction (not all do—for example, SLAP widgets [27] use FTIR [28] for touch and DI for detecting objects), forcing us to distinguish between physical occluders and the hands of the users. We suggest using a heuristic based on the motion of an object—if an object has not moved for T (e.g., 10) seconds, we regard it as an occluder. Other approaches would be to require a minimum size for an object to be classified as an occluder, or to put fiducial markers on the undersides of physical objects on the display (the latter is not a general solution, however).

Finally, we should note that physical objects are three-dimensional, and thus also have a geometric extension above the table surface. This means that accurate occlusion detection should consider the line of sight of the user, whereas our implementation only utilizes the 2D footprint of an object on the tabletop surface. However, our techniques would also work with a line-of-sight based approach, and thus the technical limitations with our implementation do not affect the validity of our results. We discuss this aspect in more detail in the design implications section.

4.2 Levels of Occlusion Management

We define four levels of increasing occlusion management:

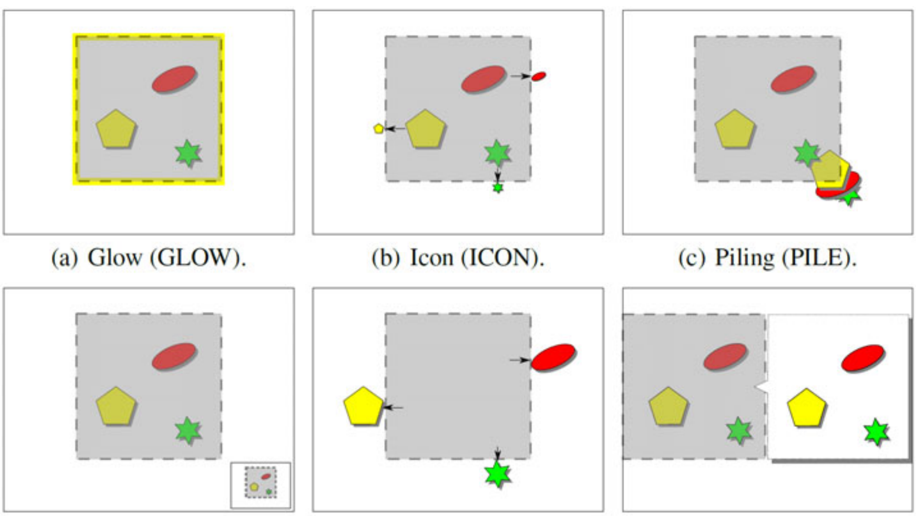
– **No knowledge**: A virtual object is fully occluded by a physical object, and the user is unaware of its existence. The virtual object will remain hidden until the user happens to move or lift the occluding object.

- **Awareness**: A user that is aware of an occluded virtual object knows that one or more virtual objects are occluded by a physical object, but does not necessarily have knowledge about the number or identity of the hidden objects.
- **Identification**: A user that can identify an occluded virtual object can see (part or a copy of) the virtual object, but cannot interact directly with the object. To interact with the object, the user must lift or move the occluding physical object.
- **Access**: Having access to an occluded virtual object means that the user is able to fully view and interact with the object as if it was not hidden at all.

4.3 Design Parameters

We suggest four additional design parameters for classifying occlusion management:

- **Space Consumption**: Amount of space utilized by the technique in addition to the space utilized by the physical objects themselves.
- **Visual Clutter**: Degree of visual clutter produced. With occluded objects no longer visible, how much new clutter is introduced by the technique?
- **Spatial Transformation**: Impact on the geometric properties of the visual space, including distortion, displacement, and rotation.
- **Intention Support**: Intentional occlusion possible. The user may want to intentionally hide virtual objects using physical objects. Is this scenario possible?



(a) Glow (GLOW). (b) Icon (ICON). (c) Piling (PILE).

(d) Minimap (MMAP). (e) Move (MOVE). (f) Replicate (REPL).

Fig. 2. Overview of occlusion management techniques evaluated in this paper. The large gray rectangle represents the footprint of a physical object placed on a tabletop display, thus occluding three virtual objects drawn on the actual display (ellipse, star, and pentagon).

4.4 Techniques

We derive six techniques for different levels of occlusion management. In Figure 2, we use a standard situation with three virtual objects (an ellipse, star, and pentagon) being occluded by a single physical object placed on the tabletop surface.

– **Awareness-supporting**: This type of technique only makes the user aware of occluded virtual objects without actually showing *which* objects are occluded. For this reason, such techniques take the least amount of visual space of all proposed techniques. They also respect the user's intention when hiding an object.
  - *Glow (GLOW)*: Displays a colored glow around the outline of physical objects that are occluding one or more virtual objects (Figure 2(a)).

– **Identification-supporting**: These techniques provide not only awareness, but also allow users to identify the occluded virtual objects. For this reason, the space consumption and visual clutter is often higher than for awareness-supporting techniques. However, while these techniques typically give hints about the occluded content (through miniatures or portions of the objects), the user is still required to manipulate the physical object to interact with the virtual objects.
  - *Icon (ICON)*: This technique displays a fixed-size small visual representation—i.e., an icon—of an original occluded object (Figure 2(b)). The icon will be displayed on the nearest edge of the occluding physical object in order to convey some information on where the hidden virtual object is placed.
  - *Piling (PILE)*: This technique mimics a traditional office desk where people pile papers and books into neat stacks (Figure 2(c)). The technique organizes hidden virtual objects into piles under the occluding physical object, thus showing a small portion of the original objects as a hint to the user (similar to [22]).
  - *Minimap (MMAP)*: The *minimap* technique shows a miniature overview of the virtual space with physical objects outlined (Figure 2(d)). The user can refer to the overview to find the location of hidden objects.

– **Access-supporting techniques**: This family of techniques gives the user full input and output access to hidden virtual objects. To make this possible, these techniques have high space consumption and visual clutter. However, the result is an entirely occlusion-free visual space with no physical object interaction necessary; on the other hand, users cannot intentionally cover virtual objects with physical items.
  - *Move (MOVE)*: Moves hidden objects to the nearest edge of its occluding object (Figure 2(e)). Items can be connected to their original positions using a virtual spring, causing them to spring back when the occluder is removed.
  - *Replicate (REPL)*: Inspired by Vogel et al. [6], this technique replicates the entire screen content of a particular physical occluder in a callout located in unoccluded space as close to the occluder as possible (Figure 2(f)). Objects are revealed in their natural size and their relative position is maintained.

We do not claim that the above six techniques are exhaustive examples of the design space of physical occlusion management; additional possibilities include object indices, distortion, and labels, etc. We also do not claim that these techniques are novel; in fact, some explicitly derive from existing techniques such as Shift [7], hybrid piles [22], and occlusion-aware interfaces [6]. However, we do think these techniques are a representative sampling of the occlusion management design space.

5 Designing the Evaluation

Because our work is the first (to the best of our knowledge) to study techniques for managing occlusion between physical and virtual objects on a tabletop display, we wanted to perform a canonical evaluation of the problem. To inform the design of this evaluation, we first conducted a formative evaluation involving graduate students from our university. Below we discuss the design decisions made based on this study.

5.1 Evaluation Platform

For the purposes of evaluating the above techniques, we implemented a graphical framework in Java for multitouch tabletop displays based on the Piccolo 2D graphics API and the TUIO [29] touch event protocol. The framework detects both finger touches as well as physical objects on the surface—an object that has not moved for a given time (10 seconds, in our case) is regarded as an occluder. We encourage the reader to view the paper's video for footage of our occlusion management software.

5.2 Formative Evaluation

We conducted a formative evaluation of the testing platform with three participants. All participants performed a jigsaw puzzle task where they reassembled a picture split into 3×3 tiles back into the original image. The tiles were randomly placed on the tabletop surface but were not rotated (in fact, tiles could only be moved, not rotated). We also randomly placed four sheets of paper (physical occluders) on the tabletop surface. Participants performed one reassembly task for each of our techniques.

Because of the formative nature of this evaluation, we did not collect any correctness or time measurements. However, we took careful notes of participant behavior. The qualitative feedback was very positive despite tracking inaccuracies inherent to our tabletop. While the jigsaw puzzle task with sheets of paper as occluders worked well, we found it difficult to control the placement of the occluders on the tabletop surface. Furthermore, even though we asked participants not to pile or remove occluders from the display, a common strategy was to initially move them out of the way before solving the task, thus neutralizing the impact of any specific occlusion management technique. This was an issue we needed to address in the full experiment.

5.3 Design Decisions

Based on the formative evaluation, we made the following decisions:

- **Low-level visual search tasks**: Given our intent to build a solid foundation for physical/virtual occlusion management, we decided to select fairly low-level tasks involving *visual search* [30] for our evaluation. Because visual search generally is split into guided and unguided search (depending on whether the user has a priori knowledge of the spatial position of the target or not), we decided to perform two separate experiments, one for each of these. We argue that these tasks are building blocks for higher-level tasks involving more complex objects like text documents and applications, so our results should therefore generalize even to such objects.

- **Static occlusion**: Existing work [6, 8] on physical occlusion between a user's arm and the display deals with intrinsically dynamic situations with moving targets and occluders. However, the primary problem on tabletops is *static* occlusion caused by physical objects placed on the table surface. For this reason, and for the sake of simplicity, we do not consider dynamic situations in our evaluation. Nevertheless, our occlusion management techniques handle dynamic occlusion equally well as static occlusion, so we think that the results should again generalize to other settings. In fact, our framework implementation requires physical objects to be static on the tabletop surface for at least 10 seconds to be regarded as occluders. Without being able to otherwise differentiate interaction touches and occluders, we think is a reasonable strategy for realistic occlusion management implementations.
- **Small targets**: Occlusion can occur at many scales (such as for small icons vs. large documents) as well as on many levels (i.e., partial vs. total occlusion). On the basis that large and complex visual objects can generally be decomposed into many smaller objects, we opted to design our tasks for small and generic targets.
- **No physical objects**: Based on our experience from the formative evaluation, we chose not to include any actual physical objects in this study, but instead simulated them using virtual objects placed in the top layer of the visual space. This enabled us to fully control the placement of physical occluders for the purposes of the experiment. While this is naturally counterintuitive in a study on physical/virtual occlusion, we do not believe this affects the validity of our results in any way—the efficacy of the techniques themselves are independent of the source of occlusion.
- **Immovable occluders**: Again drawing on our observations from the formative evaluation, we decided not to allow occluders to be moved to avoid having users spend the first portion of each trial moving items to one side of the visual space. While this is clearly an artificial constraint imposed for mostly technical reasons, we can defend this by the fact that in many cases, it is either not practical to move objects away to clear space for interaction (such as for heavy books, large stacks of paper, keyboard or monitors), or the physical items are intimately tied to the task the user is performing (citing research papers, reading manuals, writing notes, etc).

6 Experiment 1: Unguided Visual Search

Our objective with this study was to derive the properties of our occlusion management techniques for unguided visual search. In perceptual studies, *unguided visual search* is defined as an active scan of a visual environment for a particular target whose location is not previously known [30]. The goal was to study technique performance under different space and object conditions. In particular, we opted **not** to compare techniques to a baseline with no occlusion management because our pilot testing made it obvious that the comparison was in favor of all of the new techniques.

6.1 Apparatus

We conducted the experiment on an 1.2m×0.9m (approximately 81 inch) DI [26] multitouch tabletop display equipped with with two DLP projectors, each with

1280×800 resolution (for a total of 1600×1280). The projectors were powered by a computer running Microsoft Windows 7.

6.2 Participants

We recruited 12 paid volunteers (8 males, 4 females) for the experiment. Ages ranged from 21 to 29 (average 25.3, median 23) years, and all participants had normal or corrected-to-normal vision with no color deficiency (self-reported).

6.3 Task

We designed our experiment as a basic instance of unguided visual search [30], where the participants were asked to find a known *target object* in an unstructured visual space. The visual space consisted of both virtual and physical objects:

- **Virtual objects**: These objects are graphical elements drawn on the display, and include the target as well as a set of *distracting objects* that have a similar visual appearance as the target. More specifically, **all** virtual objects were 3×3 rectangular grids of white cells, with one cell colored red. For the target, the colored cell was always the center one, whereas objects with other positions of the colored cell were distractors (Figure 3 gives an overview). The rationale for this design was that targets should not be preattentively distinguishable from distractors.
- **Physical objects**: These represent *physical occluders* (such as books, stacks of paper, or coffee mugs) placed on the tabletop surface, thereby potentially hiding virtual objects on the display. To allow us to fully control the number and placement of occluders, these were actually **not** physical in our experiment; instead, they were represented by gray rectangles (300×300 pixels in size) placed in a layer above the virtual objects. Tapping on an occluder's rectangle temporarily *lifted* the object to make occluded content visible; tapping again on the screen would return that occluder to its original position. Only one occluder could be lifted at a time. We enforced a 1-second delay between lifts to model real occluders.

The space was configured so that all virtual objects were randomly placed behind a physical occluder, and thus initially occluded. Furthermore, all objects were placed so that they did not overlap (except between virtual objects and occluders) and were within reach of a participant without having to move; however, some techniques like MOVE could potentially cause an object to be moved out of reach or to partially overlap another object. Because of the high number of virtual objects, we could not rely on chance to make some physical occluders not occlude any object, so we controlled this by enforcing one occluder to be a dummy occluder, i.e., with no virtual object underneath. This gave some benefit to GLOW by cutting down search space.

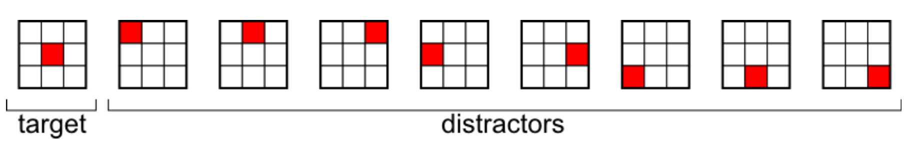


Fig. 3. Virtual objects in Experiment 1. The object on the left is a target, others are distractors.

6.4 Experimental Conditions

We included three factors in our study: Technique, Object Size, and Object Number.

- **Technique:** We included all of the occlusion management techniques above as a factor *T*: GLOW, ICON, PILE, MMAP, MOVE, and REPL.
- **Object Size:** Because identification-supporting techniques often rely on miniature versions of virtual objects, we included Object Size *S* as a factor in the experiment with two levels: *large* and *small*. Our pilot testing helped us set values to these levels: small objects were 42×42 mm (60×60 pixels) and fully recognizable at normal magnification, and large objects were 56×56 mm (80×80 pixels) or about a quarter of physical occluders (210×210 mm or 300×300 pixels). In particular, the small size was selected so that the miniature versions of targets would not be trivially identifiable using the ICON, MMAP, and PILE identification techniques.
- **Object Number**: Techniques with high visual clutter and space consumption are presumably sensitive to the total number of virtual objects in a visual space. To test this hypothesis, we included an Object Number factor *N* with two levels: *few* and *many*. Again, pilot testing helped us find values for these: 10 objects vs. 20.

6.5 Experimental Design

We used a full factorial within-participants design with the following factors:

	12	participants
×	6	**Techniques** *T* (see above)
×	2	**Object Sizes** *S* (Small, Large)
×	2	**Object Number** *N* (Few, Many)
×	10	repetitions (training excluded)
	2880	total trials (240 per participant)

Trials were organized in blocks for each technique. Block order was balanced using a Latin square across participants to counteract learning effects; other factors were randomized within blocks. The experiment platform collected completion time.

6.6 Procedure

Participants received training with each technique before each block until they indicated that they were ready to proceed. Trials were interleaved with an intermission screen during which participants could rest, also ensuring that their hands were in a neutral position prior to each trial. When starting a trial (and the timer), the screen changed to show the visual space with physical and virtual objects. Participants were instructed to complete the task as fast as possible. The trial ended (stopping the timer) when the participant found and tapped the correct target (or its copy in the REPL technique). Tapping on the wrong virtual target made it flash red to indicate the error.

After finishing each technique block, participants were given a structured interview where they were asked about that technique. A full experimental session, including training and post-block structured interview, lasted approximately 60 minutes.

6.7 Hypotheses

H1 Access-supporting techniques (MOVE, REPL) will yield faster performance than identification-supporting techniques (ICON, MMAP, PILE).

H2 Identification-supporting techniques (ICON, MMAP, PILE) will yield faster visual search performance than awareness-supporting ones (GLOW).

6.8 Results

For each trial, we measured the completion time, the number of occluder lifts, and the number of taps on objects. For lifts and taps, we found no significant differences between techniques, and so we disregard these measures for the rest of this analysis.

Table 1. Significant effects of factors on time (Experiment 1, RM-ANOVA)

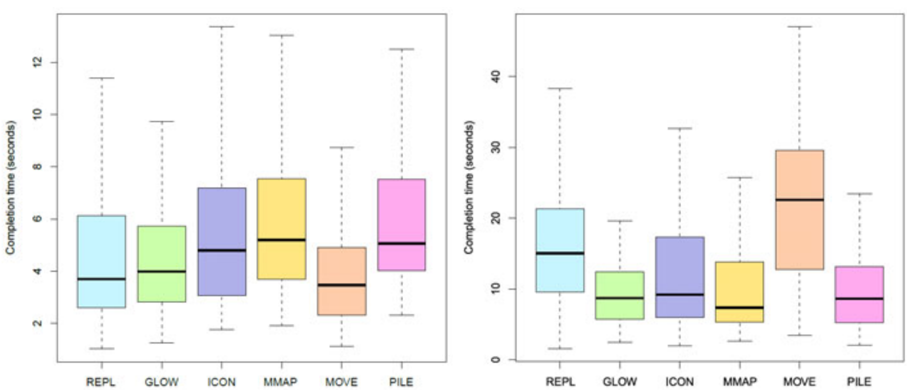
Table 2. Significant effects of factors on time (Experiment 2, RM-ANOVA)

Factors	df, den	F
Technique (*T*)	5,55	**16.53
Object Size (*S*)	1,11	**23.61
Object Num (*N*)	1,11	**60.14
*T * S*	5,55	*3.05
*T * N*	5,55	**5.05
*S * N*	1,11	**17.87
*T * S * N*	5,55	0.05
Technique (*T*)	5,55	**23.29
Occl. degree (*O*)	1,11	0.025
*T * O*	5,55	2.12

Significance results:

* = $p \leq .05$

** = $p \leq .0001$



(a) Experiment 1. (b) Experiment 2.

Fig. 4. Completion times for occlusion management techniques for both experiments.

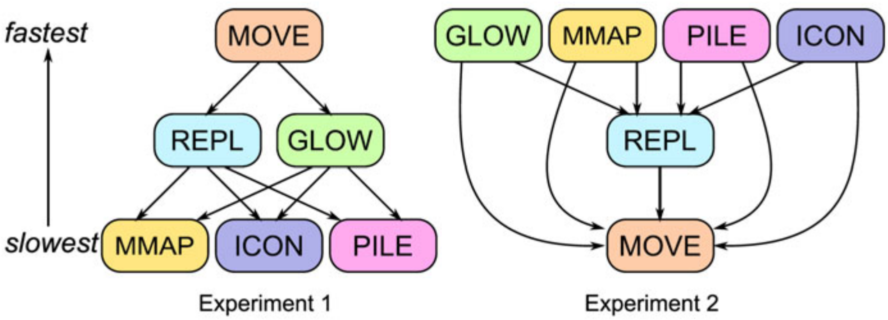


Fig. 5. Significant time differences ($p < .05$) for techniques in Experiment 1 (left) and 2 (right)

Figure 4(a) shows the completion times for each technique (three standard deviation outliers removed per participant and condition). The distribution of completion times was not normal, so we analyzed its logarithm using a repeated-measures analysis of variance (RM-ANOVA, normality and homogeneity of variances assumptions checked with Shapiro-Wilks and Bartlett's tests, respectively, and both valid after log-transformation)—see a summary of completion time effects in Table 1.

We used Tukey's HSD test to analyze pairwise differences for techniques on completion time—significant results ($p < .05$) are summarized in Figure 5 (left). In particular, MOVE performed significantly faster than any other technique ($p < .05$), followed by REPL and GLOW, and finally MMAP, ICON, and PILE as the slowest.

6.9 Discussion

Our results show that for unguided visual search, MOVE and REPL perform significantly better than ICON, PILE, and MMAP. While we have only tested a sample of the occlusion management design space, we think that this means that access-supporting techniques (MOVE, REPL) perform better than identification-supporting techniques (ICON, PILE, MMAP) for unguided visual search. This confirms **H1**, and it is clearly due to the fact that MOVE and REPL make all targets visible and selectable without requiring any occluders to be lifted. Also, the drawback of these techniques—displacing objects from their original positions—had no impact on the nature of the task. Identification-supporting techniques, on the other hand, generally show a smaller version of the targets, making recognition more difficult.

For **H2**, we were surprised to find that GLOW performs significantly better than ICON, PILE, and MMAP, thus rejecting the hypothesis. This may partially be due to the reason given above: recognizing the target is harder for these techniques. With GLOW, participants immediately started lifting the occluders to find the target object, whereas for ICON, PILE, and MMAP, they would essentially have to perform two visual search tasks: first for miniature versions on the whole space, and then again under the occluder they lifted. This was also supported by feedback from interviews.

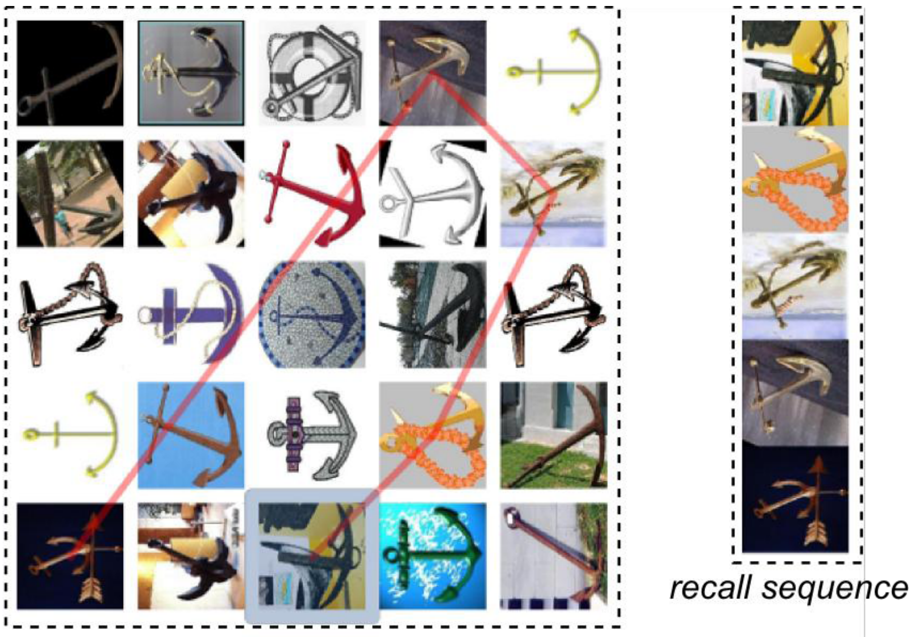


Fig. 6. Typical scenario for Phase I in Experiment 2

7 Experiment 2: Guided Visual Search

Beyond the pure unguided visual search in the first experiment, we wanted also to study the impact of occlusion management for *guided visual search* [30] where the spatial location of objects is important. We thus designed Experiment 2 as a game of Concentration with face-up photographs. Beyond task, we used the same apparatus and procedure as in Experiment 1. Below we discuss differences between the studies.

7.1 Participants

We recruited 12 paid volunteers (10 males, 2 females) for the experiment. Ages ranged from 22 to 28 (average 24.3, median 23) years, and all participants had normal or corrected-to-normal vision with no color deficiency (self-reported).

7.2 Task

Guided visual search tasks require that the participants have some a priori knowledge of target location. We achieved this using a revisitation task consisting of two phases:

I. **Learning:** Users were shown a 5×5 grid of photographs (each 100×100 mm, or 150×150 pixels, in size and visible at all times). A randomly selected photograph would be highlighted. By tapping and holding the highlighted photograph for a minimum *dwell time* (1.5 seconds), the highlight would move on to a new photograph. This was repeated 5 times. We also showed a red path connecting all of the photographs to visit to further aid participants in learning the locations.

II. **Recall:** After having visited (and presumably memorized) the location of 5 photographs, the user was asked to revisit the photographs (still face-up and visible) in the recall phase. During this time, a set of physical occluders (again simulated using gray rectangles) was added to the display in random locations, thus obscuring part of the grid. As before, occluders could be lifted, one at a time, by tapping. We enforced the photograph sequence from the learning phase, so participants could only proceed to the next photograph in the sequence by selecting the correct photograph (i.e. no out-of-order selections were possible).

Figure 6 shows a typical scenario during Phase I (learning).

We timed participants in Phase II, but there was no time limit on Phase I. The sequence of photographs was shown on the right side of the screen with the next photograph to visit highlighted, and out-of-order selections were not possible. We did this to eliminate errors in the experiment, allowing us only to analyze completion time.

Finally, photographs for each trial were selected from the same random category of the CalTech-101 image dataset [31]; this ensured that all images had the similar motif. We manually eliminated images that were almost identical, however.

7.3 Experimental Conditions

We included two factors: Technique *T* (same as Experiment 1) and Degree of Occlusion *O*. The degree of occlusion was simply the number of physical occluders in Phase II (recall).We used the same size occluders as before (210×210 mm or 300×300 pixels) and included two levels: *low* (1 occluder) and *high* (2).

7.4 Experimental Design

We used a full factorial within-participants design with the following factors:

	12	participants
×	6	**Technique** *T* (see above)
×	2	**Degrees of Occlusion** *O* (Low, High)
×	4	repetitions (training excluded)
	576	total trials (48 per participant)

As before, trials were organized in blocks for each technique, and block order was counterbalanced using a Latin square (*O* randomized). The experimental platform collected completion time as well as the number of taps and lifts per trial.

7.5 Hypothesis

H3 Identification-supporting techniques (ICON, MMAP, PILE) and awareness-supporting techniques (GLOW) will yield faster guided visual search performance than access-supporting techniques (MOVE, REPL).

7.6 Results

Figure 4(b) shows the completion times for each technique. We again observed that completion time measurements did not obey a normal distribution, so we analyzed its logarithm using an analysis of variance (RM-ANOVA, normality and homogeneity of variances assumptions checked as above and valid after log-transformation) to find significant effects of the factors on completion time—see Table 2 for results.

To compare techniques directly, we performed a Tukey's HSD posthoc test. Results of this analysis can be found in Figure 5 (right). In particular, all other techniques were significantly ($p < .05$) faster than REPL and MOVE, and REPL was in turn significantly ($p < .05$) faster than MOVE (almost a reversal from Experiment 1).

7.7 Discussion

The results from Experiment 2 confirm hypothesis **H3**—techniques with less impact on the spatial position of targets performed better than access-supporting techniques that drastically affect the spatial arrangement on the screen. This is not a surprising result given that guided visual search relies on a priori knowledge of target locations.

Our results also show that REPL performed significantly faster than MOVE, indicating that the design of access-supporting techniques is an important factor to consider. In this case, the fact that REPL maintains the relative positions of objects in the locality of occluded space is probably the reason for this difference in performance. In other words, both MOVE and REPL transform the absolute positions of objects, but REPL at least maintains relative positions.

8 Implications for Design

Based on our findings from both of the experiments, there is no one technique that performs optimally for both guided and unguided visual search. There is clearly a tradeoff between a technique's impact on virtual objects on the display, and its capability to mitigate occlusion due to physical objects. In such circumstances, a balance must be struck between these two factors. The GLOW technique performs only second to MOVE in Experiment 1, and is among the fastest techniques in Experiment 2. It is also relatively lightweight to implement and does not significantly alter the interaction in existing applications. For these reasons, we recommend the use of GLOW as a starting point for managing physical/virtual occlusion on tabletops.

Having said that, there is nothing that prevents us from combining techniques to achieve better performance. For example, the minimap (MMAP) technique is a fairly nonintrusive addition to any application, and may actually serve additional purposes for overview+detail in maps, visualizations, and graphical editors. It is also true that these techniques represent only a sample of the occlusion management design space. Although we based our designs on a systematic exploration of this space, it is possible that other designs exist that would perform better. This is a topic for future research.

Furthermore, our experiments tested only low-level visual search, but general tabletop applications consist of many other types of tasks at different abstraction levels, as well as potential targets of many different shapes and sizes. However, we argue that visual search is a central activity in interactive computing, so our results should generalize to many real-world tabletop applications. Our results for relatively small objects should also generalize to components of larger objects. Still, it would be interesting to study occlusion management in longitudinal and more realistic settings.

8.1 Detecting 3D Occlusion

This work does not consider the fact that there is more to physical objects than their 2D footprint on the surface of the tabletop display, and that physical objects also extend along the vertical axis. This means that objects may be hidden from the user's viewpoint by physical objects even if the display area they inhabit is not actually covered—in other words, the occlusion problem between physical and virtual objects on tabletop displays is 3D and not 2D. Provided that we had full knowledge of the visible and hidden areas of the display from the user's viewpoint, we could adapt our techniques to also work in this setting. Our work is a first step towards fully occlusion-aware tabletop interfaces, but more research is needed to reach this ultimate goal.

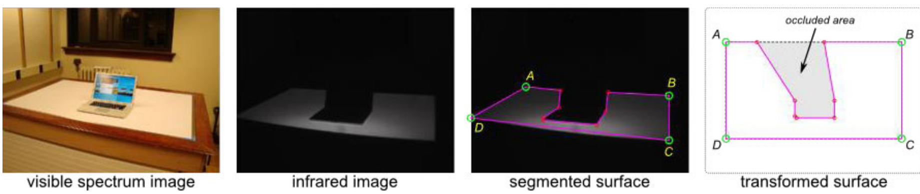


Fig. 7. 3D occlusion detection process using an infrared camera mounted on the user's head

We have experimented further towards reaching this goal by mounting an infrared webcam on a headset so that the camera sees what the user sees (similar to Vogel [8]). With a DI tabletop, the surface of the table will glow with infrared light—except in places where the surface is occluded from the user's point-of-view (Figure 7). This allows us to separate the visible area of the tabletop as a 2D shape. By transforming the 2D shape of visible surface to the coordinate space of the table, we can derive occlusion information that can be used as input for any of the techniques proposed here.

9 Conclusion and Future Work

We have presented a novel approach to manage occlusion between physical items resting on tabletop displays and virtual objects projected on the display. Our primary objective is not to find an optimal occlusion management technique, but rather to characterize the performance of different techniques under different spatial characteristics and tasks. Results from two user studies on guided and unguided visual search, respectively, clearly show that the different design alternatives have different strengths and weaknesses, and one outcome from our work is perhaps that the more lightweight approaches are more promising for the general case.

In the future we plan on continuing to study occlusion management for tabletop displays. Much work remains to be done for detecting physical/virtual occlusion in 3D. We also envision techniques that distort the visual space to aid visual search.

Acknowledgements. This material was partially supported by the U.S. Department of Homeland Security's VACCINE Center under award no. 2009-ST-061-CI0001.

References

1. Wigdor, D., Penn, G., Ryall, K., Esenther, A., Shen, C.: Living with a tabletop: Analysis and observations of long term office use of a multi-touch table. In: Proceedings of the IEEE Workshop on Tabletop Displays, pp. 60–67 (2007)
2. Steimle, J., Khalilbeigi, M., Mülhäuser, M., Hollan, J.D.: Physical and digital media usage patterns on interactive tabletop surfaces. In: Proc. ACM ITS, pp. 167–176 (2010)
3. Baudisch, P., Gutwin, C.: Multiblending: displaying overlapping windows simultaneously without the drawbacks of alpha blending. In: Proc. ACM CHI, pp. 367–374 (2004)
4. Ishak, E.W., Feiner, S.K.: Interacting with hidden content using content-aware free-space transparency. In: Proc. ACM UIST, pp. 189–192 (2004)
5. Ramos, G., Robertson, G., Czerwinski, M., Tan, D., Baudisch, P., Hinckley, K., Agrawala, M.: Tumble! Splat! Helping users access and manipulate occluded content in 2D drawings. In: Proc. ACM AVI, pp. 428–435 (2006)
6. Vogel, D., Balakrishnan, R.: Occlusion-aware interfaces. In: Proc. ACM CHI, pp. 263–272 (2010)
7. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: Proc. ACM CHI, pp. 657–666 (2007)
8. Vogel, D., Cudmore, M., Casiez, G., Balakrishnan, R., Keliher, L.: Hand occlusion with tablet-sized direct pen input. In: Proc. ACM CHI, pp. 557–566 (2009)
9. Cotting, D., Gross, M.H.: Interactive environment-aware display bubbles. In: Proc. ACM Symposium on User Interface Software & Technology, pp. 245–254 (2006)
10. Bell, B., Feiner, S.: Dynamic space management for user interfaces. In: Proc. ACM UIST, pp. 239–248 (2000)
11. Bell, B., Feiner, S., Höllerer, T.: View management for virtual and augmented reality. In: Proc. ACM UIST, pp. 101–110 (2001)
12. Porter, T., Duff, T.: Compositing digital images. Computer Graphics 18(3), 253–259 (1984)
13. Hancock, M.S., Booth, K.S.: Improving menu placement strategies for pen input. In: Proc. GI, pp. 221–230 (2004)
14. Inkpen, K., Dearman, D., Argue, R., Comeau, M., Fu, C.L., Kolli, S., Moses, J., Pilon, N., Wallace, J.R.: Left-handed scrolling for pen-based devices. International Journal of Human Computer Interaction 21(1), 91–108 (2006)
15. Forlines, C., Balakrishnan, R.: Evaluating tactile feedback and direct vs. indirect in pointing and crossing selection tasks. In: Proc. ACM CHI, pp. 1563–1572 (2008)
16. Leithinger, D., Haller, M.: Improving menu interaction for cluttered tabletop setups with user-drawn path menus. In: Proc. IEEE Tabletop, pp. 121–128 (2007)
17. Apitz, G., Guimbretière, F.: CrossY: a crossing-based drawing application. ACM Transactions on Graphics 24(3), 930 (2005)
18. Ramos, G., Balakrishnan, R.: Fluid interaction techniques for the control and annotation of digital video. In: Proc. ACM UIST, pp. 105–114 (2003)
19. Zeleznik, R.C., Miller, T.: Fluid inking: augmenting the medium of free-form inking with gestures. In: Proc. GI, pp. 155–162 (2006)

20. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., To, P.: Occlusion-aware menu design for digital tabletops. In: Ext. Abstr. of ACM CHI, pp. 3223–3228 (2009)
21. Shen, C., Hancock, M.S., Forlines, C., Vernier, F.D.: CoR2Ds: Context-rooted rotatable draggables for tabletop interaction. In: Proc. ACM CHI, pp. 1781–1784 (2005)
22. Khalilbeigi, M., Steimle, J., Mühlhäuser, M.: Interaction techniques for hybrid piles of documents on interactive tabletops. In: Ext. Abstr. of ACM CHI, pp. 3943–3948 (2010)
23. Shneiderman, B.: Interface—touch screens now offer compelling uses. IEEE Software 8(2), 93–94, 107 (1991)
24. Wellner, P.: Interacting with paper on the DigitalDesk. CACM 36(7), 86–96 (1993)
25. Harrison, C.: Appropriated interaction surfaces. Computer 43(6), 86–89 (2010)
26. Çetin, G., Bedi, R., Sandler, S. (eds.): Multitouch Technologies. NUI (2009)
27. Weiss, M., Wagner, J., Jansen, Y., Jennings, R., Khoshabeh, R., Hollan, J.D., Borchers, J.O.: SLAP widgets: bridging the gap between virtual and physical controls on tabletops. In: Proc. ACM CHI, pp. 481–490 (2009)
28. Han, J.Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In: Proc. ACM UIST, pp. 115–118 (2005)
29. Kaltenbrunner, M.: reacTIVision and TUIO: A tangible tabletop toolkit. In: Proc. ACM ITS, pp.9–16 (2009)
30. Treisman, A.: Search, similarity, and integration of features between and within dimensions. Journal of Experimental Psychology: Human Perception & Performance 17(3), 652–676 (1991)
31. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In: Proc. Workshop at IEEE CVPR (2004)

Usage and Recognition of Finger Orientation for Multi-Touch Tabletop Interaction

Chi Tai Dang and Elisabeth André

Augsburg University, Human Centered Multimedia,
Department of Computer Science, Universitaetsstr. 6a, 86159 Augsburg, Germany
{LastName}@informatik.uni-augsburg.de

Abstract. Building on the observation that finger orientation is an inherent part of human's interaction in the real world, exploiting finger orientation for multi-touch tabletop interaction would facilitate more natural interaction techniques. We motivate this by means of examples where the finger orientation improves or enriches interaction. Afterwards, we present a simple and fast approach to detect the finger orientation reliably for multi-touch tabletop interaction. The steps involved are computationally cheap and therefore suit the needs of tracking software operating under time-critical conditions. We show that the presented approach enables the detection of finger orientation also for fingers that touch the tabletop surface only slightly. Further, recognition rates on real data gained from the camera within a multi-touch tabletop are presented in order to give a measure for the precision and reliability of the presented approach.

Keywords: Finger Orientation, Multi-Touch, Tabletop, Tracking, Interaction.

1 Introduction

Over the last years, various multi-touch sensing technologies evolved and have matured out of their infancy, for example vision-based techniques [1-5] or techniques based on electrical capacitance [6-8]. Such technologies attracted a lot of attention not only from the research community, but also from the general public as well as the industry. Consequently, commercial products based on these technologies became publicly available, for example in the form of horizontal tabletops [6, 9, 10, 11], large displays [12], desktop computers [13, 14], notebooks [13, 14], or smartphones and tablet-sized devices [8]. Particularly smartphones and small multi-touch devices are now commonplace and more and more users become familiar with multi-touch interaction. Commercial horizontal tabletops are not as widespread as small mobile devices mainly because of their higher costs. However, they laid the foundation for multi-touch interaction design and still pose a rich playground for interaction research due to their experimental character and the possibilities of the bigger input space for interaction.

The novelty of those technologies is primarily attributed to interaction techniques and applications involving direct-touch with bare fingers and also in combination with

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 409–426, 2011.

physical artifacts representing digital information. Such interaction techniques are entitled to be natural since users may use their fingers to interact with digital objects in the same way as they would do with physical objects. Indeed, directly touching objects and manipulating them with multiple fingers pose a natural form of interaction that benefits from users' manual dexterity. Direct interaction with fingers allows to use a multi-touch tabletop computer straight away without learning how to use the input device. Humans use their fingers every day for any number of tasks and therefore they are used to interact with their fingers in the real world. To some extent, humans consciously make use of directions in this interaction to complement or convey their intention, for instance when pointing at objects or indicating a direction.

From a software application's point of view, the sensing technologies provide touch properties which have to be interpreted correctly by the application so as to enable the expected natural interaction. Common touch properties used therefor include finger position [6, 15], contact shape [16, 17], or contact area's size [18]. In [19], the authors empirically evaluated and discussed several finger input properties and the usage of those for multi-touch tabletop interaction. Their evaluation yielded guidelines for widget-design and several proposals for widget-designs where finger orientation, that is the direction which the finger points to, plays an important role. To our knowledge, the finger orientation has not been used for interaction in multi-touch applications as also claimed in [19]. One possible reason may be the unavailability of finger orientation in freely available tracker software, such as [15] or [20], which provide only the aforementioned common touch properties. Reliable methods to detect finger orientation unambiguously and with high precision would foster integration of finger orientation detection into tracker software.

Apparently, employing finger orientation to extend or complement natural interaction would offer valuable potential for a more natural form of interaction, since finger orientation is an inherently integrated aspect in human's interaction in the real world.

This paper is structured as follows: In the next two sections, we motivate the use of finger orientation for natural interaction by giving design examples for user interfaces and interaction techniques. Section 4 introduces into the technical background followed by a description of a naive approach and our new approach for detection of finger orientation. In Section 6, we present recognition rates for the presented approaches and discuss benefits and issues. Finally, we summarize and conclude our work in Section 7.

2 Usage of Finger Orientation

Incorporating finger orientation into natural interaction poses chances for improvements for several issues pertaining to touch user interfaces and interactions: manipulation, occlusion, selection, and adaptation. Both interaction techniques and user interfaces can be designed to account for finger orientation as exemplified follows.

Manipulation. Wang and colleagues [21] present a variety of interaction techniques, such as orientation-sensitive widgets, that would be enabled by robust techniques to recognize finger orientations. Such widgets take the finger orientation into account and enable function selection combined with parameter adjustment requiring only little space for interaction. Another example for this interaction technique could be a rotary switch that offers only two states to pin or release objects on the workspace, see Figure 1.

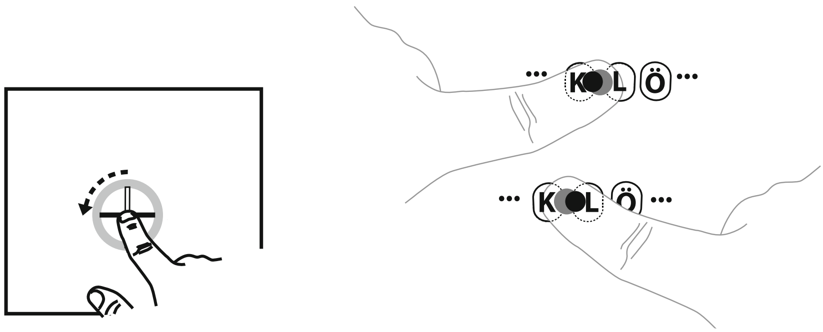


Fig. 1. Rotary switch *(left)*; Adaptation of reported input point *(right)*

Occlusion. Information about finger orientation would allow us to determine areas that are potentially occluded by the hand and adapt the display of graphical objects accordingly. For instance, widgets could be re-oriented in such a way that they are visible for the most part. Further, techniques such as *occlusion-aware pop-ups* or *occlusion-aware dragging* presented for pen or stylus interaction in [22] could be afforded by means of the finger orientation and realized for interaction with bare fingers. Thereby, hierarchical menus or tooltips might appear in non-occluded area or occluded text segments could be shown in callouts to support selection tasks.

Selection. If the orientation of fingers was recognized, users would be able to use their fingers of either one hand or both hands for pointing at objects displayed on the surface allowing them to select distant objects in a more natural manner as depicted in Figure 2 at the left-hand side. Users could also use thumb and index finger of one hand to span an open angle for object selection. Another option opens up if users make use of both hands to span two open angles as sketched in Figure 2 at the right-hand side. The two selected areas may intersect with each other and create a third selection area which might be used for object selection.

Adaptation. Vogel and colleagues [23] show that the reported input point from a finger touch differs from the intended target location by an offset. This offset is due to the fact that users perceive an input point different than the real target location. Based on this observation, they suggest an adaptation to correct the reported input point by this offset. Here, the finger orientation could provide the direction in which the adaptation should be applied. As an example, the interaction with a virtual keyboard on small touchscreens would benefit from such an adaptation as sketched in Figure 1. Without adaptation, the reported input point (gray filled circle) would result in ambiguous key selection whereas with an adaptation, the reported input point would better fit the perceived input point (black filled circle).

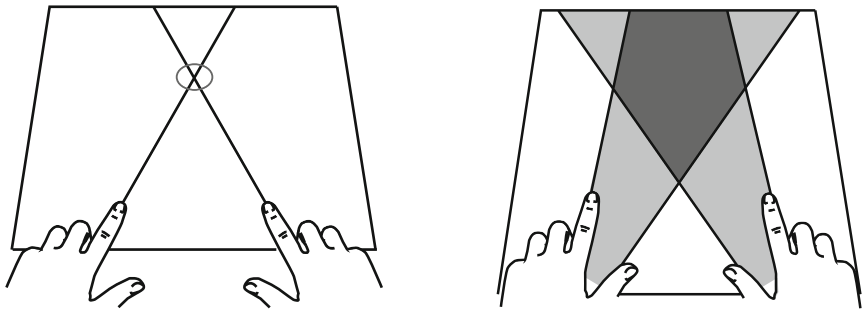


Fig. 2. Distant object selection *(left)*; Fingers spanning open angles for selection *(right)*

From a pragmatic point of view, user interfaces and interaction techniques that integrate finger orientation would help mitigate one of the main issues in natural interaction, namely the arm fatigue issue [24, 25]. By reducing the overall hand and arm movements for manipulation or selection tasks, users' hand and arm fatigue would be diminished as well. Not only single interaction techniques can be improved but also higher-level recognition tasks, such as the distinction between one- and two-handed interactions [26], would be enabled by reliable detection of finger orientation. In [26], we present a mechanism to distinguish hands based only on the location and orientation of finger contact areas.

3 Related Work

Much work has been done on detecting finger orientation in 3D space employing multiple cameras or color images, for example [27, 28, 32], whereas only less work covers detection of finger orientation for 2D touch sensing technologies or in combination with infrared images.

Malik and colleagues [29] presented the Visual Touchpad which utilized two color cameras mounted above the touchpad to detect user's hands and fingers. They identified fingers' positions through computer vision methods to find the fingertips on a hand contour. The hand contour is also used to determine the finger orientation for each fingertip. Their approach is based on color images and a direct view on the hands which is quite different from prevalent multi-touch sensing technologies employing infrared images and a bottom view on the sensing surface.

Wang et al. [21] proposed an approach to unambiguously determine the finger orientation that is based on the contact areas produced by finger touches. They fit an ellipse into the contact shape and use the longer ellipse axis for determination of the finger orientation. Moreover, they observe the center point's variation of the contact areas when a finger lands on the surface to disambiguate the finger orientation. Their approach is suitable for sensing technologies that provide only finger contact areas, whereas sensing technologies such as Diffused Illumination offer more potential for detection of finger orientation with a higher precision. In their work, they also show that finger orientation is a useful input property that can be employed to enhance user interactions.

A rather simple and inexpensive way to integrate finger orientation in multi-touch tabletop interaction was conducted by Marquardt et al. [30]. They employed the Microsoft Surface table [9] and a glove which was tagged with several fiduciary markers. The tabletop system was able to detect the markers together with their orientation, thus allowed them to derive finger orientation and to identify individual parts of the hand and their orientations. Wearing gloves is contrary to natural interaction, but Marquardt's approach allows for rapid prototyping of interaction techniques and they indicated that integrating more properties from fingers and hands provide rich opportunities for interaction design.

We pointed out that finger orientation poses meaningful chances to extend or complement natural interaction. In order to be used for user interaction, finger orientation must be determined reliably and with high precision. For this purpose, we present and discuss a simple approach that reliably detects the finger orientation for tabletop setups which employ sensing techniques similar to Diffused Illumination. This approach even detects the finger orientation for difficult cases, where the finger touches the surface only slightly.

4 Diffused Illumination

A lot of multi-touch sensor technologies employing infrared light emerged during recent years [1, 2, 3, 5, 33, 34]. In this paper, we refer to the Diffused Illumination approach which operates on raw images that are comparable with those created by technologies such as DSI or LLP [5]. Diffused Illumination setups have infrared sources mounted in the interior of the table which emit infrared light towards the tabletop surface. Objects such as fingers or physical artifacts on or above the tabletop surface reflect the infrared light back into the table. This reflected infrared light exposes a camera sensor inside the table which delivers images that show the illuminated objects. Therefore, when users' fingers approach the tabletop surface, the captured images show the whole hand with its fingers where the brightness of the hands' and fingers' pixels indicate the closeness to the tabletop surface, for example Figure 3 on the left-hand side. Hence, they offer rich possibilities for object detection through computer-vision methods that take the pixel values into account.

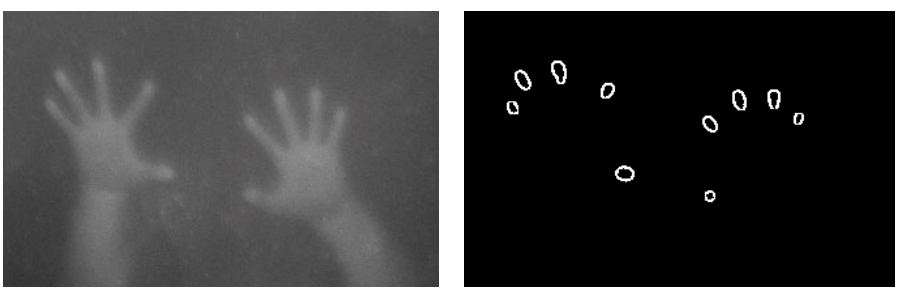


Fig. 3. Infrared image *(left)* and corresponding contour image *(right)*

The contact area's pixels of a finger that touches the surface are brighter than the pixels of non-contact areas. Hence, a typical process chain to find finger contact areas and to locate the fingers' coordinates exploits this fact and utilizes a brightness threshold to distinguish contact area from non-contact area. The steps comprise of converting the raw camera image into a blob image based on the brightness threshold and afterwards converting the blob image into a contour image. Within the resulting contour image, each contour represents a contact area of one finger that touches the tabletop surface as shown in Figure 3 on the right-hand side. Finally, the coordinates of each finger contact can be determined as the corresponding contour's center location.

5 Finger Orientation Detection

In this section, we will first outline a straightforward way to determine finger orientation which serves as a baseline for the performance comparison in Section 6.3. The remainder of this section starting from Section 5.2 illustrates our new approach in detail in order to ease integration in tracker software.

5.1 A Naive Approach

A naive approach to determine finger orientation bases on the aforementioned contour image. The steps therefor can be accomplished easily with high-level functions that are part of the computer vision package OpenCV [31]. We will further use the term *ellipse method* to denote the approach described in the following.

When considering the contour image in Figure 3, we can spot each finger contour as an ellipse representing the finger contact area. Apparently, such a matter of fact enables to fit an ellipse into each detected contour and take the angles between x-axis and the corresponding longer ellipse axis to determine finger orientations as exemplarily sketched in Figure 4 for only one contour.

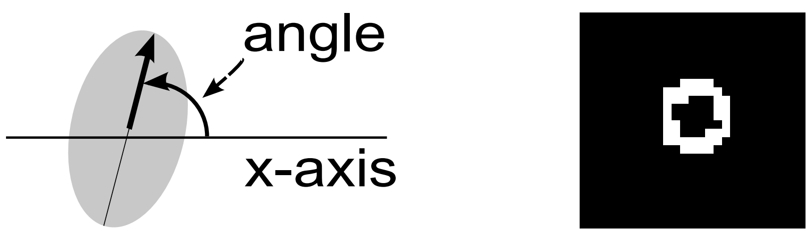


Fig. 4. Angle to *x-axis (left)*. A circular contour *(right)*.

However, this approach suffers from inherent weaknesses since it relies only on the contour of a finger contact. For instance, it produces ambiguous results for circular contours because the axes in a circle can point towards any direction. Furthermore, a detected finger orientation could be wrongly skewed by 180° since the longer axis of an ellipse possesses two possible directions.

180° Adjustment. With the aid of the raw camera image, we are able to resolve the finger orientation in case it is wrongly skewed by 180°. We detect this kind of ambiguity by walking the longer ellipse axis into both directions while looking for the direction that shows up a non-finger pixel at first. The thereby detected direction is where the fingertip ends, thus we adjust the afore-detected finger orientation if it is wrongly skewed. Henceforth, we will use the term *ellipse method + 180-adjust* to denote this addition to the *ellipse method*.

Discussion. *Ellipse method* and *ellipse method + 180-adjust* represent a straight forward way, but there are common cases in which finger interaction produces problematic contours for both methods. For instance, let us consider a child with small fingers touching the surface or finger touches that stem from users touching the surface only slightly. In these cases, the detected contour is very small and features a circular shape in the worst case which leads to imprecise or high deviant finger orientations as exemplarily depicted in Figure 4. This is due to the fact that for small contours, the fewer pixels that contribute to the contour, the more effect one pixel has on the detected finger orientation. This is even more worse if we consider that camera noise always randomly affect pixels of the contour which results in jumping values for the detected finger orientation.

To overcome these issues and furthermore increase precision and stability of the detected finger orientation, we make use of the difference in brightness of proximate pixels kept in the raw image. We draw on that information to determine the finger contour and derive the finger orientation from only a part of the finger contour.

5.2 A Simple and Precise Detection Algorithm

When considering images produced by Diffused Illumination setups, we can identify each finger with its outer contour as depicted in Figure 5. What each finger contour has in common are two quasi axially symmetrical lines that converge circularly at the fingertip.

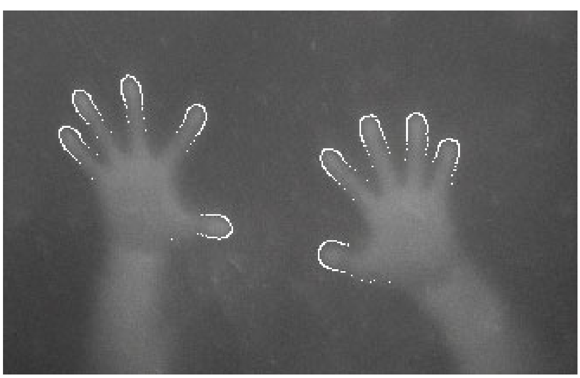


Fig. 5. Finger contours marked with *white lines*

Our approach exploits these symmetrical running lines to calculate the finger orientation. For each finger, we are able to detect these two lines and they always point into the direction of the finger orientation, thus the thereby identified finger orientation is reliable. Our algorithm comprises of three essential steps, each consecutively applied in the given order to each detected finger position. The detection of finger position was described in Section 4.

1. Detect the finger contour.
2. Determine the symmetrical lines.
3. Determine the angles in which the tracks point to and average them.

1. Detection of Finger Contour. As a first step, the outer finger contour for a given finger position has to be identified. This task is decisive for the remaining steps because the points of the contour intrinsically contribute to the precision of the finger orientation to be determined. Initially, we span a circle with the radius R pixels around the finger contact position and distribute points on that circle at an interval of 5° as depicted in Figure 6. We have chosen $R = 40$ pixels for our images in order to cover at least two times the finger width of small people's fingers and one time and a half the finger width of people with chubby fingers. This value must be adjusted for other tabletop setups depending on the resolution of the raw images and the projection size of the tabletop. For higher resolution cameras, the amount of points on the circle might be increased for a higher precision of detection.

The next step is to perform a search for a contour pixel starting from the finger contact's center position to each point that was distributed on the circle. To be precise, we process 72 paths at an interval of 5°. Within each search run, we compare the pixel value of each point on the path with the pixel value of the center position. Once the difference of their values exceeds a certain threshold, the search terminates and the pixel coordinate on the path is noted in a list.

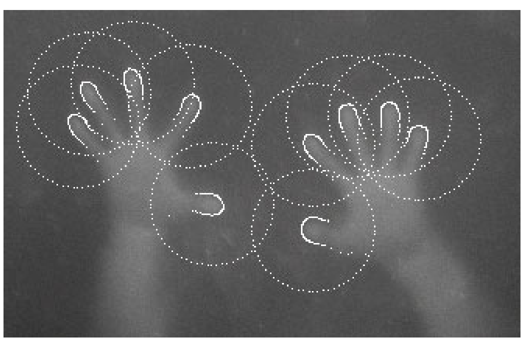


Fig. 6. *Points on circles* spanned around the finger contacts

For our implementation, we used the value 18 for the threshold. This threshold may vary for other DI settings depending on the brightness and contrast of the captured images. Here, adaptive threshold calibration based on a histogram of the image could compensate for environments with varying light conditions. The result of this step is a list which denotes whether a contour point was found or not for each of the 72 points on the circle. In particular, the list has the following properties:

Definition 1. *Properties of contour-point list*

1. *The ordered entries enumerated from 1 to 72 correspond to the points on the circle from 0° to 355° at an interval of 5°.*
2. *Each entry contains the coordinate (x, y) of a contour pixel.*
3. *If no contour pixel was found, then the entry contains the value (-1, -1).*

This procedure guarantees that in theory the finger contour found contains only the contour of the finger we consider and not a contour point of an adjacent finger. Since we start the search from the finger contact's center position, the first pixel that terminates the search must belong to the same finger. Furthermore, most of the searches terminate quickly because the distances between the finger contour and the finger contact's center position are short, see Figure 6.

2. Determine Symmetrical Lines. The following step operates only on the list defined in the previous section and determines the two quasi-symmetrical lines that contribute to the finger orientation.

In what follows, we treat the list as a ringbuffer where the subsequent entry of entry 72 points at entry 1 and vice versa. Furthermore, the defined names in italic type denote indexes into the list corresponding to the usage in Figure 7 unless otherwise noted. The name *list* denotes the afore defined list of points, where a single point can be retrieved by means of squared brackets as used in the programming language C.

At first, we perform a search for the biggest range that only consists of values of *(-1, -1)*. This range is defined as the gate $g_{start}, \ldots, g_{end}$ and represents the part where the finger is connected to the hand as depicted in Figure 7. The complementary range $c_{start} = g_{end} + 1, \ldots, c_{end} = g_{start} - 1$ is defined as the finger contour.

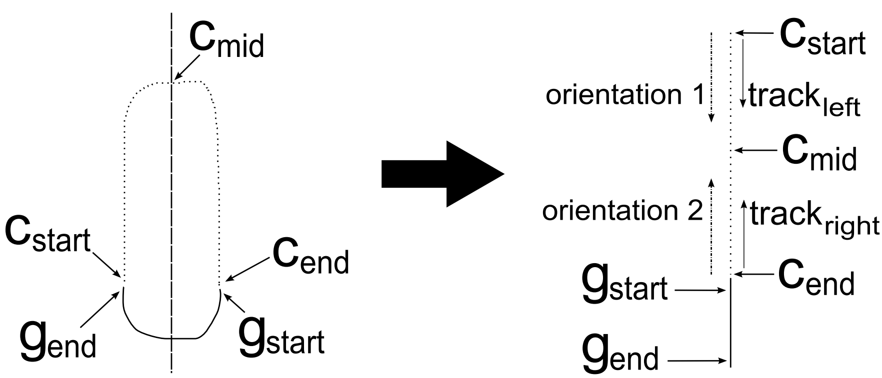


Fig. 7. Finger contour with start and end points *(left)*. List of contour points and its allocation *(right)*.

When considering the finger contour as two parts with c_{mid}: $c_{start} < c_{mid} < c_{end}$, there are two contour tracks that have to be cut back in order to remove the part representing the fingertip. The fingertip part can only be used for detection of finger orientation if the contour tracks are absolutely symmetrical to each other. This cannot be guaranteed due to different finger or hand pose which may destroy symmetry of the fingertip contour. Therefore, the fingertip part of the contour has to be removed as far as possible. We empirically determined that considering only 60% of the points in each contour track suffices to remove the fingertip points. That is the amount of points to include in each contour track is defined as $amount_c = 0.6 * (c_{end} - c_{start}) / 2$. This can be used independent of image resolution because higher resolution would result in more contour pixels, but the relationship between finger and fingertip remains the same. Considering only 60% of the finger contour is a tradeoff between including the contour that contributes to the finger orientation and omitting the fingertip contour. For the following steps, we further define the left contour track as $track_{left} = c_{start}, \ldots, c_{mid} - amount_c$ and the right contour track as $track_{right} = c_{end}, \ldots, c_{mid} + amount_c$, see Figure 7.

We have to catch a rare case in this step: if the list does not contain a gate range, then we have an elliptical finger contour as exemplified in Figure 8.

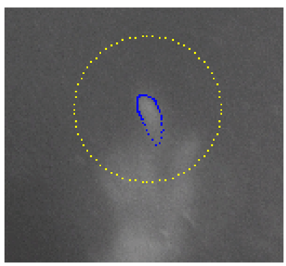


Fig. 8. Elliptical finger contour caused by low contrast

This case happens rarely and is caused by low contrast of the raw image which leads to falsely detected contour points stored in the list. In such a case, we skip the next step and proceed with the *ellipse method + 180-adjust* to detect finger orientation, which was described in Section 5.1.

3. Determine Average Angle. Both afore ascertained contour tracks of a finger contact contribute to the finger orientation based on the properties of the list defined in Definition 1. The properties imply that c_{start} is the starting point of the left contour track and c_{end} is the starting point of the right contour track. Because of that, we also know that the tracks $track_{left}$ and $track_{right}$ point into the direction where the finger is pointing to as well. Therefore, the last step serves to calculate the angles to the x-axis for $track_{left}$ and $track_{right}$ and average them afterwards to determine the final finger orientation.

Here, two methods with differing complexity may be applied. A computationally complex method bases on linear regression, whereby the points of a contour track are considered as a point cloud. A linear regression model is then fitted on this point cloud using least squares and the resulting slope of the regression line contributes to the finger orientation. This method is applicable for time-critical conditions if sufficient processing power is available.

A much simpler and faster approach that requires linear time draws on the slopes for each point of a contour track. For this approach, the coordinate denoted by the index c_{start} or c_{end} is defined as the anchor point respectively depending on the contour track to work on. That is either $anchor_x = list[c_{start}]_x$, $anchor_y = list[c_{start}]_y$ if we consider $track_{left}$ and otherwise $anchor_x = list[c_{end}]_x$, $anchor_y = list[c_{end}]_y$. The following instructions have to be applied on $track_{left}$ and $track_{right}$ independently. At first, the differences in x-value and y-value from the anchor point to all other contour track points are summed up and the resulting summed x-values and y-values are considered as the slope for the line running through the point cloud for the contour track. This is formally specified with equations 1, 2 and 3 where t_{start} and t_{end} are the indexes into the contour point list of the corresponding contour track. Finally, both slopes have to be averaged to determine the finger orientation.

$$sum_{\Delta x} = \sum_{i=t_{start}}^{t_{end}} list[i]_x - anchor_x \tag{1}$$

$$sum_{\Delta y} = \sum_{i=t_{start}}^{t_{end}} list[i]_y - anchor_y \tag{2}$$

$$slope = \frac{sum_{\Delta y}}{sum_{\Delta x}} \tag{3}$$

The simpler approach admittedly is prone to errors due to wrong initialization of the anchor point. However, the results we present in Section 6.3 show that this simple method still produces recognition rates that are quite close to those of the linear regression method.

6 Recognition Rates

In order to obtain recognition rates that have relevance for real data and that would occur in real tabletop interaction, we have captured a raw videostream from an infrared camera that was mounted in the table directed towards the surface. The video shows hands and fingers from a user who touches the surface the same way as he would do to interact with an application. Thereby, he uses combinations of one hand and two hands and with different finger combinations multiple times.

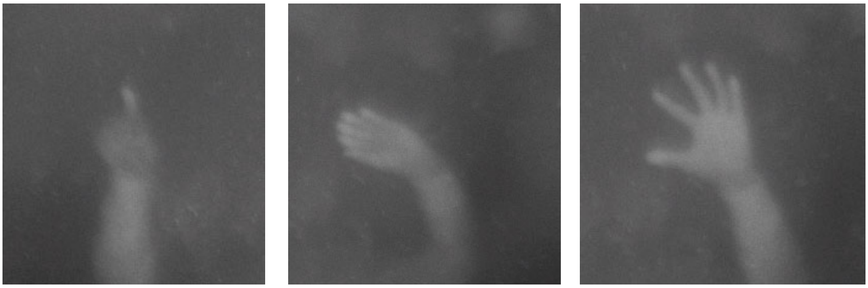


Fig. 9. From left to right: *(1)* move object with index finger, *(2)* move object with multiple fingers, *(3)* grasping with all fingers

For example, the finger combinations included combinations used to move objects with the index finger (Figure 9, Nr. 1) or multiple fingers (Figure 9, Nr. 2), grasping with all fingers (Figure 9, Nr. 3) or zooming with fingers of one hand (Figure 10, Nr. 1) or both hands (Figure 10, Nr. 2). As a result of this, the video contains multiple repeats of parallel landing and lifting fingers of multiple adjacent fingers. While the camera captured the video, the tabletop showed a black screen and gave no visual feedback to the finger contacts. Furthermore, the fingers in some of the captured images were blurry due to quick continuous hand and finger movements. Hence, the video shows images that occur in multi-touch tabletop setups for realistic continuous interaction situations. In addition to the video, we created reference finger orientations for each frame of the video that were used to evaluate the accuracy of the detected finger orientations.

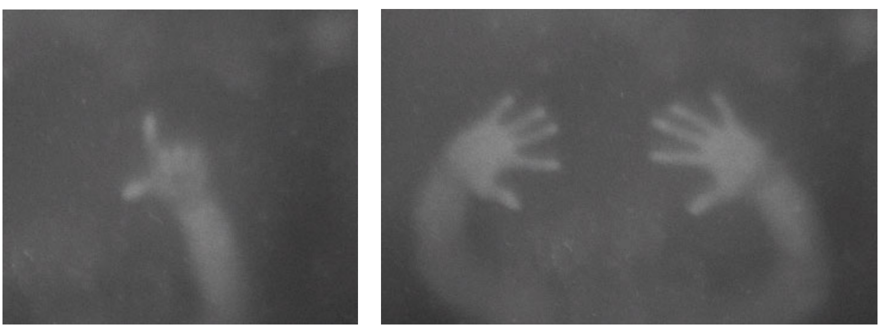


Fig. 10. From left to right: *(1)* zoom with fingers of one hand, *(2)* zoom with fingers of both hands

We chose a different method to test the performance of our approach than Wang did in [21] because of two reasons. First, Wang evaluated his finger orientation approach with an FTIR table which poses higher contrast and less noise thus creates lower challenges than DI tables. Second, he evaluated his approach by investigating to what extent the system's response was in line with the user's objectives when conducting a set of pre-defined tasks with the index finger.

In the present work, we propose a different evaluation approach which compares the finger orientation found by our algorithm with the finger orientation perceived and determined by three independent human judges. The advantage of our method is that it enables a task-independent evaluation and gives a measure for the precision that is more focused on continuous interaction.

6.1 Reference

The video for the reference finger orientations was available in an uncompressed format to preserve the raw data and comprises of 749 frames captured at 30 fps with a resolution of 640x480 pixels for each frame. The resolution covers a physical surface space of 60x80 cm. For creating the reference, each frame was converted to a blob image and the center position of each recognized blob was stored with its frame number in an XML-file. In all, 2007 finger contacts were detected. To ease the annotation of finger orientations and to generate the reference automatically, we have developed a graphical tool with the following functionality. Each frame along with the position of the detected finger contacts was presented to the annotator and the annotator had to manually adjust the finger orientations, which were detected through the *ellipse method + 180-adjust* approach. Overall, three persons annotated 6021 finger orientations and hence for each finger contact three finger orientations were created. The annotators were instructed to repeat an adjustment as often as required if they were not absolutely sure about the correctness of the visually perceived finger orientation and their adjusted finger orientation. The data gained through the annotations were used to build the reference finger orientations by calculating an average finger orientation for each finger contact.

6.2 Accuracy of the Reference Orientations

The reason why we averaged annotations procured from three different persons is that different persons may determine slightly different finger orientations due to camera noise, image contrast and image resolution. Our tool enabled the annotator to adjust the finger orientation angle with a precision of two decimal points as depicted in Figure 11.

Figure 11 also shows that the raw images are blurry and with low contrast. This makes it difficult to pinpoint an absolute finger orientation. Therefore, the annotation tool supplied a colored visual line alike to a ruler which assisted the annotator in gauging a proper finger orientation by choosing the mid of the acceptable angle ranges. This process enabled us to procure data from multiple annotators and take the averages for the reference. Overall, the reference finger orientations had a standard deviation of 3.34°.

6.3 Results

The precision of our approach is illustrated by comparing its recognition rates with the recognition rates of the *ellipse method* and the *ellipse method + 180-adjust* that

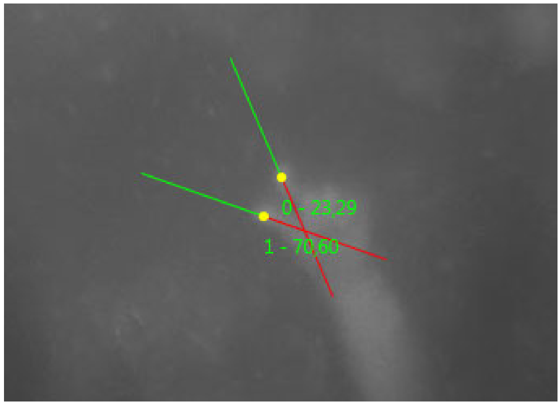


Fig. 11. Annotated index-finger and thumb

are explained in Section 5.1. As for our approach, we implemented both variants described in Section 5.2 to find a line running through the contour track. The term *Contourtrack* denotes our approach within this section. Table 1 shows the recognition rates for all four approaches and their standard deviation from the finger orientation in the reference. The table lists the recognition rates for four cases: 5°, 10°, 15°, and 25°. The recognition rate in the 5° column shows the percentage of recognized finger orientations at which the difference to the reference is less than 5°. The same interpretation applies to the 10°, 15°, and 25° column.

Table 1. Recognition rates and standard deviation for four methods: 1 = *ellipse method*, 2 = *ellipse method + 180-adjust*, 3 = *Contourtrack* + simple slope, 4 = *Contourtrack* + regression line.

Method	< 5°	< 10°	< 15°	< 25°	*SD*
1	41.5%	60.49%	68.81%	74.39%	86.46°
2	49.58%	74.34%	84.65%	92.53%	23.19°
3	75.29%	93.02%	96.86%	99.3%	6.17°
4	75.24%	94.87%	97.81%	98.9%	6.43°

The naive *ellipse method* shows low recognition rates (60.49% at 10°-error) and a high standard deviation of 86.46° which mainly stems from the wrongly 180° twisted finger orientations. This method hugely benefits from correcting the wrongly 180° twisted finger orientations (*ellipse method + 180-adjust*) as shown in the second row of Table 1. The extension improved the recognition rates (74.34% at 10°-error) and reduced the standard deviation significantly to 23.19°. However, the recognition rates of the naive methods *ellipse method* and *ellipse method + 180-adjust* show that their precision is not reliable enough to be used in realistic applications.

In the third row and the fourth row, the recognition rates of our new approach are shown, which are much better than the naive ellipse methods. Both methods' values show good results which exceeds 93% for a maximum error of 10°. If we accept a maximum error of 25°, then both methods provide recognition rates over 98%. Both standard deviations are less than or equal to 6.43° which is only 3.09° worse than the standard deviation in the human reference data with 3.34°.

Discussion. The precision of our approach increases with higher image resolution and image contrast. The more pixels that can be used for a finger contour, the better the precision. The better the image contrast, the better a finger contour can be detected. Preprocessing steps to reduce image noise or to improve image contrast would improve the performance, but we have not included such steps so as to keep the approach simple and fast. This way, the approach can be used as a lightweight add-on to already established tracker software. The deviation of our detected finger orientations from the reference orientations was mainly attributed to insufficient contrast of the raw image. Figure 12 illustrates two cases where the finger contour could not be determined correctly. In these cases, the difference between a finger pixel value and a background pixel value was too small, thus finger and background was not distinguishable with the threshold in use.

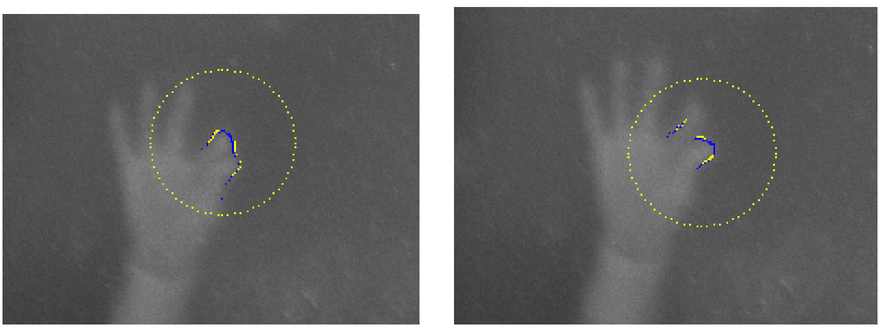


Fig. 12. Error cases with wrongly detected finger contour

Here, increasing the image resolution and image contrast improves the recognition rates of our approach. Nevertheless, the empirical data clearly shows that the presented approach detects finger orientation reliably and with high precision even with low-resolution images ($1cm^2$ covered by less than 64 pixels) and low contrast. Our implementation was evaluated on a PC with Core2Duo (2.83GHz) running at 30fps where it added only marginal latency (~0.19ms per finger contact for *Contourtrack* + regression line). Hence, latency does not have a negative impact on the perceived speed of user interaction.

In Wang's approach [21], few limitations exist which are handled correctly by our approach. First, their algorithm usually detects a wrong finger orientation if users touch the surface with the side face of the thumb due to the center displacement of the contact area that is different from the other fingers. Second, if users perform a "sliding down" gesture while touching the surface, the center displacement is again different from the center displacement of an index finger's landing process resulting in wrong detection of finger orientation. Our approach is based on the finger contour which shows the correct finger orientation in such cases. The Microsoft Surface [9] tabletop also supports detection of finger orientation and would have been a suitable candidate for performance comparisons. Unfortunately, it lacks of application support for providing unprocessed raw images from each single camera, which is required for a meaningful performance comparison. Therefore, we have chosen to utilize a self-made multi-touch table employing the Diffused Illumination technique.

7 Conclusions

Our work enriches natural interaction by considering finger orientation for user interface design as well as for interaction techniques. As we have shown, both can be extended in a meaningful sense regarding manipulation, occlusion, selection, adaptation and also for high-level tasks. Natural interaction will benefit from human's inherent understanding of direction, in particular finger orientation, that they practice every day. A basic requirement to enable further research in this topic is that the detection of finger orientation has to be reliable, precise and stable for continuous interaction. Therefor, our presented approach (*Contourtrack*) to detect finger orientation has proven to be accurate enough with recognition rates over 93%. Because of the approach's low algorithmic complexity, it suits the needs of time-critical processing thus foster availability of finger orientation in tracker software.

Acknowledgments. The work described in this paper is partially funded by the EU under research grant eCUTE (Reference: 257666).

References

1. Takeoka, Y., Miyaki, T., Rekimoto, J.: Z-touch: an infrastructure for 3d gesture interaction in the proximity of tabletop surfaces. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010, pp. 91–94. ACM, New York (2010)
2. Han, J.Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In: UIST 2005: Proceedings of the ACM Symposium on User Interface Software and Technology, pp. 115–118. ACM, New York (2005)
3. Hofer, R., Naeff, D., Kunz, A.: Flatir: Ftir multi-touch detection on a discrete distributed sensor array. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 317–322. ACM, New York (2009)
4. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., Saarikko, P.: It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In: CHI 2008, pp. 1285–1294. ACM, New York (2008)
5. NUIGroup: Multitouch techniques, `http://nuigroup.com/forums/viewthread/1982`
6. Dietz, P., Leigh, D.: Diamondtouch: a multi-user touch technology. In: UIST 2001: Proceedings of the ACM Symposium on User Interface Software and Technology, pp. 219–226. ACM, New York (2001)
7. Rekimoto, J.: Smartskin: an infrastructure for freehand manipulation on interactive surfaces. In: CHI 2002, pp. 113–120. ACM, New York (2002)
8. Apple inc., iphone, ipad, magic trackpad, `http://www.apple.com`
9. Microsoft surface, `http://www.surface.com`
10. Smart table, Smarttechnologies, `http://www.smarttech.com`

11. Archimedes solutions, http://www.multi-touch.de
12. Evoluce, Multitouch lcd monitors, http://www.evoluce.com
13. Dell, Latitude xt2, dell studio one 19, http://www.dell.com
14. HP Touchsmart tx2, touchsmart 600-1050de, http://www.hp.com
15. NUIGroup: Community core vision, http://ccv.nuigroup.com
16. Moscovich, T.: Contact area interaction with sliding widgets. In: UIST 2009: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 13–22. ACM, New York (2009)
17. Cao, X., Wilson, A.D., Balakrishnan, R., Hinckley, K., Hudson, S.E.: Shapetouch: Leveraging contact shape on interactive surfaces. In: Tabletop 2008, pp. 129–136 (2008)
18. Benko, H., Wilson, A.D., Baudisch, P.: Precise selection techniques for multi-touch screens. In: CHI 2006: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1263–1272. ACM, New York (2006)
19. Wang, F., Ren, X.: Empirical evaluation for finger input properties in multi-touch interaction. In: CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1063–1072. ACM, New York (2009)
20. NUIGroup: Touchlib, a multi-touch development kit, http://www.nuigroup.com/touchlib
21. Wang, F., Cao, X., Ren, X., Irani, P.: Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In: UIST 2009: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 23–32. ACM, New York (2009)
22. Vogel, D., Balakrishnan, R.: Occlusion-aware interfaces. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 263–272. ACM, New York (2010)
23. Vogel, D., Baudisch, P.: Shift: a technique for operating pen-based interfaces using touch. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems, CHI 2007, pp. 657–666. ACM, New York (2007)
24. Wigdor, D., Perm, G., Ryall, K., Esenther, A., Shen, C.: Living with a tabletop: Analysis and observations of long term office use of a multi-touch table. In: Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems, TABLETOP 2007, pp. 60–67 (2007)
25. Yee, W.: Potential limitations of multi-touch gesture vocabulary: Differentiation, adoption, fatigue. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 291–300. Springer, Heidelberg (2009)
26. Dang, C.T., Straub, M., André, E.: Hand distinction for multi-touch tabletop interaction. In: ITS 2009: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, pp. 101–108. ACM, New York (2009)
27. Jennings, C.: Robust finger tracking with multiple cameras. In: Proceedings International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems 1999 (1999)
28. Hung, Y.P., Yang, Y.S., Chen, Y.S., Hsieh, I.B., Fuh, C.S.: Free-hand pointer by use of an active stereo vision system. In: Proceedings of Fourteenth International Conference on Pattern Recognition 1998, vol. 2, pp. 1244–1246 (1998)
29. Malik, S., Laszlo, J.: Visual touchpad: a two-handed gestural input device. In: ICMI 2004: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 289–296. ACM, New York (2004)
30. Marquardt, N., Kiemer, J., Greenberg, S.: What caused that touch? expressive interaction with a surface through fiduciary-tagged gloves. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces ITS 2010, pp. 139–142. ACM, New York (2010)

31. OpenCV, open computer vision library, `http://sourceforge.net/projects/opencvlibrary`
32. Tianding, C.: A solution of computer vision based real-time hand pointing recognition. In: 27th Chinese Control Conference, CCC 2008, pp. 384–388 (2008)
33. Jackson, D., Bartindale, T., Olivier, P.: Fiberboard: compact multi-touch display using channeled light. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, pp. 25–28. ACM, New York (2009)
34. Izadi, S., Hodges, S., Butler, A., Rrustemi, A., Buxton, B.: Thinsight: integrated optical multi-touch sensing through thin form-factor displays. In: Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond: the Future of Displays and Interacton, EDT 2007. ACM, New York (2007)

Tangoscope: A Tangible Audio Device for Tabletop Interaction

Jörg Edelmann[1], Yvonne Kammerer[1], Birgit Imhof[1],
Peter Gerjets[1], and Wolfgang Straßer[2]

[1] Knowledge Media Research Center,
Konrad-Adenauer-Straße 40, 72072 Tübingen, Germany
[2] University of Tübingen, Department of Graphical Interactive Systems (GRIS),
Sand 14, 72076 Tübingen, Germany
{j.edelmann,y.kammerer,b.imhof,p.gerjets}@iwm-kmrc.de

Abstract. Tabletop installations allow multiple users to playback digital media simultaneously. With public speakers, however, simultaneous auditory content gets superimposed, leading to a confusing and disturbing user experience. In this paper, we present *Tangoscope*, a tangible audio output device for tabletop interaction. As we augmented headphones with a visual marker-based identification mechanism, with the Tangoscope each user is provided with individual auditory content. To allow for an instruction-free and intuitive usage of the audio device we employed the metaphor of a real stethoscope. A first user study indicated the self-explaining use of the Tangoscope.

Keywords: Tabletop interaction, audio interface, individual audio content, natural user interface, multi-user interaction.

1 Introduction

Camera based Tabletop systems like “Microsoft Surface” or the “reactable” [1] allow users to interact with the surface directly by using their fingers and hands. Moreover, these systems enable users to interact with the interface through the manipulation of real objects, taking advantage of the natural physical affordances of these objects [2]. These systems are capable to process multiple simultaneous input and provide therefore a good foundation for multi-user applications [1, 3]. Since user interactions occur directly in screen space on the display, these systems are deemed to be easy-to-learn and intuitive. With large screen estates, several users can simultaneously explore multimedia contents (e.g., videos) displayed on the tabletop system. However, if users simultaneously explore different interaction objects for which audio is provided (e.g., auditory text or audio channels of videos), public audio can be very confusing and disturbing. This is because users are not only exposed to their requested auditory information, but also to those of other users. In order to avoid such interferences, thus, certain auxiliary mechanisms are required to convey private auditory content to individual users of a shared tabletop application [4].

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 427–434, 2011.

2 Private Auditory Information in Shared Tabletop Applications

One approach to convey private auditory information to each individual user, is to provide separate shielded speakers mounted in front of each user [5, 6]. Although being successful applications for collaborative music generation systems [5] or for the provision of auditory feedback [6], several individual speakers conveying different auditory contents simultaneously might still be disturbing for the exploration of complex multimedia materials as, for example, in a museum setting. Furthermore, users cannot easily listen in to audio content of other users or move freely around the installation.

Morris et al. [4] introduced a system that uses earbuds in a single ear to convey private auditory information (i.e., mp3 songs) to specific users via individual sound channels. By using earbuds users are only exposed to the song they have currently requested. This system, however, employs a DiamondTouch [7] table that allows user identification. Each user is sitting on a custom chair, through which he/she is capacitively coupled to the table, thereby allowing the identification of touch events of a certain user. As a consequence, users are not able to move around freely during tabletop interaction, but have to stay seated on their assigned chair.

With common camera-based tabletop systems, user identification as provided by the DiamondTouch System is not possible, as they merely track the various contact points on the tabletop surface by objects such as fingers (e.g., Microsoft Surface) or visual tags. Therefore, in order to determine the receiver of certain audio content, such vision-based interactive tabletops must be augmented with additional identification mechanisms. This paper describes an approach to convey private auditory information to specific tabletop users for systems that cannot identify the corresponding user of each touch event. We propose a tangible audio device with a visual marker to assign audio requests to a connected headphone and thereby to the designated user. Moreover, in order to make the interaction with the tangible audio device as self-explaining and intuitive as possible, we tried to use a suitable metaphor for audio retrieval, that is, a stethoscope. We chose this metaphor, because the use and function of a stethoscope were assumed to be well-known to everyone, thus enabling users to reuse existing skills for interaction with the new device.

The following section describes our interactive multi-touch tabletop system that aims at providing users with private auditory information through the use of *Tangoscope* – an audio output device built from a real stethoscope (Fig. 1a left and middle). Furthermore, we present a user study which tests the intuitive understanding and use of the Tangoscope by comparing it to a technically equivalent “earbuds-tangible audio output device” which is not conveying the stethoscope metaphor (Fig. 1a right).

3 System Description

3.1 Tabletop Setup

We implemented a prototype of the audio device on a custom built standing-height tabletop setup with a display size of 1.2x1.35 m. The image is back-projected by two tiled Full-HD projectors allowing shadow-free user-interaction at a resolution of 1920x2160 pixels. The capture of user-input in terms of touch and tangible interaction is provided by a vision-based approach where the surface is illuminated with infrared light which gets diffused by the projection layer. Objects and fingers placed upon the table are captured from below by an infrared sensitive camera with a resolution of 864x768 pixels at 120 Hz. With the large size and high resolution of the display and the simultaneous interaction possibilities, the system is well suited for multiple simultaneous users. To extract the touch and visual marker information the camera image is processed similar to [8] utilizing the graphics processing unit to provide low feedback latency. The tracking software is able to differentiate between touch input and visual markers as well as between different markers. Like common vision-based setups, the system, however, is not able to distinguish between different users by touch input.

Multiple different audio channels on a single computer require either specialized audio hardware or multiple regular sound cards. Our implementation utilizes four external USB sound cards with each of them being connected via cable to one audio output device. To create working prototypes of the Tangoscope that match the desired metaphor as close as possible we altered real stethoscopes by replacing the regular earplugs with in-ear headphones and attaching a unique visual marker to the membrane of the head (Fig. 1 b).

Fig. 1. (a) Original stethoscope (left), *Tangoscope* (middle) and earbuds with marker only (right); **(b)** Unique visual markers attached to stethoscope heads

3.2 Interaction Techniques for the Tangoscope

Audio in Scatter-View Applications. Scatter-view specifies a typical presentation mode of a collection of visual material on tabletop systems. The items are presented like cards lying on a table and can be transformed with multi-touch gestures. Though applications usually allow the simultaneous playback of multiple video items, the audio channels interfere with each other. By placing the stethoscope head upon an item, the application can now stream the desired audio channel to the respective headphones, while other videos can continue playing for visual comparison. Users can also switch audio channels by easily swapping the stethoscope head to another item.

The rotation of the head can be additionally used to control other audio parameters such as volume or playback speed.

Spatial Exploration of Virtual Audio Sources. Beside the possibility of starting and stopping single audio channels, the application can also utilize the position of the stethoscope head upon visual material to allow the exploration of spatially distributed audio material. When moving the head between two sources, the audio seamlessly blends from one source to the other. Combined with multi-touch gestures for transformation, the user can spread the virtual space to listen more isolated to a specific audio source or shrink the material to hear a mix of the different sources.

Multi-User Applications. The described applications easily scale with multiple simultaneous users. Each user is then presented his/her private audio channel depending on the position of his/her stethoscope head identified by his/her own unique visual marker (Fig. 2). Users can listen in to other users' content easily by placing the head of their stethoscope next to that of another user.

Fig. 2. Multiple users accessing individual auditory content

3.3 Application Scenarios

Even though ownership of each touch point is not provided by the Tangoscope, audio assignment can be accomplished with this device on systems that cannot identify users. In the following we want to outline some tabletop applications where the Tangoscope can add further interaction possibilities with audio content.

Public Installations: Public speakers are usually not applicable for tabletop systems in public spaces like museums or libraries due to noise disturbance. Ordinary headphones, however, require an additional user identification mechanism. By using the Tangoscope, users can listen to individual audio content without disturbing other users or visitors. Furthermore, the Tangoscope allows for an instruction-free usage that is especially needed for interfaces and devices in public places.

Tabletop Games: Multi-player tabletop games like “Dungeons & Dragons” lack the possibility of private information for individual participants, as the whole screen is visible for all players. Here, the Tangoscope can be used to provide players with private information like mission briefings or status messages.

Group Tasks: For collaborative tasks on a tabletop system like [3] where co-located group members work in parallel, the Tangoscope can provide easy access to individual audio information without disturbing other users. Furthermore auditory content can be easily shared with arbitrary other group members who can listen in by placing the stethoscope head on the same content.

Collaborative Music Creation: In music creation applications like the reactable [1], the Tangoscope could provide monitoring functionality for the different participants. The device can be used further to listen to different mixtures by adjusting the volume of the instruments depending on the distance to the stethoscope head on the surface.

4 User Study

Stethoscopes are commonly known as medical devices which are usually used to monitor a person’s heartbeat. The Tangoscope incorporates this metaphor to provide an intuitive approach to individual audio content for tabletop interaction. In a first user study we therefore addressed the question whether users are able to transfer without any further instructions the familiar functionality of a real stethoscope from a medical scenario to audio retrieval on a tabletop system. Since the benefit of individual audio channels in a multi-user scenario was already studied by [4] we focused on the value of the stethoscope metaphor. To create a technically comparable control condition we attached a box with a visual marker to in-ear headphones similar to the ones used for the Tangoscope (Fig. 1a). This device provides indeed the same interaction possibilities from a technical point of view, but as only difference to the Tangoscope omits the stethoscope metaphor.

4.1 Methods

Across two groups we contrasted the use of the Tangoscope against the earbuds-tangible device, resulting in a between-subjects experimental design. Twenty doctoral students and staff from a German research institute (average age: 28.9 years; 11 female) volunteered for this study and were randomly assigned to either the Tangoscope or the earbuds-tangible device condition (10 per condition). None of the participants had prior experience with the audio devices used in this study. Twelve participants had limited prior experiences with interactive tabletops.

Twenty music videos of the current MTV charts, were presented as still images on the tabletop surface (Fig. 2). Videos could be started by placing the visual marker upon one of the images. During video playback the corresponding audio was provided through the headphones of the Tangoscope or the standard earbuds, respectively. If a user removed the tangible object from the surface, video and audio were stopped. Furthermore, users could move, rotate, and scale the video-items with multi-touch gestures. Time-stamped records of all user interactions with the interactive tabletop were logged by the application. Furthermore, a video camera was used to record participants during task processing.

Participants' task was to explore the music videos on the tabletop and to select their personal top 5 videos. In order to analyze their spontaneous exploration of the system, they were neither explained how the videos could be played, nor were they told about the existence and usage of the audio device. Tabletop interaction time was limited to five minutes per participant. The experiment was conducted in a laboratory setting, with participants being tested in individual sessions. After having received task instructions, participants were brought to the laboratory. There they encountered the interactive tabletop with 20 images of the music videos distributed across the tabletop surface. Either the Tangoscope or the earbuds-tangible device was lying on the right-hand side of the tabletop. The experimenter, who stayed in the back of the room during task processing, did not give any comments or advice, with one exception: if participants did not consider the audio device at all in the first two and a half minutes, the experimenter told them that they could use the device. After five minutes participants' interaction with the tabletop was interrupted.

To examine participants' interaction behavior we measured the events related to the audio device, that is, the number of times video/audio was played, time to first playback, and the total duration of video/audio playback. In addition, video data allowed analyzing the time until the audio device was picked up by the participants and the time taken from pickup to the first playback. Finally, participants were asked to rate the ease-of-use with the audio device on a 5-point scale (5 = highly agree).

4.2 Results

To compare participants' interaction behaviors and subjective ratings between the Tangoscope condition and the earbuds-tangible device condition 2-tailed t-tests were conducted. Table 1 shows the mean values of the measures for both conditions. T-tests revealed several significant differences with regard to audio usage in the two conditions. During the five minutes of exploration, Tangoscope users played video/audio significantly more often than users in the earbuds-tangible device condition (earbuds users), $t(18) = -2.23$, $p = .04$. Moreover, time to first playback was significantly shorter for Tangoscope users than for earbuds users, $t(18) = 2.08$, $p = .05$. In contrast, users did not differ with regard to the total duration of video/audio playback, $t(18) = -1.48$, ns. Furthermore, video data showed no differences between conditions regarding the time needed until the audio device was picked up by the users, $t(18) = -1.33$, ns. Four Tangoscope users and six earbuds users received the hint that they were allowed to use the audio device, as they did not consider the audio device at all in the first two and a half minutes of exploration, $\chi^2(1) = 0.80$, ns. Once having picked up the audio device, however, it took Tangoscope users significantly shorter to play video/audio than earbuds users, $t(10.34) = 2.76$, $p = .02$. Finally, users in both condition considerably agreed that the audio device was easy to use ($t(18) = -1.59$, ns).

Table 1. Mean values of the measures for both conditions

	Tangoscope	**Ear-buds**
# audio playbacks	37.70	24.90
time to first playback (in s)	115.24	176.01
total duration of audio playback (in s)	99.89	73.15
time to first pickup (in s)	104.70	71.20
time to first playback after pickup (in s)	24.60	104.60
ease of use (1-5)	4.50	3.60

5 Conclusion and Future Work

Results of the study indicate the natural and intuitive use of the Tangoscope as an audio output device for tabletop interaction. As compared to earbuds users, Tangoscope users needed less time to the first playback and played video/audio significantly more often. Thus, even without any instructions on how to use the audio device, the Tangoscope unfolded its advantages in comparison to the technically fully equivalent earbuds-tangible device. In both conditions quite a few users, however, completely ignored the audio device until the experimenter told them that they were allowed to use the device. Yet, at least after this hint was given all Tangoscope users understood how to use the device almost immediately, whereas three earbuds users could not find out how to use the device until exploration time was over, not even after the experimenter additionally gave them a second advice to use the "box". Three earbuds users tried to use the "box" as a remote control pressing on the dots of the visual marker (Fig. 1) showing that users usually search for meaningful functions. Particularly, the short time Tangoscope users needed from pickup to the first playback mirrors that users are well able to map the meaning and usage of a stethoscope as a medical device to the scenario of tabletop interaction with digital media.

Since this first user study solely addressed the question whether the metaphor is suitable for an audio device for tabletop interaction, more effort needs to be put in comparing the Tangoscope to other technical solutions like directed speakers in combination with tangible objects or restricted workspaces. For future work we are also planning to explore the Tangoscope device in multi-user scenarios. We assume that in such scenarios users can easily switch between conversation and listening by hanging the device around their neck. Another interesting research question is whether the device could help visually impaired people to access information on tabletop installations by providing auditory feedback of the underlying digital content.

Our current implementation of the Tangoscope needs to be connected via cable to the tabletop system for audio transmission. To make the device more flexible for large tabletop systems and multiple simultaneous users, wireless operations of the headphones of the Tangoscope would be helpful. This might be accomplished by using Bluetooth-headphones which are coupled to the tabletop computer. However, we are not aware of an available Bluetooth stack which allows for multiple simultaneous audio output devices. Another approach to provide wireless operation could be realized by plugging the headphones of each Tangoscope into a mobile device which receives audio playback commands via Wi-Fi from the tabletop system.

The Tangoscope audio device provides a feasible solution to the problem of superimposed audio output for multi-user tabletop applications. By utilizing the metaphor of a stethoscope it allows for instruction-free usage and we believe it is therefore well suited for tabletop installations particularly in public space.

References

1. Jordà, S., Geiger, G., Alonso, M., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proc. TEI 2007, pp. 139–146. ACM Press, New York (2007)
2. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proc. SIGCHI 1997, pp. 234–241. ACM Press, New York (1997)
3. Morris, M.R., Fisher, D., and Wigdor, D. Search on Surfaces: Exploring the Potential of Interactive Tabletops for Collaborative Search Tasks. Information Processing and Management (Spring 2010) (in press)
4. Morris, M.R., Morris, D., Winograd, T.: Individual audio channels with single display groupware: effects on communication and task strategy. In: Proc. CSCW 2004, pp. 242–251. ACM Press, New York (2004)
5. Blaine, T., Perkis, T.: The Jam-O-Drum Interactive Music System: A Study in Interaction Design. In: Proc. DIS 2000, pp. 165–173. ACM Press, New York (2000)
6. Hancock, M.S., Shen, C., Forlines, C., Ryall, K.: Exploring non-speech auditory feedback at an interactive multi-user tabletop. In: Proc. GI 2005, pp. 41-50, CHCCS (2005)
7. Dietz, P., Leigh, D.: DiamondTouch: A Multi-User Touch Technology. In: Proc. UIST 2001, pp. 219–226. ACM Press, New York (2001)
8. Edelmann, J., Fleck, S., Schilling, A.: The DabR – A multitouch system for intuitive 3D scene navigation. In: 3DTV CON - The True Vision, pp. 1-4 (2009)

Supporting Social Protocols in Tabletop Interaction through Visual Cues

Mirko Fetter[1], Tom Gross[1], and Maxi Hucke[2]

[1] Human-Computer Interaction Group, University of Bamberg
{Mirko.Fetter,Tom.Gross}@uni-bamberg.de
[2] Darmstadt University of Technology
Maxi.Hucke@stock-homburg.de

Abstract. Multi-touch tabletops provide new means for co-located people to work together on a task by directly manipulating objects and tools on a single display in unison. Despite their benefits they also entail new challenges. One major concern is how to help users avoid conflicting actions. Previous work discusses if social protocols are sufficient to regulate coordination, and if policies are needed to enforce specific behaviours. Our study on different variants of a tabletop game shows that providing visual cues on ownership can help to follow social protocols and therefore reduce the need for policies.

I 1ntroduction

Multi-touch tabletops provide new means for co-located people to work together on a task by directly manipulating objects and tools on a single display in unison. Despite their benefits they also entail new challenges. One major concern is to help users avoiding conflicting actions. For example, users could accidentally manipulate others' documents, perform actions simultaneously that are incompatible with each other, or alter global preferences that influence other users. Since all users' actions compete for the same display space at the same time, solutions need to support users in minimising the number of conflicts while keeping effort for coordination low.

Previous work basically takes up two opposite positions for addressing conflicts: One states that social protocols [5]—socialised norms of good and polite behaviour—take effect automatically, and help to regulate or even avoid conflicts. The other position points out that social protocols are not sufficient and as a consequence imposed rules and policies have to be established.

An example for the first position is Greenberg and Marwood's work [3] on concurrency control in real-time distributed groupware. They argue that social protocols enable users to naturally mediate and coordinate interactions in order to minimise conflicts while working together with a groupware system. Also Tse et al. [8] noticed in two case studies with multiplayer games on multi-touch tabletops that natural social protocols regulated some aspects of the game (e.g., turn taking).

On the other hand Morris et al. [5] suggest coordination policies for groups using a shared tabletop display to reduce conflicts, assuming that social protocols are not sufficient. Such policies however come along with an administrative overhead for the users.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 435–442, 2011.

Besides their advantages both positions have their drawbacks for the user, whether it is the chance of accidental conflicts where social protocols cannot proactively come into action, or the burden of a coordination overhead in order to avoid conflicts by imposing policies. We suggest a balanced solution by providing visual cues to users in order to activate social protocols. In the following we discuss how visual cues can help to avoid conflicts on multi-touch tabletops. We report on an empirical study using visual cues on ownership in a multi-touch game to reduce conflict.

2 Concept: Visual Cues to Avoid Conflicts

Providing awareness information is a fundamental concept in order to support group activity in shared workspaces [2]. In co-located single display environments that allow direct and synchronous manipulation for several users, physical and social cues provide users with a basic level of awareness on others' actions and currently used artefacts and objects. However, as multi-touch tabletop applications allow people to easily switch between personal and group work [7], users are often not able to monitor all activities taking place when concentrating on a personal task. Further, people are sometimes not able to judge the impact of their action towards others and hence lack the necessary information allowing them to act without interfering with other people's activities. Basically these two aspects can be seen as causes for conflicts that arise in co-located work and where social protocols fall short of coming into action [3, 5].

We suggest that improving the awareness of users by providing visual cues is beneficial in activating social protocols and thus is helpful to reduce conflicts in coordination. It can reduce the need for more restrictive methods. In our example we assume that ownership is related to social protocols and norms since out of a commonly agreed moral concept of respecting ownership we can derive behaviour that is socially acceptable. Thus, when people are aware of ownership structures they will act according to these social protocols and not break them intentionally. We suppose that in a multi-touch tabletop environment perceived affordance of ownership can be supported through visual cues. In our case ownership is beneficially communicated via distinct borders of territories and colours of artefacts.

Therefore, we came up with the hypothesis that by providing visual cues that reveal ownership of digital artefacts people are triggered in holding on to social protocols, and thus show adequate behaviour in cooperative and competitive tabletop interaction.

3 The Study

We conducted a study based on a multi-touch tabletop game that uses basic visual cues in order to subconsciously activate social norms that are related to ownership. In this section, we describe the implemented game and multi-touch tabletop, the experimental setup and its execution, as well as the data analysis with its results.

3.1 The PuhBox Game on the cueTable

The PuhBox game is a multi-touch game for four players split into two teams and is loosely based on the classic arcade game Pong by Atari. The goal of this cooperative-competitive game is trying to score a goal while hindering the opponent to shoot a ball behind the own goal line. The difference to the classic version of the game is the fact that the PuhBox game instead of two paddles provides 16 small squared boxes, which are equally divided between the two teams by placing them close to the player's position when the game starts. Each player can freely move these boxes by dragging. Once the ball hits a box, it bounces back. So, by rearranging the boxes, the players are able to protect their goal line, as the ball cannot be manipulated directly.

The rectangular playing field consists of two goal lines on the opposite shorter sides of the table and two reflective borders on the longer sides. When a goal is scored the ball disappears from the playing field and reappears in the scoring teams' playing field, as this team is next to serve. To serve, the ball is hit with a box and gets an impulse. At the beginning of the game, the serving team is chosen randomly. Each team starts with eight boxes spread over their side of the playing field. The game ends when the first team reaches 21 points. The current score is presented in each team's field near the centre of the playing field.

Every eight seconds one randomly chosen box slowly dissolves at its current location and reappears in the centre. This behaviour is called teleporting. The box indicates teleporting by fading out over a few seconds until it completely disappears. The fading mechanism helps the players to anticipate the teleporting and to take action (e.g., by rearranging the boxes). Further, the time span of fading out was implemented to help each player recollecting from which half of the playing field the box was teleported when reappearing in the middle. Teleporting was introduced to make the PuhBox game more dynamic.

The PuhBox game was developed and implemented for the PuhBox hard- and software [4] and runs on our own low-cost implementation of a FTIR multi-touch table. The PuhBox Framework is written in Java and encapsulates the low-level image processing, from capturing over filtering to blob tracking and finally gesture recognition, from the application layer by an event-based architecture. The PuhBox game leverages the PuhBox Framework.

3.2 Experimental Setup and Execution

In the following we give insight into the experimental setup and its execution. To examine our hypothesis we implemented three variants of the PuhBox: No Colour (NOC), Background Coloured (BGC), and Box Coloured (BOC) (cf. Figure 1). In the NOC variant no visual cues are given to the players. The screen is plain black, representing the whole playing field. In addition, all 16 boxes are coloured blue and there is no middle line. In the BGC variant visual cues regarding ownership are implemented in form of distinct borders of territories. A line surrounds both halves; the left side is red, the right side is green. In the BOC variant, additionally visual cues regarding artefact ownership are implemented through distinct colouring of the boxes, where half of the boxes are coloured in red, the other half in green.

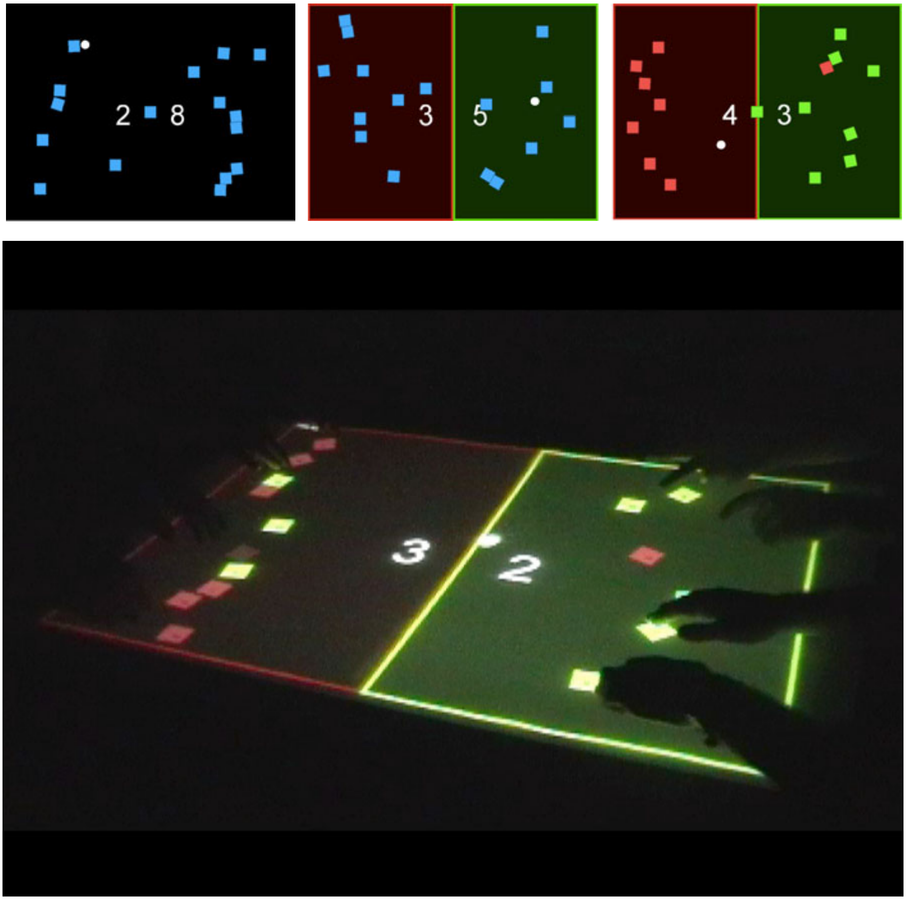


Fig. 1. PuhBox variants No Colour (NOC), Background Coloured (BGC), and Box Coloured (BOC); participants playing the BOC variant

The study was carried out in seven days. The subjects (*N*=64) were between 18 and 48 years (*M*=25.6, *SD*=4.9), with 19 females and 45 males. Most of them had never used a multi-touch table; none of them had played or seen PuhBox before. For each trial a group of four subjects formed two teams. The assignment to each game variant occurred randomly resulting in 16 trials split up between the three variants (5 NOC, 5 BGC, and 6 BOC). In each trial the teams played two matches of the same variant.

The trials took place in a separate room. An overhead video camera recorded the participants' interaction on the table, conversations and utterances. Interaction logs recorded the position of all 16 boxes every five seconds; the logging interval has been proven beforehand to be adequate for recording the relevant interaction. Finally, all subjects filled out a questionnaire after the two matches, addressing the perception of visual cues and fairness ratings, as well as personal details like age, gender.

At the beginning the two teams positioned themselves on the opposite shorter edges of the PuhBox. Before the two recorded matches started the participants were instructed how to interact with the multi-touch table. A trial run allowed them to get a feeling for the interactions, while goal and basic handling of the game were explained. No specific instructions were given regarding territories and the ownership of boxes in all three variants. This condition was fixed over all three variants in order to see how the visual cues are able to evoke the impression of ownership structures and accordingly lead to an adaption of the behaviour.

3.3 Data Analysis and Results

In the following we describe our data analysis as well as the results regarding interaction logs and questionnaire data.

Interaction Logs

As mentioned above the interaction logs captured the position of all 16 boxes every five seconds during a match. From the log data we calculated parameters to explore our hypothesis. Since each match is individual in length all interaction log variables were standardised. The transfer density for steals TD_S and for recaptures TD_R were calculated as a ratio of match length and the sum of steals (boxes taken away from their home side) and recaptures (boxes taken back to their home side) within each match. The number of boxes on the home side *NH* was standardised by downsampling the number of measuring points of the interaction logs to the shortest match (25 points within 120 sec) and downsampling the values of all other matches to 25 measuring points. By taking the time a box spent on its home side and calculating the proportion of being home considering the length of a match we computed the percentage of time on home side *PH* for each box. Averages for *NH* and *PH* over all boxes were calculated per match.

In a first step, we applied an ANOVA with repeated measures to detect possible learning effects between the two matches and to check whether these differences interact with the game variant. Since there were no significant within-subjects effects of match number and no interaction effect of match number by variant of the game we computed average values for all interaction log variables from both matches.

In a second step, we conducted an ANOVA to identify group differences based on the three game variants. Looking at the parameters (cf. Table 1) there were no significant differences in TD_S [$F(2,13) = 1.886$, $p = .191$] and TD_R [$F(2,13) = 1.792$, $p = .205$], but a strong tendency that interaction in the NOC variant is characterised by a greater density than in both other conditions. Steals and recaptures of boxes seem to be more frequent when no visual cues on ownership are given. Significant differences between the three game variants could be found both for *NH* [$F(2,13) = 9.521$, $p < .01$] and *PH* [$F(2,13) = 9.607$, $p < .01$]. The Sheffe post hoc test revealed that the game interaction in the BOC game variant shows higher averages compared to both other conditions. So overall, more boxes are kept on their home side for a longer period of time when visual cues like distinct borders of territories and distinct colouring of artefacts convey the impression of ownership.

Table 1. Means (SD) for each game variant

Interaction logs	Variant of the game		
	NOC	BGC	BOC
Transfer density of steals TD_S	10.97 (2.7)	17.64 (4.9)	20.07 (11.7)
Transfer density of recaptures TD_R	17.52 (6.2)	29.82 (10.8)	29.80 (15.9)
Number of boxes on home side NH	9.54 (0.9)	10.42 (0.6)	12.62 (1.7)
Percentage of time on home side PH	60.23 (5.8)	66.35 (3.3)	80.44 (11.2)

Questionnaire

Our questionnaire captured subjective assessments of the matches concerning the perception of the visual cues and judgement of fairness. Since the values of our interaction analysis are based on a group level we computed average values and cumulative values from the four members of each group. For the analysis of perceived visual cues we computed the sum of the answers whether the players had perceived their game field halves as their territories with exclusive access P_T (0=no, 1=yes) and the average of the impression whether boxes (i.e. artefacts) belonged to one team P_A (1= not at all, 4=totally). We also computed the average of fairness ratings of the other team F_O as well as of one's own fairness F_S. We conducted an ANOVA to see whether the groups perceived the visual cues differently between the game variants and whether the groups report differently about socially accepted behaviour in terms of fairness. Looking at the perception of the visual cues (cf. Table 2) we found significant differences between the three game variants only for P_A [$F(2,13) = 5.763$; $p < .05$]. A Sheffe post hoc test revealed that players in the BOC condition have a stronger impression than players in the NOC condition. For P_T we can only describe a trend between the game variants, which is in the expected ranking order (BOC ~ BGC > NOC). Looking at the subjective ratings of fairness (cf. Table 2) we did not find significant differences between the game variants, but above-average values in all groups as well as marginal higher ratings for the opponent's fairness than one's own.

Table 2. Means (SD) of the assessments of the matches for each game variant

Questionnaire Data	Variant of the game		
	NOC	BGC	BOC
Perception of territory ownership P_T	1.40 (1.1)	2.40 (1.8)	2.67 (1.6)
Perception of artefact ownership P_A	1.80 (0.5)	2.10 (0.9)	3.13 (0.6)
Fairness of opponent F_O	3.95 (0.9)	4.55 (0.8)	3.67 (0.8)
Fairness of oneself F_S	3.95 (0.5)	4.20 (0.5)	3.50 (1.0)

Further we examined the relationships between the subjective assessments of the matches and the interaction logs. Significant correlations ($p<.05$) between awareness information and interaction logs could be found only in the BOC condition where P_T shows a positive relationship to TD_S (r=.82) and TD_R (r=.82) meaning that a stronger perception goes along with greater intervals stealing and recapturing boxes and conversely. Also, we found significant correlations between F_O and interaction logs only in the BOC condition. Significant positive relations (p<.05) can be shown for *NH* (r=.81), *PH* (r=.82) and TD_S (r=.82), meaning that teams rating the opponent as more fair interact in a way that the number of boxes and their time spent on the home side are higher, and the intervals between stealing boxes are greater and vice versa. The same tendency was found for the TD_R (r=.79, p<.10). In contrast, in all three game variants only tendencies ($\alpha < .10$) regarding the relation between F_S and interaction logs were found and they differ between game variants. F_S in the BOC condition show a positive correlation with TD_S (r=.77) and TD_R (r=.76), meaning that teams rating their own fairness as high also showed greater intervals in stealing and recapturing boxes and conversely. Whereas F_S in the NOC and BGC conditions show a positive correlation with *NH* (NOC: r=.82; BGC: r=.83), meaning that teams rating their own fairness high also kept the number of boxes on their home side high.

4 Conclusions

Our study shows that providing visual cues can support users to act according to social protocols. Although the instructions for the game did not change over the three game variant, the interaction behaviour of the participant changed with different visual cues that lead to the perception of ownership. Thus, visual cues like boundaries and colours in our study had an impact on the actual behaviour of the participants.

As in previous work [6] on coordination techniques on tabletop groupware we showed that a controlled experiment in the form of a game can show effects of coordination techniques. The usage of a game supports the immediate involvement of the participants in the experiment in opposite to a more abstract task. However, this comes with the trade-off of reducing the applicability of this technique in a more natural real world task.

Nevertheless, our suggestion is to give users visual cues to support the natural regulation of interactions through the subliminal activation of social protocols. As a consequence users demonstrate behaviour, which can commonly be perceived as socially acceptable. As a side effect this kind of visual cues also supports users to easily uncover a deviation from social protocols. The chance of being observed could further lead to a reduction of unaccepted behaviour. In our case, the perception of ownership was always mutual for both teams over the different variants. Hence, the fairness rating over the three variants did not vary that greatly. In the variants where the perception of ownership was low, taking boxes from the other side was not perceived as wrong from both teams; in the variants were the perception of ownership was higher, taking boxes from the other side was perceived as improper behaviour from both teams and accordingly rather rare. But when a infringement took place, the teams could much better identify that territorial or artefact ownership was ignored and react with protest.

Generally visual cues can help to reduce the social-technical gap [1] instead of widening it with imposed policies and locking mechanisms. Visual cues that reveal underlying information are a simple but fruitful way to reduce accidental infringements against social norms while preserving a high degree of freedom for all users. Accordingly, applying coordination policies should be considered as a last step when all possibilities for improving interaction through providing additional information to the users have been exhausted.

Acknowledgments. We thank the members of the Cooperative Media Lab; special thanks to Sascha Liebsch, the participants of the study and the anonymous reviewers.

References

[1] Ackerman, M.S.: The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. International Journal of Human-Computer Interaction 15(2&3), 179–203 (2000)

[2] Dourish, P., Bellotti, V.: Awareness and Coordination in Shared Workspaces. In: Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work - CSCW 1992, Toronto, Ontario, Canada, October 31 - November 4, pp. 107–114. ACM Press, New York (1992)

[3] Greenberg, S., Marwood, D.: Real Time Groupware as a Distributed System: Concurrency Control and its Effect on the Interface. In: Proceedings of the 1994 ACM Conference on Computer-Supported Cooperative Work - CSCW 1994, Chapel Hill, North Carolina, United States, October 22 - 26, pp. 207–217. ACM Press, New York (1994)

[4] Gross, T., Fetter, M., Liebsch, S.: The cueTable: Cooperative and Competitive Multi-Touch Interaction on a Tabletop. In: Extended Abstracts of the Conference on Human Factors in Computing Systems - CHI 2008, Florence, Italy, April 5-10, pp. 3465–3470. ACM Press, New York (2008)

[5] Morris, M.R., Ryall, K., Shen, C., Forlines, C., Vernier, F.: Beyond "Social Protocols": Multi-User Coordination Policies for Co-located Groupware. In: Proceedings of the 2004 ACM Conference on Computer-Supported Cooperative Work - CSCW 2004, Chicago, Illinois, USA, November 6-10, pp. 262–265. ACM Press, New York (2004)

[6] Pinelle, D., Barjawi, M., Nacenta, M., Mandryk, R.: An Evaluation of Coordination Techniques for Protecting Objects and Territories in Tabletop Groupware. In: Proceedings of the Conference on Human Factors in Computing Systems - CHI 2009, Boston, MA, USA, April 4-9, pp. 2129–2138. ACM Press, New York (2009)

[7] Scott, S.D., Grant, K.D., Mandryk, R.L.: System Guidelines for Co-located, Collaborative Work on a Tabletop Display. In: Proceedings of the Eigth European Conference on Computer-Supported Cooperative Work - ECSCW 2003, Helsinki, Finland, September 14-18, pp. 159–178. Kluwer Academic Publishers, Netherlands (2003)

[8] Tse, E., Greenberg, S., Shen, C., Forlines, C.: Multimodal Multiplayer Tabletop Gaming. ACM Computers in Entertainment 5(12) Article No.12 (August 2007)

Effects of a Tabletop Interface on the Co-construction of Concept Maps

Stefan Oppl and Chris Stary

Department of Business Information Systems – Communications Engineering
Kepler University of Linz
Freistädterstrasse 315, 4040 Linz, Austria
{stefan.oppl,christian.stary}@jku.at

Abstract. Concept Mapping is a method for externalizing and reflecting knowledge about real world phenomena. In cooperative settings, concept maps can also be used to aid cooperative learning activities and the development of a common understanding about the mapping subject. This process imposes requirements on tool support that have hardly been addressed in existing concept mapping tools. We present a tabletop interface designed to meet these requirements. In an empirical study, the positive effects on the cooperative mapping process facilitated by the proposed system have been shown in comparison to a traditional, screen-based system.

Keywords: Concept mapping, tangible interface, tabletop, cooperation, alignment of meaning, sense making.

1 Introduction

Concept maps are established means to organize and represent information. They can be used to support the process of eliciting, structuring, and sharing knowledge. According to their objective, to enable meaningful learning (cf. [1]), most of the applications of concept maps can be found in education. Concept maps, besides other means to represent information, use concepts as entity to structure items of interest. A concept can be a central term, an expressions or a metaphor. It represents a coherent unit of information for the person using it. In the course of mapping, concepts are put into mutual context by specifying concrete relationship, leading to a network of concepts, i.e. the concept map.

Based on the findings related to MiMs (Montessori-inspired Manipulatives) grasping information using tangible user interfaces could leverage the way to develop intelligence (cf. [2], [3]). In our research we target the social and cognitive dimensions of building capacities. We aim to support cooperative concept mapping using a tangible tabletop interface. To do so, we have to examine the fundamentals of concept mapping and its usefulness in different stages of learning (section 2). When reviewing existing tools, lack of support for the specifics of cooperative concept mapping can be identified (section 3). Our tabletop concept mapping system has been designed and evaluated to overcome this deficiency (section 4 to 6). The contribution

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 443–460, 2011.

of this paper can be found in the detailed empirical examination of the effects the setting using the tangible tabletop system has on both, the individual's activities in cooperative mapping process and the outcome of the mapping process for the involved individuals.

2 Concept Mapping and Its Effects on Learning

Persons express their items of interest and the relationships by means of language constructs. As language allows expressing human thoughts, cognitive structures and the network of concepts explicitly representing semantics seem to correspond in one way or another [3] [4].

The strengths of concepts mapping lie in the focused, but still meaningful representation of domains of interests. According to the (subsumption) theory of Ausubel [1] it is the type of structure (i.e. a network of nodes and relationships) allowing learners to capture large amounts of relevant information from textual and verbal inputs, assuming acquisition of new knowledge is dependent on what is already known.

2.1 Acquisition

As the prior knowledge of learners needs to be addressed effectively before entering learning processes, concept maps can serve as a means reflecting already acquired knowledge. According to Chang et al. [5] concept mapping as a graphic organizer can assist text learning through enhanced text comprehension and summarization abilities. When BouJadue et al. [6] have used concept maps as chemistry homework tools positive impact could be found with respect to organizing information, fostering metacognition, and engaging learners in building individual knowledge structures.

2.2 Co-construction

Sutherland et al. [7] have suggested concept mapping as a methodological catalyst for organizational learning. Concept mapping offers a way to simultaneously understand complex systems in terms of both intra- and interpersonal relationships. In their evaluation study, concept mapping has been used by two stakeholder groups along a process of structured conceptualization. In that case concept mapping has facilitated the development of a jointly authored conceptual framework to be used in future program planning, development, and evaluation. The revealed knowledge conversion has also been argued by Fischer et al. [8] in the context of collaborative learning, as concept maps allow to deal with homogeneity/heterogeneity of information items, such as learning prerequisites, in a constructive way.

Okada et al. [9] could show the benefits of using concept maps as sense-making representations of conversations. Hereby, concept maps have been used as visual thinking technologies to create knowledge structures about and for conversations mediated through videoconferencing. In the course of negotiating collaborative concept mapping has been supplemented with oral narratives. In this way, each participant could represent the collective sense-making process. The participants could negotiate individual meanings of a concept in a straightforward way.

3 Existing Support for Concept Mapping

There exists a variety of tools for concept mapping, among them the CMapTools [10], Inspiration, (www.inspiration.com), SemNet [11], Smart Ideas (www.smarttech.com), Visimap (www.visimap.com), and Thinkmap (www.thinkmap.com).

Focusing on the cooperative application of concept mapping for the co-construction of meaning, tool support has to provide both, equal access to the mapping surface, and means to explicitly support the cooperation-. Existing tools hardly address these issues. The CMapTools support spatially distributed asynchronous and synchronous co-construction of concept maps by providing each participant with an instance of the toolset running on dedicated computers, which can be connected via a network [10]. However, negotiation processes occurring in the course of co-construction processes are more effective when carried out in a face-to-face setting [12]. In such a setting, not only the content of the concept map (and an accompanying chat or forum, respectively) can be used to convey information, but also facilities like gesture and facial expression support the process of negotiating a common understanding of the learning object.

Traditional user interfaces based on screen, keyboard and mouse provide exclusive access to the input channels of the concept mapping support system. They hinder the co-construction-process by lacking equal access for all participants. They rather require for all inputs to pass the “filter” of the user currently using mouse and keyboard.

In mapping-environments not supported by computers, equal access is typically enabled by a mapping surface accessible to all, such as a table [13]. The use of a table provides immediacy of manipulation, which is attained by the physical creation of the model. Participants immediately refer to a physical representation rather than abstract items. They create and alter the model in a dialogue-based way until reaching consensus about what is represented.

Mental models of individuals are externalized, questioned and can be modified at the same time. However, due to the potential complexity of concept maps (especially for more complex mapping subjects), computer-support is considered crucial [14]. As an example, in the case of spontaneous idea elaborations relationships are not traced automatically when concepts are moved in traditional settings. In computer-supported settings this task can be handled by the system, thus allowing for spontaneous changes without affecting prior results. Rolling back follows recorded modifications, which ensures consistency.

Both, the requirements of equal access to the mapping surface and the need for computer-support can be met by tabletop interfaces. In this work, we present our approach for a tabletop interface to support cooperative concept mapping. Baraldi et al. [15], Tanenbaum & Antle [16], Do-Lenh et al. [17] and Maldonado et al. [18] have proposed similar approaches.

Baraldi et al. [15] have used gesture recognition on a multitouch table in order to support concept mapping. They however did not explicitly focus cooperative use cases. Tanenbaum & Antle [16] have also focused on the individual creation of concept maps using a tabletop system. While the concept map itself has been projected onto the table-surface, manipulation has been performed using tangible tokens and constraints, in particular for self-checks in knowledge-structuring and learning tasks. Maldonado et al. [18] present a tabletop interface for cooperative concept mapping. They use a multi-touch interface and a purely digital visualization of the concept map. Their evaluation, however, focused on a qualitative study of usability aspects and did not focus on the effects of the tabletop for the construction of the concept map. Do-Lenh et al. [17] have physically represented the concepts of the map for collaborative usage – users are able to interact with the system by directly manipulating the concept representations. Additional digital tools also accessible through the table surface are used to further augment the concept map (i.e. naming concepts using a virtual keyboard). Do-Lenh et al. have explicitly addressed and examined the tabletop's effects on collaboration. Their findings are put in the context of our work below (cf. section 'Related Studies').

4 Tangible Tabletop Concept Mapping

The system we present here is based on a tabletop interface. We use tangible tokens for content representation and interaction with the user (see Fig. 1). Physical tokens have been chosen for implementation in favor of a multi-touch system, because the physical effects of graspable representations facilitate knowledge externalization and communication (e.g. [12], [13]).

Users place tokens and associate them accordingly to form a concept map. The tokens act as carriers for concepts. All interactions between the users and the system occur on the surface to enable simultaneous manipulation of the model. The table surface and a secondary, traditional display are used for visual information output. The system has been built upon the ReacTIVision-Framework [19] for element tracking on the tabletop system.

For interaction the system provides a set of tokens available in different shapes and colors. Each shape represents an arbitrary concept type and has to be assigned by its users. The different token types allow for semantic structuring of the map by assigning a specific meaning to each type. To create associations between concepts, the representing tokens have to be shortly moved in close spatial vicinity (see Fig. 2). The participants of a mapping session name connections and concepts by using a wireless keyboard.

Fig. 1. System Overview, showing a simple procedural concept map on both, the table surface and the secondary display (computer screen)

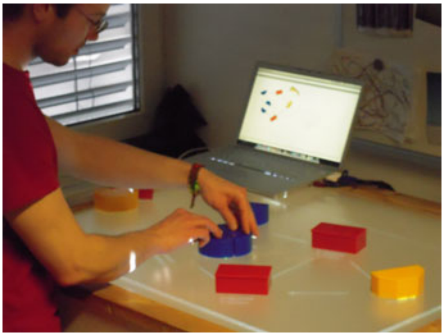


Fig. 2. Labeling & Associating

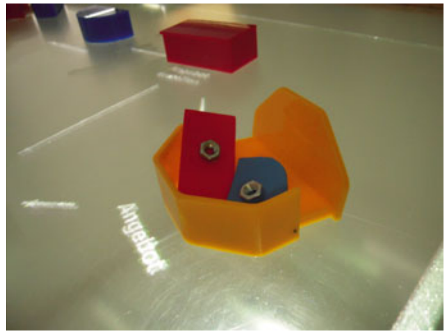


Fig. 3. Using Tokens as Containers

To overcome the restrictions of the limited size of the table surface and to allow for additional structuring, all tokens have been designed and constructed as containers (see Fig. 3). A (sub-)map can be saved to a embeddable token by placing the token in a designated area of the surface (“tray”). The (sub-)map then can be attached to a concept by putting the embeddable token briefly next to an open container token and simply placing it inside afterwards. Placing tokens inside a container significantly adds weight and also produces audible feedback when grasping the container. In this way, awareness about additional, currently hidden, information becomes available.

Support for exploration of knowledge domains is explicitly provided by history and reconstruction support. The system automatically tracks the creation history of the concept map and allows for navigation through the time line (similar to the concept of design history [20]). The reconstruction of former states of the map, as required in case of (experimental) changes, is supported by the system by displaying step-by-step directions on which concepts to remove, move or add to reestablish the former concept map. While undoing changes this way still cause more effort than in purely digital tools, our system is the first explicitly addressing the requirement of undoing changes in physical modeling environments, and providing an implementation to that respect.

The system has been designed to be controllable by several people simultaneously. People group around the table, which is about 110 cm in height. In order to foster interaction, activity and changing roles among the users, they need to stand upright rather than being seated. In the course of collaboration they position themselves where their current focus of interaction lies.

The active table surface is 100 x 80 cm in size and allows for placement of about 15 to 20 concept tokens at a time. The software subsystem is able to track and visualize an arbitrary number of interactions in real time. Even when in non-mapping-states (like history navigation or concept naming), the system keeps track of changes to the map on the table surface and updates its internal representation of the map accordingly. Changes can then be propagated to the output channels (table surface, secondary displays or attached remote displays) as soon as the live mapping visualization is active again.

5 Evaluation

We first detail the finding of hypotheses and the study design before reporting the results and outcome.

5.1 Hypotheses

To be able to outline the evaluation design, the metrics of "effectiveness" have to be specified. According to Okada et al. [9], a concept map should serve as a mediator for conversations. Sutherland et al. [7] argue to consider concept maps as a means to develop a common understanding of the represented problem and to integrate different viewpoints in a constructive way. As basic requirement the system must be open with respect to the concept mapping process itself, i.e. has to allow the creation of arbitrarily complex concept maps representing arbitrary content. If these three requirements can be met when using the presented system, it can be considered to support cooperative concept mapping effectively:

- H1: The tabletop system allows for the creation of arbitrarily complex concept maps.
- H2: The tabletop system enables cooperation among the participants in the mapping process.
- H2a: Using the tabletop system leads to a higher degree of cooperation during map construction compared to using a screen-based tool.
- H3: The tabletop system enables the development of a common understanding of the mapping topic by supporting the alignment of meaning.
- H3a: Development of a common understanding is facilitated when using the tabletop system in comparison to a screen-based system.

Hypothesis 1 addresses the fundamental requirement of being able to use the system for concept mapping. This hypothesis is also used to test, whether the feature to create hierarchical maps (using the container metaphor) is an adequate replacement of an unlimited modeling surface (as available on a screen-based mapping tool). Hypotheses 2 und 3 capture the specific requirements for cooperative learning and reflection processes. The latter are supplemented with an additional sub-hypothesis, in order to contrast our system with traditional, screen-based concept mapping tools. The goal of the sub-hypotheses is to test, whether tabletop systems allow for better, i.e. more effective, support for cooperative concept mapping than screen-based systems.

5.2 Related Studies

Do-Lenh et al. [17] have evaluated the usage of a tabletop concept mapping tool in cooperative settings. In an additional study [21], the same authors have also evaluated the learning outcome of collaborative tabletop concept mapping in comparison to the CMapTools as a representative of traditional, screen-based concept mapping tools.

Regarding the usage process, the authors have found out, that the tabletop system fosters a continuous flow of cooperation among the participants during the whole process of the creation of the concept map. Cooperative creation of a concept map using the CMapTools is characterized by alternating phases of largely individual content creation (i.e. having on participant taking the mapping initiative) and cooperative phases of reflection of the concept map.

Regarding the learning outcome, the authors were not able to justify the positive impact of a tabletop system on the learning outcome. On both, individual and group level, the participants using the screen-based system reached higher learning performance than the participants using the tabletop system. They however found a 'qualitatively better collaboration' in cases where the tabletop system had been used. They also argue that the physical setting could have influenced the intensity of collaboration, as the physical degrees of freedom might affect group dynamics. In this respect, their setting differs from ours, as people were sitting around the tabletop surface in their setup, whereas participants are standing upright around the tabletop in our case.

5.3 Study Design

Context. Two studies were carried out in the course of a lab on business process modeling. Students attending this lab learn to represent knowledge about business processes using different three process modeling languages. The goal of the lab is to recognize the fact that different scenarios of use for process models require different forms of representations. During the lab, students are introduced to three graphical modeling languages and have to apply each language by modeling a common business scenario. In their final task, the students have to collaboratively reflect on their experiences and contrast the modeling languages regarding their strengths and weaknesses for different scenarios of use. They document their findings using a concept map, which they co-create in the course of their reflection and discussion process (for a sample, see Fig. 4). This task was subject to both studies.

The scope of the first study was to examine the effects of the tabletop interface on the collaborative concept mapping process (contributing to the evaluation of H1, H2 and H3). 54 students of business information systems in the first part of their undergraduate studies participated in the study (46 male, 8 female). None of the participants had any prior knowledge in process modeling or concept mapping. To complete the concept mapping task, they were arbitrarily grouped into 18 teams of 2 or 3 persons, who arranged around the tabletop system.

In the second study, the tabletop concept mapping scenario was compared to a collaborative mapping setting using a traditional, screen-based system (mainly focusing to the evaluation of H2a and H3a but also contributing to H1, H2 and H3). 49 students in the same phase of their studies as described above participated in the study (40 male, 9 female). To complete the concept mapping task, they were arbitrarily grouped into 23 teams of 2 or 3 persons. 11 groups created the concept map using the tabletop system again arranging themselves around the table (standing, no seats), while 12 groups used the CMapTools toolset collectively sitting in front of a computer screen.

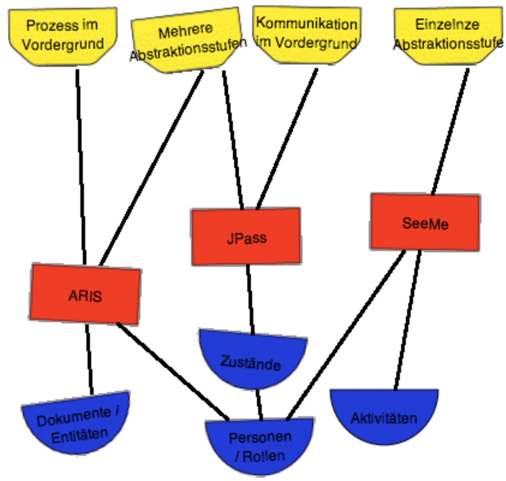


Fig. 4. Sample Concept Map (screen-based viewer of the tabletop system, concepts labeled in German, association labels have not been used by the modelers)

Measures. Data was collected during and after the mapping sessions from three sources: (1) Video recordings of the mapping process from two spatial perspectives – one focusing on the participants, the other covering the mapping surface in detail, (2) recordings of the evolution of the created concept maps (by the system), (3) post-mapping questionnaires on usability, usefulness and perceived outcome.

In total 7 quantitative and 5 qualitative measures were extracted from the collected data. In the following, we describe the measures and how they have been used to test the hypotheses:

Quantitative measures:

- Hypothesis 1 – Creation of Concept Maps:
  - *Size of concept map (S)* – measured by number of concepts used. S can be used as a measure for the complexity of the map.
  - *Speed of map creation (SC)* – measured in seconds per concept used. SC can be used as a measure for the system's ease of use.
  - *Number of concept types used (CT)* – calculated by counting the number of different semantic types of concepts used. CT can be used as a measure for amount of content structuring.
  - *Connectedness (C)* – measured in connections per concept used. Can be used as a measure for the complexity of the map.
- Hypothesis 2 – Cooperation:
  - *Distribution of activity time among participants (ATD)* – extracted from videos; considering not only physical initiative but activity in general (including discussion, negotiation, giving directions for mapping). Cooperative activity times are distributed equally among

the involved participants. ATD serves as a measure for involvement of the participants.

  - *Turn Taking of physical initiative (TPI)* – extracted from videos; focusing on physical initiative and counting how often participants switched tool control. Simultaneous physical activity *(SPA)* on the tabletop was recorded separately (not possible in the case of CMapTools). TPI can be used as a measure for the amount of interaction occurring among the participants.

- Hypothesis 3 – Development of common understanding:
  - *Fraction of mapping time used for discussion (FDT)* – extracted from videos. Complete mapping time was distributed among time used for map creation, time used for technical and administrative discussion and time used for discussion about the content of the map. The latter is used to calculate this measure. FTD is a measure of the amount of time dedicated to activities for the alignment of individual views.

Qualitative measures (extracted from questionnaires and videos):

- Hypothesis 1 – Creation of Concept Maps:
  - *Perceived system restrictions during the mapping process (pR)* – users were asked which aspects of the system they perceived to restrict the mapping process, i.e. not letting them express what they intended to externalize. pR not only is a measure of perceived restrictedness but also gives hints on how interaction with the system can be enhanced. For the redesign of interaction, the subjective findings were cross-checked with the video recordings showing the interaction with the system in order to identify the actual pitfalls and misunderstandings causing interaction to break.
- Hypothesis 2 – Co-creation:
  - *Perceived cooperation during the mapping process (pCP)* – users here described the amount and specifics of cooperation during the mapping process. This included explicit questions on implicit or explicit role taking and switching during the process as well as open questions on how well cooperation was perceived in general.
  - *Observed cooperation during the mapping process (oCP)* – extracted from videos. Cooperation was qualitatively assessed based upon the way individual contributions where integrated in the common concept map. It includes omitted contributions, misunderstandings or cooperative effort to create or alter content.
  - *Perceived ability to contribute the common concept map (pCO)* – this measure targets perceptions of the individual contribution to the mapping result. Users had to explicitly describe in which aspect of the concept map (e.g. content, structuring, …) they felt to have contributed to most. The results of this measure have been cross-checked with the result of pCP, as many users described the roles and contributions of their fellow group members there.

- Hypothesis 3 – Development of common understanding:
  - *Perceived impact of mapping process and result on individual understanding (pIU)* – users were asked, whether and to which extent the mapping process and the mapping result contributed to their individual understanding of the topic the concept map was created on. The questions were formulated in a way that encouraged users to describe, in which aspects or areas their understanding changed. pIU thus requires introspection, reflecting on and externalizing one's own cognitive structures.
  - *Perceived impact of mapping process and result on development of a common understanding (pCU)* – in addition to their perception of changes in their individual understanding, users were asked whether they found the mapping process and/or result helpful for the development of a common understanding about the mapping topic across all participants. Again, users were asked to describe, where they perceived changes and based on which perceptions they identified these changes.

Along with pIU, data for pCU was gathered not only directly after the mapping session but also for a second time about 8 weeks after the session to examine long-term impact of the concept mapping session.

Measures based on *perceived understanding* might appear to be inferior when compared with "objective" measures e.g., based on the expert interpretation of mapping results. In the following, we describe why our approach to measure mental model alignment via individual perception nevertheless is valid. Changes in individual cognitive structures (which are a consequence of cooperatively creating a concept map), however, cannot be measured directly, but only via additional externalized structures (such as, again, concept maps) [22]. Diagnosing changes requires repeatedly letting people externalize their views about the same phenomena. This creates methodical problems such as better results in later sessions due to learning effects. These problems can only be addressed with certain quantitative statistical approaches (such as "Hierarchical Linear Modeling" [22]). Measuring "alignment" would, however, require being able to not only quantify changes but also gain qualitative data about how the cognitive structures have changed. Currently available approaches, that are methodologically sound and correctly deal with the challenges of change diagnosis, do not provide qualitative data.

Although considered suboptimal when examining changes in cognitive structures [22], subjective measures still appear to be the most suitable instrument when judging cognitive structure alignment, as they at least provide the required qualitative data. The use of subjective measures, however, is a limitation of this study, as reliability of the collected data cannot be judged for these measures.

Testing. Hypothesis 1 is tested using the measures *Size of concept map (S)*, *Speed of map creation (SC)*, *Number of concept types used (CT)*, *Connectedness (C)* and *Perceived system restrictions during the mapping process (pR)*. S, CT and C are calculated for maps created with the tabletop and the CMapTools. As the CMapTools have already turned out to support Concept Mapping (cf. Canas et al., 2004), similar results for identical mapping tasks indicate appropriate mapping support by the tabletop. pR provides additional hints on potential restrictions on concept mapping imposed by the tabletop system.

Hypotheses 2 and 2a are tested using the measures *Distribution of activity time among participants (ATD)*, *Turn Taking of physical initiative (TPI)*, *Perceived cooperation during the mapping process (pCP)*, *Observed cooperation during the mapping process (oCP)* and *Perceived ability to contribute the common concept map (pCO)*. All measures are calculated for mapping sessions using the tabletop system and the CMapTools, respectively. Both ATD and TPI are measures for the involvement of the participants. Similar individual ATD values indicate a similar amount of participation in the mapping processes. TPI and its sub measure SPA directly indicate the strength of cooperation during the mapping process. pCP, oCP and pCO give hints on cooperation support based on subjective perceptions of the participants.

Hypotheses 3 and 3a are tested using Fraction of mapping time used for discussion (FDT), Perceived impact of mapping result on individual understanding (pIU) and Perceived impact of mapping result on development of a common understanding (pCU). FDT shows the fraction of time dedicated to discussing the content of the concept map. Discussion of mapping content is an integral part of the development of a common understanding. High values for FDT thus indicate potential for the development of a common understanding. The actual impact of system usage is only measured qualitatively using pIU and pCU. Both measures are again based on the subjective perceptions of the participants.

6 Results

Size (S): The average size of a tabletop concept map considering both studies is 14.68 concepts ($SD = 7.9$, $n = 29$). In the second study, the tabletop concept maps ($M = 12.18$, $SD = 8.84$, $n = 11$) were significantly smaller than those created with CMapTools ($M = 25.92$, $SD = 7.04$, $n = 12$) for the same mapping task (one-sided Wilcoxon-test for unpaired samples: $W = 8$, $p < 0.001$).

Speed of Creation (SC): The average time used to place or alter a specific concept was calculated by dividing the overall modeling time (total time minus time for discussion and technical or administrative issues) through the total number of placed concepts. On the tabletop (considering both studies) the average time per concept was 40.3 s ($SD = 14.2$, $n = 29$). In the second study, the speed of creation for tabletop concept maps ($M = 32.5$ s, $SD = 11.26$ s, $n = 11$) was significantly faster than for those create with CMapTools ($M = 48.12$ s, $SD = 17.91$ s, $n = 12$) for the same mapping task (one-sided Wilcoxon-test for unpaired samples: $W = 83$, $p = 0.021$).

Concept Types (CT): The tabletop basically only provides 3 different types of tokens. However, using more than 3 concept types was possible, once sub models with different semantics were embedded. Considering both studies, the average number of concept types used on the tabletop was 3.14 ($SD = 1.04$, $n = 29$). In the second study, the number of concept types of tabletop concept maps ($M = 3.64$, $SD =$

1.50, n = 11) did not significantly differ from that of maps create with CMapTools (M = 3.50, SD = 2.11, n = 12) for the same mapping task (two-sided Wilcoxon-test for unpaired samples: W = 75, p = 0.587).

Connectedness (C): Connectedness is calculated by dividing the overall number of connections through the overall number of concepts. On the tabletop (considering both studies) the average connectedness was 0.96 (SD = 0.41, n = 29). In the second study, the connectedness of tabletop concept maps (M = 1.09, SD = 0.36, n = 11) did not significantly differ from the connectedness of maps create with CMapTools (M = 1.23, SD = 0.20, n = 12) for the same mapping task (two-sided Wilcoxon-test for unpaired samples: W = 64.5, p = 0.476).

Distribution of Activity Time (ATD): On the tabletop (considering both studies), distribution of activity time for groups of three people (n = 19) was 49.3 % for participant A, 33.1 % for participant B and 17.6 % for participant C (participants were sorted in decreasing order from A to C). For groups of 2 persons (n = 10) was 60.4 % (SD = 12.2 %) for participant A and 39.6 % (SD = 12.2 %) for participant B.

In the second study, only groups of two persons have been analyzed for reasons of sample size (CMapTools: 11 samples for 2 persons, 1 sample for 3 persons, Tabletop: 9 samples for 2 persons, 2 samples for 3 persons). Distribution of activity time (percentual values for participant A and participant B, where A is the person with higher activity) on the tabletop system (M: 57.5 % - 42.5 %, SD: 11.3 % - 10.4 %, n = 9) did not significantly differ from the distribution of activity time for groups using CMapTools (M = 56.6 % - 43.4 %, SD = 7.4 % - 9.6 % , n = 11) for the same mapping task (two-sided Wilcoxon-test for unpaired samples calculated on the percentual differences of the activity time of both participants: W = 51, p = 0.86).

Physical Turn Taking (TPI): In the 29 sessions carried out on the tabletop, 17 cases have been identified (SPA), in which no explicit physical turn taking was carried out, as at least two participants were working simultaneously on the tabletop. In an average mapping time of 16 minutes and 40 seconds (SD = 10m 1s), 2 turn taking activities have been identified in 6 of the remaining sessions, another 6 sessions did not show any physical turn taking activities. In the second study, 7 cases of simultaneous physical activity have been identified on the tabletop (n = 11), while no such case was present for CMapTools (n = 12). Exclusive physical access to the mapping controls by one person (i.e. no turn taking at all) was identified in 2 cases on the tabletop (n = 11) and in 8 cases for CMapTools (n = 12). On the tabletop, the remaining 2 cases both showed 2 turn taking activities (in an average mapping time of 7m 7s, turntaking on after 3m 33s). For the CMapTools, 2 turn taking activities were identified in two cases, 4 and 8 turn taking activities were identified one time each (in an average mapping time of 19m 22s, turn taking on average happens after 4m 42s).

Discussion Time (FDT): On the tabletop (considering both studies) the average fraction of discussion time per concept was 73.13 % (SD = 6.43 %, n = 29). In the second study, the fraction of discussion time of tabletop concept maps (M = 76.08 %, SD = 8.84 %, n = 11) was significantly higher than that of maps create with

CMapTools ($M = 57.15$ %, $SD = 7.49$ %, $n = 12$) for the same mapping task (one-sided Wilcoxon-test for unpaired samples: $W = 120$, $p < 0.001$).

In the following we discuss the results of the **qualitative measurement** with respect to the hypotheses. For further results the readers are referred to [23].

6.1 Discussion

The results of hypotheses testing described in the section above require some further discussion and interpretation, in particular when putting evaluation results in mutual context.

H1 – The tabletop does not allow for the creation of arbitrarily complex concept maps: Hypothesis 1 had to be rejected – the tabletop system does not allow the creation of arbitrary concept maps. Compared to maps created with the CMapTools, the size of the maps was significantly smaller (Measure S). Measure pR also showed that the two major restrictions of the tabletop system were perceived to be the too small size of the mapping surface and the too big size of the mapping elements, respectively. The amount of concept types used did not significantly differ from maps created with the CMapTools (Measure CT). The amount of concept types limited to three (due to the current hardware implementation) was perceived as a problem by some of the users. For both, model size and semantics, the option to embed sub models does not seem to be an appropriate alternative to arbitrarily sized maps created on a single layer.

From a usability point of view, however, the tabletop does not slow down the creation of a concept map. The speed of creation actually was significantly faster for tabletop maps (Measure SC). Moreover, values for Connectedness did not show any significant difference (Measure C). Results of the measure pR also confirm that the system is considered usable from a technical point of view (except certain features of the original version, which caused confusion due to misleading UI metaphors. These features underwent a redesign based on the initial findings and were not perceived to be incomprehensible after redesign – for a more detailed elaboration on this aspect, cf. [24]).

Overall, H1 had to be rejected because of the design decision to use tangible elements for content representation. If the requirement of providing a concept mapping environment with capabilities equivalent to screen-based tools is a central one (which in the case presented here it was not), choosing a multi-touch surface with a digital representation of the concept map might be a more suitable option. It however remains unclear which effect a multi-touch-based system would have on cooperation and the alignment of individual understanding – from a conceptual point of view, physical representation should provide added value for these aspects (e.g. [12], [13]).

H2 – The tabletop enables cooperation among participants: Hypothesis 2 was verified – the tabletop system does support cooperative concept mapping. While the amount of activity time is not equally distributed (Measure ATD), tests show that all participants were significantly involved. Further studies using scenarios with up to 6 participants show corresponding time distributions – involvement does not decrease for even more participants. Simultaneous mapping activities (Measure SPA) have been identified in the majority of sessions using the tabletop system. In contrast to the sessions using CMapTools, exclusive physical access to the mapping controls could hardly be identified on the tabletop. Qualitative measures (Measures pCP, oCP and pCO) consistently showed the tabletop system to be perceived supportive with respect to cooperation. Explicitly, the physical setup (standing around a table) and the ability to physically reference and modify the map were mentioned and observed as positive factors several times. oCP on the tabletop system also showed that a restricting factor for simultaneous interaction – the editable but still exclusive access to the keyboard used for labeling – in fact seems to foster cooperation, as labels are discussed intensively before being entering. Using CMapTools, oCP does not show this tendency – in most of the cases the participant in control of the keyboard decides on the label autonomously. oCP provides hints to a possible explanation for this difference: usage patterns show, that on the tabletop, placement and association of elements and labeling them is a distributed activity shared among at least two participants, while for CMapTools the exclusive access to both, keyboard and mouse, leads to temporary full control of the mapping process by a specific participant (see also description of measure TPI above).

H2a – The tabletop leads to more cooperative activities than the screen-based system: As for Hypothesis 2a, TPI shows an average duration for one turn which is significantly lower for the tabletop than for the CMapTools. Simultaneous physical activity (Measure SPA) can be identified in more than two thirds of the tabletop sessions and is not present in the CMapTool-sessions. In turn, exclusive access to the input channels can be identified in 75% of the CMapTool-sessions while it has only be observed in 18% of the tabletop-sessions. Video analysis (oCP) confirms the effects of exclusive access to input channels (like mouse and keyboard) by a single user and shows, that the static nature of sitting around the system (as opposed by standing in the case of the tabletop) seems to cause the differences for the measures TPI and SPA. Activity time distribution, however, did not differ significantly. However, in 20% of the screen-based-sessions, participants noted that they were not able to contribute to the concept map at all. This was not the case in any of the tabletop sessions. Accordingly, the process of cooperation was perceived suboptimal in some cases for the screen-based system.

The results for hypotheses 2 and 2a are in line with the study of Rogers et al. [25] and Lucchi et al. [26], in which the effects of different interface types (screen-based, multitouch, tangible tabletop) have been examined in more detail than in the work presented here.

H3 – The tabletop supports the development of a common understanding: The fraction of time dedicated to discussion (Measure TDT) as a quantitative measure used to test Hypothesis 3 show, that on average two thirds of the overall mapping time are use used for activities contributing to negotiation of meaning. This amount is significantly lower in case of the screen-based system. Qualitative feedback (Measures pIU and pCU) consistently shows, that the system was perceived to have positive impacts on the development of a common understanding (with over 75% stating that it was easy to find consensus about the content to be represented in the map). Statements on pIU and pCU did not show any fundamental changes over time. People described similar perceptions when questioned directly after the mapping session and 8 weeks afterwards, where the latter were more elaborate due to perceptions of the impact in daily work practices.

H3a – Development of a common understanding was not facilitated when using the tabletop in comparison to the screen-based system: The perceived effects of on the development of a common understanding tested in Hypothesis 3a (Measures pIU and pCU) did not show any significant difference between the two examined tools. In both cases the majority of participants reported on the perception that it was "easy to come to a consensus" or to "find a common understanding" (80% for the CMapTools, 78% for the tabletop system).

The result for the hypotheses 3 and 3a however must not be overrated, as the concept maps created before the cooperative mapping session did not show fundamental differences in the understanding of the conceptual relationships within the mapping domain. Thus the development of a common understanding was likely to be perceived in a straightforward way. Future evaluations will require focusing on a more controversial task in order to be able to examine the effects of the systems in more detail.

6.2 Interpretation of Overall Results

Putting the results for each hypothesis in mutual context, leads to additional findings: The far less detailed concept maps created with the tabletop (as compared to the screen-based-system) lead to a significantly higher amount of discussion during the concept mapping process. When using the CMapTools participants have focused more on the process of representation than on the alignment of different individual understandings. Results of the examination of cooperation during the mapping process additionally show that groups using the CMapTools mostly have a dedicated "system operator" who actually creates the map. Regardless of the amount of interaction occurring during mapping, video analysis of the mapping sessions has revealed for all sessions using CMapTools that operating participants act as "filters", in particular, omitting content input by others which they consider not appropriate. The risk of representing the perception of information of a single participant in the map and leaving it unquestioned thus seems to be higher when using screen-based-tools compared to the tabletop.

Consequently, the restrictions of the tabletop system seem to contribute to the well perceived results on cooperative alignment of meaning – higher effort for map creation and the restricted modeling space seem to lead to more focused – however, not comprehensive – maps, which serve as an anchor point for further discussion. In the context of the design of tabletop-based interaction in groups in general, these findings direct research to a detailed study of the objectives of tabletop-supported applications. Whenever the goal is to serve as a facilitator and mediator for inter-personal communication and direct information exchange, tangible representations of the relevant data might be appropriate. In case an application mainly targets towards comprehensive representation of information, a non-tangible way of representing data (i.e. using touch-based interaction on a digitally enhanced surface) might be a better choice. Tangible representations still pose restrictions on the complexity of data to be represented. They cannot be overcome by developing workarounds, as shown for the container metaphor in this paper.

7 Conclusions

Concept Mapping is a method for externalizing and reflecting knowledge about real world phenomena. In cooperative settings concept maps can also be used to aid cooperative learning activities and the development of a common understanding about the mapping subject. This process imposes requirements on tool support that have hardly been addressed in existing concept mapping tools. In this paper, we have presented a tabletop interface designed to meet these requirements. In the empirical study the positive effects of the proposed system have been shown in comparison to a traditional, screen-based system.

The major limitation of this study turned out to be the selected mapping task, which has caused less controversy as expected. Due to the already existing congruence of the individual understandings of the mapping subject across the set of participants before the concept-mapping task, effects of the system on cooperative alignment of meaning could hardly be identified. Future work will have to build upon mapping topics and participants that lead to higher controversy and more effort of find a common understanding. Methodologically, one way to evaluate the development of a common understanding should be found that goes beyond analyzing individual perceptions. Candidates for such methods are dialogue-consensus-methods using structure-elaboration-techniques [13], and methods for diagnosing learning-related changes in mental models [22].

Overall, the proposed system shows positive effects on the cooperative construction of concept maps. While the physical restrictions of the tangible approach prevent the system from facilitating concept mapping as originally described in literature, the same restrictions foster discussion and steps towards cooperative alignment of meaning. For the selected sample topic the study shows that the tabletop approach and a screen-based approach support concept mapping on different levels while leading to similar outcomes regarding the development of a common viewpoint. The different qualities of support will have to be explored in future studies.

References

1. Ausubel, D.P.: Educational Psychology, A Cognitive View. Holt, Rinehart and Winston, Inc., New York (1969)
2. Antle, A.N.: The CTI framework: informing the design of tangible systems for children. In: Proceedings of the 1st international Conference on Tangible and Embedded interaction, TEI 2007, Baton Rouge, Louisiana, February 15 - 17, pp. 195–202. ACM, New York (2007)

3. Feltham, F.G.: Do the blocks rock: a tangible interface for play and exploration. In: Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing For Habitus and Habitat, OZCHI 2008, Cairns, Australia, December 08 - 12, vol. 287, pp. 188–194. ACM, New York (2008)
4. Quillan, R.M.: Semantic memory. In: Minsky, M. (ed.) Semantic Information Processing, pp. 216–270. MIT Press, Cambridge (1968)
5. Chang, K.E., Sung, Y.T., Chen, I.D.: The effect of concept mapping to enhance text comprehension and summarization. The Journal of Experimental Education 71(1), 5–23 (2002)
6. BouJadue, S., Atthie, M.: The effect of using concept maps as study tools on achievement in chemistry. Eurasia Journal of Mathematics, Science & Technology Education 4(3), 233–246 (2008)
7. Sutherland, S., Katz, S.: Concept mapping methodology: A catalyst for organizational learning. Evaluation and Program Planning 28, 257–269 (2005)
8. Fischer, F., Mandl, H.: Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. Journal of the Learning Sciences 14(3), 405–441 (2005)
9. Okada, A., Tomadaki, E., Buckinham Shum, S., Scott, P.: Fostering open sensemaking communities by combining knowledge maps and videoconferencing. Upgrade IX(3), 27–36 (2008)
10. Cañas, A., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M., Carvajal, R.: Cmaptools: A knowledge modeling and sharing environment. In: Concept Maps: Theory, Methodology, Technology, Proceedings of the 1st International Conference on Concept Mapping. Universidad Pública de Navarra, Pamplona (2004)
11. Fisher, K.M., Faletti, J., Patterson, H.A., Thornton, R., Lipson, J., Spring, C.: Computer-based concept mapping: SemNet software - a tool for describing knowledge networks. Journal of College Science Teaching 19(6), 347–352 (1990)
12. Hornecker, E.: Understanding the benefits of graspable interfaces for cooperative use. In: Proceedings of Coop, pp. 4–7 (2002)
13. Dann, H.-D.: Variation von Lege-Strukturen zur Wissensrepräsentation. In: Scheele, B. (ed.) Struktur-Lege-Verfahren als Dialog-Konsens-Methodik, Ein Zwischenfazit zur Forschungsentwicklung bei der rekonstruktiven Erhebung subjektiver Theorien, Aschendorff, Arbeiten zur sozialwissenschaftlichen Psychologie, vol. 25, pp. 2–41 (1992)
14. Novak, J., Cañas, A.J.: The theory underlying concept maps and how to construct them. Technical Report IHMC CmapTools 2006-01, Florida IHMC (2006)
15. Baraldi, S., Del Bimbo, A., Valli, A.: Interactive concept mapping through gesture recognition on a tabletop. In: Proceedings of CMC2006, San Josè, Costa Rica (2006)
16. Tanenbaum, K., Antle, A.: Using physical constraints to augment concept mapping on a tangible tabletop. In: Proceedings of the International Conference on Education and Information Technology, pp. 539–547 (2008)
17. Do-Lenh, S., Kaplan, F., Sharma, A., Dillenbourg, P.: Multi-finger interactions with papers on augmented tabletops. In: TEI 2009: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 267–274. ACM, New York (2009)
18. Maldonado, R.M., Kay, J., Yacef, K.: Collaborative concept mapping at the tabletop. In: ACM International Conference on Interactive Tabletops and Surfaces, ITS 2010, pp. 207–210. ACM, New York (2010)
19. Kaltenbrunner, M., Bencina, R.: reacTIVision: a computer-vision framework for table-based tangible interaction. In: TEI 2007: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, pp. 69–74. ACM Press, New York (2007)

20. Klemmer, S., Thomsen, M., Phelps-Goodman, E., Lee, R., Landay, J.: Where do web sites come from? capturing and interacting with design history. Human Factors in Computing Systems, CHI Letters 4(1) (2002)
21. Do-Lenh, S., Kaplan, F., Dillenbourg, P.: Paper-based concept map: the effects of tabletop on an expressive collaborative learning task. In: Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction, British Computer Society, pp. 149–158 (2009)
22. Ifenthaler, D.: Diagnose lernabhängiger Veränderung mentaler Modelle - Entwicklung der SMD-Technologie als methodologisches Verfahren zur relationalen, strukturellen und semantischen Analyse individueller Modellkonstruktionen. PhD thesis, University of Freiburg (2006)
23. Oppl, S.: Unterstützung expliziter Articulation Work. PhD-Thesis, Technical University of Vienna (2010)
24. Oppl, S.: Konsistente Verwendung von Metaphern als Erfolgskriterium für komplexe Tangible User Interfaces. In: Workshop-Proceedings der 9. fachübergreifenden Konferenz Mensch und Computer 2009 (2009)
25. Rogers, Y., Lim, Y., Hazelwood, W., Marshall, P.: Equal Opportunities: Do Shareable Interfaces Promote More Group Participation Than Single User Displays? Human-Computer Interaction 24(1-2), 79–116 (2009)
26. Lucchi, A., Jermann, P., Zufferey, G., Dillenbourg, P.: An empirical evaluation of touch and tangible interfaces for tabletop displays. In: Proceedings of TEI 2010. ACM Press, New York (2010)

The Continuous Interaction Space: Interaction Techniques Unifying Touch and Gesture on and above a Digital Surface

Nicolai Marquardt[1], Ricardo Jota[2], Saul Greenberg[1], and Joaquim A. Jorge[2]

[1] Dept. Computer Science, University of Calgary, Calgary, Alberta, Canada T2N 1N4
{nicolai.marquardt,saul.greenberg}@ucalgary.ca
[2] VIMMI Group / INESC-ID IST / Technical University of Lisbon, 1000-029, Portugal
rjc@vimmi.inesc-id.pt, jaj@inesc.pt

Abstract. The rising popularity of digital table surfaces has spawned considerable interest in new interaction techniques. Most interactions fall into one of two modalities: 1) direct touch and multi-touch (by hand and by tangibles) directly on the surface, and 2) hand gestures above the surface. The limitation is that these two modalities ignore the rich interaction space between them. To move beyond this limitation, we first contribute a unification of these discrete interaction modalities called the *continuous interaction space*. The idea is that many interaction techniques can be developed that go beyond these two modalities, where they can leverage the space between them. That is, we believe that the underlying system should treat the space on and above the surface as a *continuum*, where a person can use touch, gestures, and tangibles anywhere in the space and naturally move between them. Our second contribution illustrates this, where we introduce a variety of interaction categories that exploit the space between these modalities. For example, with our Extended Continuous Gestures category, a person can start an interaction with a direct touch and drag, then naturally lift off the surface and continue their drag with a hand gesture over the surface. For each interaction category, we implement an example (or use prior work) that illustrates how that technique can be applied. In summary, our primary contribution is to broaden the design space of interaction techniques for digital surfaces, where we populate the continuous interaction space both with concepts and examples that emerge from considering this space as a continuum.

Keywords: Touch, gestures, surfaces, interactive tabletops, 3D interaction, tangibles, portable devices, continuous interaction space.

1 Introduction

The advent of highly interactive digital surfaces has motivated researchers to develop a rich set of accompanying interaction techniques. While there are now a broad variety of techniques, hand input (sometimes holding tangibles) dominates. Most prior research has focused on two modes of recognizing hand input.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 461–476, 2011.

(a) *On the surface* includes touch interactions directly on the reachable parts of the display, usually using fingers and hands [7] [10] [19] [40], or tangible objects [2] [32]. Interactions typically include selecting, grabbing, throwing, rotating, and moving.

(b) *Above the surface* includes free-hand gesture recognition that occurs in the space above the surface. These interactions typically point to, select and access content not reachable by the user's direct touch [23] [24] [33], or map gestures to particular actions [17] [20] [30].

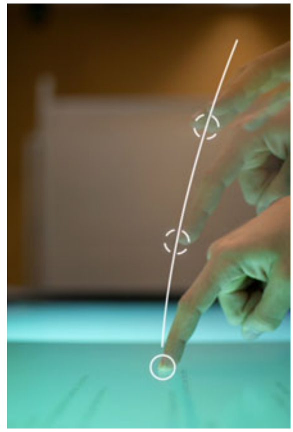


Fig. 1. Continuous interaction from the space above the tabletop to direct touch

The problem is that the vast majority of this prior research has explored hand interactions only within these two distinct modes. The limitation is that these two modalities ignore the rich interaction space between them. Our perspective differs. We propose and contribute a unification of these discrete interaction modalities called the *continuous interaction space*. The idea is that many interaction techniques can be developed to not only fit into a particular modality, but that they can leverage the space between them. That is, we believe that the underlying system should treat the space on and above the surface as a continuum, where a person can use touch, gestures, and tangibles anywhere in the space, and naturally move between them (illustrated in Figure 1).

In order to illustrate possible forms of interaction in the continuous interaction space, we introduce and contribute a variety of techniques that exploits the space between these modalities. We constructed these interactions in a way that takes full advantage of this unified space between touch on the surface and the space above it. For each interaction category, we implement an example or refer to prior work that illustrates how that technique can be applied. Thus, we also fit in earlier approaches of extending the interaction around digital surfaces; such as Hilliges' [16] and Wilson's [37] techniques for natural interaction with physically simulated digital objects, Baudisch's [2] stackable tangible blocks, or Subramanian's [30] interaction layers above a surface. While the techniques we introduce are specific to a horizontal tabletop display, the underlying concepts partially generalize to other digital surfaces such as electronic whiteboards and large wall displays.

In summary, our primary contribution is to broaden the design space of interaction techniques for digital surfaces, where we populate the continuous interaction space both with concept categories and examples that emerge from considering this space as a continuum. We don't claim that our categories are complete or our examples perfect. Rather, they show what is possible by considering the entire interaction space. The following section briefly introduces our understanding of the continuous interaction space. We then introduce our infrastructure that allowed us to develop and explore our interaction techniques. The main section of the paper introduces various categories of interaction techniques that illustrate how the unified interaction space can be applied in practice. We close with related work and a conclusion.

2 The Continuous Interaction Space

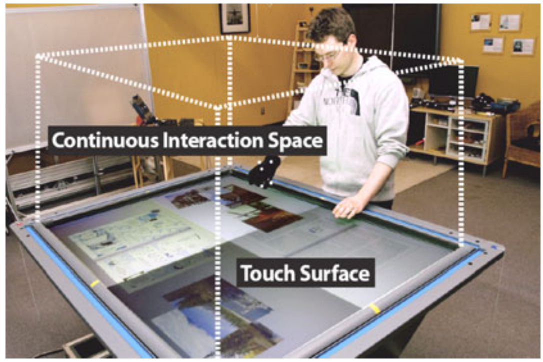


Fig. 2. The continuous interaction space above the interactive tabletop surface

We define the *continuous interaction space* as being composed of the direct touch surface and the space above, as illustrated and implemented in Figure 2. We argue that these are not two distinct spaces but instead a single interaction space. Specifically, a person can interact fluently in this 3D area, where gestural acts flow from the space above, to touch, and vice versa. As well, we believe that gestures should not be limited to interactions immediately below one's hands. That is, while the possible physical reach of a person's hands in this space is bounded to around ~1 meter above the surface, gestural acts can extend one's reach beyond these physical limits.

3 Infrastructure for Rapidly Prototyping Interaction Techniques

In order to explore possible interaction techniques in the continuous interaction space, we built an interactive tabletop system infrastructure for sensing touch and gestures. Atop this, we developed an application test bed that contains a variety of objects that people can manipulate. In particular, a person can view collections of digital content (visible on the tabletop surface in Figure 2): text documents, photos, videos, as well as abstract entities such as files and folders. The person can interact with this digital content through a variety of techniques, such as moving digital content, creating stacks of items, turning pages in digital documents (Figure 3), picking up content to reveal documents underneath (Figure 4 left), navigating videos (Figure 4 right), or browsing stacks of documents (Figure 5). These are part of the interaction techniques that we explain in Section 4[1].

The underlying infrastructure is implemented atop an interactive horizontal touch-sensitive SmartBoard surface [www.smarttech.com][2], and a Vicon motion tracking system [www.vicon.com]. This tracking system is composed of 8 high-speed infrared (IR) cameras that track reflective markers illuminated by IR spots attached to the cameras. As long as at least three cameras of the system see each individual marker, the system can infer the accurate 3D position of this marker (to around 1mm). We attached markers to gloves to track people's hands on and above the tabletop surface

[1] All except two (4.7 and 4.8) of the interaction techniques we present have been implemented in our test bed application. We included these two additional techniques as they fit into our categorization of the continuous interaction space.

[2] In order to allow displaying multiple legible documents simultaneously, the projection surface of this interactive table prototype supports a high resolution of 2800x2160 pixels.

(see right hand of the person in Figure 2). Our raw information includes the tracking of the hand model which returns the hand's and finger's yaw, pitch, and roll angle and its position in the 3D space (where a distance of 0 to the surface = touch), and the position of the surface in the 3D space. From these, we compute the pointing vector of the fingers to the surface (ray casting), and the normal vector of the hand perpendicular to the surface. We calculate the intersection of these vectors to the surface in screen pixel coordinates (so we know exactly what spot and objects on the surface these vectors are referring to). We also compute the hand's position, direction and vertical distance relative to the surface. We recognize postures such as pinching by measuring fingertip distance from one another. At a higher level, particular gestures (introduced shortly) register for updates of these hand movements.

We also track physical objects on and above the surface, where we attached similar IR reflective markers to tangible objects like wooden boxes (Figure 6), or portable tablet computers (Figure 5). Similar to the earlier described tracking of a person's hand, the system provides accurate position information of these tangibles. By comparing the distance of the tangible objects to the surface or a people's hands, we know whether the objects are placed on the surface or held in a person's hand.

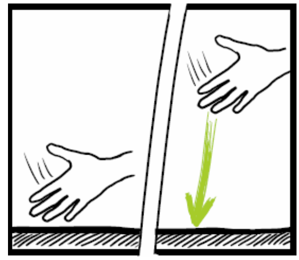
Systems such as the Vicon motion tracking are currently too expensive and unwieldy for deployment out of a lab environment. However, we believe our examples of unified interaction techniques will be implementable in the near future on much more affordable emerging technologies, e.g., shadow tracking [9], switchable diffusers [16], depth cameras [37], or any other technology that can accurately detect position and movement above the surface. Currently, these other technologies do not yet provide the necessary accurate tracking of 3D positions of hands, fingers, and objects above a tabletop surface. Our use of marker-based motion tracking allows us to rapidly and accurately prototype and explore possible interaction techniques, where we can re-implement them when these other technologies are ready.

4 Interaction Categories

To illustrate the possibilities for novel interaction techniques in the suggested continuous interaction space, we now introduce a variety of categories of interaction techniques that make use of this extended space. We explain the concept behind each category. We also introduce one or more examples applying this technique: examples are mostly implemented in our own tabletop system, but a few refer to point systems found in the prior art. Our concepts begin with gestural interaction techniques of a person's hand with digital content. As we will see, some categories not only consider the position and movements of people's hands above the surface, but also people's use of tangible objects or digital devices (usually held in one's hands or placed onto the surface) for interacting with digital content displayed on the interactive surface.

Importantly, we stress that the key idea and contributions of the interaction categories explained below is the flow of hand gestures across the interaction space as a continuum, from touch to moving above the surface. Part of doing so leverages input dimensions such as distance above the surface (where 0=touch), and hand orientation with respect to the surface. By designing such gestures, people can interact with digital content in a very rich, fluid and complex way.

4.1 Mirrored Gestures for Redundancy



Mirrored gestures are gestural pairs that *redundantly* encode identical functionality in either space. That is, a person can invoke the same action via a gesture either directly on the surface or the space above it. Mirrored gestures are related to the concept of *equal opportunity* in interface design, as suggested by Cockburn and Bryant [8]. Mirrored gestures offer *multiple representations* for gesture input, adding to the flexibility of the tabletop interface. The redundancy of functionality lets a person select the interaction technique most convenient in a particular situation, and freely choose to use either a gesture on or above the surface to interact with the digital content.

The mirrored gestures may be *different* or *similar*. Our first example uses two different gestures to produce an identical resize action: touch-based pinching with fingers touching the surface *vs.* a two-handed gesture that brackets an object via L-shaped fingers and thumb, and then stretches or shrinks the area to scale the object. Our second example uses two similar gestures to produce an identical action; in this case the two-handed bracketing gesture can be on or above the surface. A third example considers the navigation through pages in a digital book. To flip a page a person either uses touch to flip a single page through dragging (Figure 3), or does a hand waving gesture above the table that mimics page flipping.

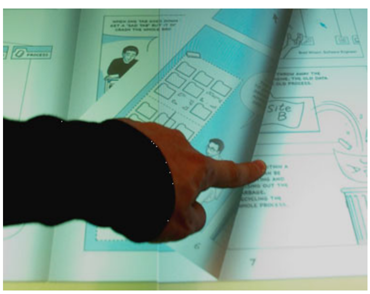
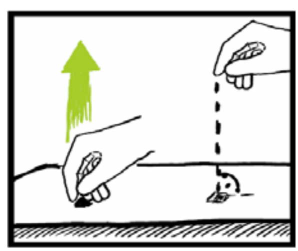


Fig. 3. *Turning* pages in a digital document through direct touch

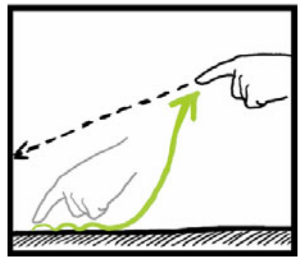
4.2 Extended Continuous Gestures to avoid Occlusion



With *extended continuous gestures*, a gesture that a person starts through direct touch on the interactive surface can *continue in the space above* the surface to avoid occlusion of the digital content visible on the tabletop display. That is, the person can lift the hand that performs the gesture off the surface, and continue the gesture in an invisible layer parallel to the touch surface.

As an example, this technique can be applied to allow precise movement of small digital content displayed on the tabletop. A person can grab the object through a 'pick' gesture while touching the surface and move it along the tabletop plane through dragging. The novelty is that this action can be continued by lifting the hand into the 3D space above the surface. Now, this reveals the exact position of the digital object on the screen that was previously occluded by the fingers touching the surface. Once moving the hand above the surface, a normal vector of the plane through the pinching fingers defines the moving position of the object on the screen. In this technique, the actual distance of the person's hand to the surface has no impact on the movement action that is performed.

4.3 Extended Reach / Raycasting Gestures



The *extended reach* or *raycasting* technique allows a person to extend their interaction range – which is physiologically limited by the reach of their hands – to remote locations on the surface. The fingers of the hand or the arm of a person extend as a pointing ray when the hand is lifted above the surface, where the intersection of this pointing ray with the tabletop plane defines the point of action in the interface. This approach is related to techniques such as Grossman's raycasting for volumetric displays [12] or Parker's *TractorBeam* pen raycasting technique on tabletop interfaces [23]. It differs in the way this technique allows fluid transitions from dragging through direct touch to raycasting remote pointing when lifting the hand off the screen, and back to direct touch.

For example, consider an extended reach gesture for moving an object around a surface. Similar to the extended continuous gesture we introduced before, a person starts by grabbing an object with a 'pick' gesture on the surface, and – while still touching – moves the object by dragging. Lifting the hand off the touch surface *affects that action's behaviour*. Now, the hand extends to a pointing ray, where the person continues to move the object, even to positions out of reach by direct touch. What is important is that these are not two separate actions. Rather, they are done as a continuous flow.

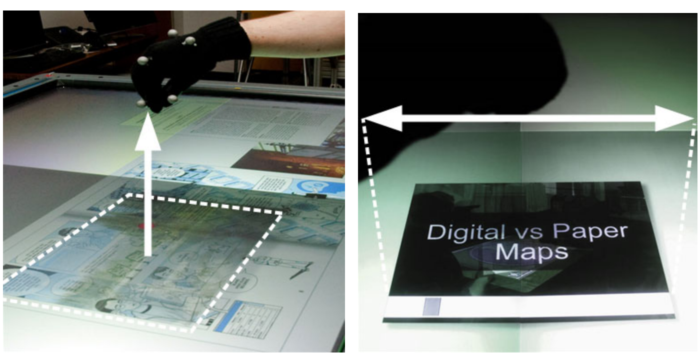
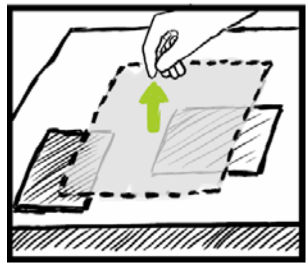


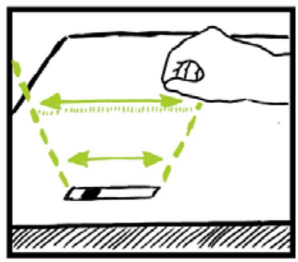
Fig. 4. *Lifting* up content to reveal documents underneath (left); *navigating* video with increased precision (right)

4.4 Lifting Gestures to Reveal Objects



The *lifting gesture* allows one to virtually lift up digital content, primarily to reveal other content lying underneath. Our example has a person interacting with a large collection of digital documents. To reveal content that is currently occluded by overlapping documents, the person can move their fingers together into a 'pick' posture, and then lift their fingers off the surface into a 'pick-up' gesture. As in real life, this lets them pick up the object above its current surface plane. The picked-up object becomes increasingly transparent to reveal the other objects that are now underneath it – transparency is a function of the vertical distance of the hand from the surface (Figure 4 left).

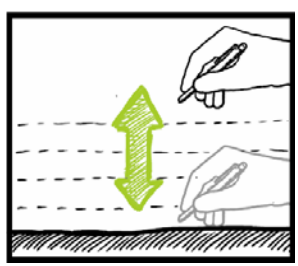
4.5 Lifting to Adjust Scale Precision



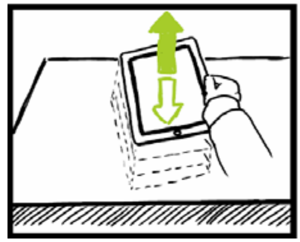
Our next interaction technique allows *lifting to adjust scale precision* to increase precision of lateral actions. Here, instead of just using the two dimensional touch surface as input to control a drag operation (e.g., moving a slider), a person can move their hand in the space above the surface to gain increased precision, where the lifting of a person's hand in the Z-axis acts as an input parameter that readjusts the scale precision of the lateral movement.

To illustrate, consider a digital video object that can be searched and played. Using direct touch, a person can drag the finger to the left or right on a video navigation slider, which causes the video to advance forwards or backwards. Yet if the object is small (or the video large), navigating is hard because small touch movements translate to large jumps in the video timeline. With lifting to adjust scale precision, a person initiates and begins the search by first touching the video slider, and then gains precision by lifting above the table (Figure 4 right). Hand height continually rescales the slider, where the same lateral movement will result in a smaller move through the video. Thus both lateral and vertical hand position controls the search.

4.6 Interaction with Discrete Layers and Stacks of Digital Objects



Subramanian [30] previously suggested that the space above the tabletop surface can be divided into *discrete interaction layers*. In our adoption of this idea, the space above the surface is divided into multiple parallel planes, each corresponding to a layer. Each of these layers can then (for example) correspond to layers of visual content in the tabletop application, or even to different interpretations of gestures within a layer (e.g., an annotation layer, an editing layer, a movement layer, etc.). A person can navigate through these layers by moving in the Z-axis above the table. In our first example, a person uses a 'pick' gesture with their fingers (or use a pen as in [30]) to select content in each of these layers, and even move content from one layer to another.



Our second example reconsiders layers as stacks. Here, a person browses the digital content ordered in a *stack of digital objects* (such as photos) by holding and moving a tablet computer in the space above the stack. The screen of the tablet computer then shows the particular photo corresponding to its particular Z position inside of the virtual stack of photos (Figure 5).

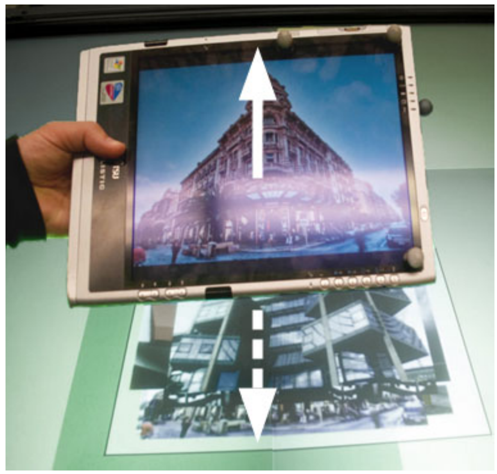
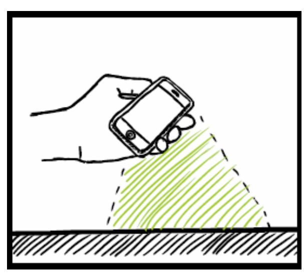


Fig. 5. *Browsing* stack of photos by holding and moving a tablet computer

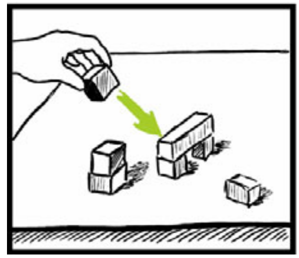
This technique has two practical limits. First, the maximum height of the virtual stack of documents is limited by the reach of a person's arms. If the person is holding a tablet (as in our 2nd example), this also limits the visibility of the tablet computer display. Second, the height of each individual layer is limited by the ability of people to select (and stay) inside of these layers. Subramanian's [30] evaluation found a layer thickness of 4 cm optimal for tasks where people have to stay in the layer. In our own implementation of browsing through a stack of documents with a tablet computer, we found a layer thickness of 1.0 cm to 1.5 cm as still practical for selecting content (and holding the device in this position).

4.7 Magic Lenses and View Ports



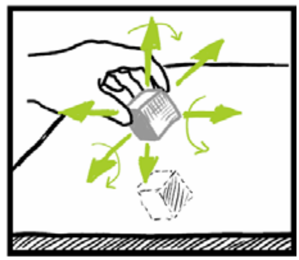
Magic lenses (inspired by Bier et al. [5]) let a person move a plane above the tabletop to see individual, modified views onto the data that is visualized on the table [17] [27]. This moving plane can be, for instance, a sheet of paper that shows information of a projector (e.g., [17]), or a portable tablet-sized device that shows information. While moving these planes through the 3D space above the surface, the content displayed on the magic lens represents the current view corresponding to the 3D position of the lens. For instance, when visualizing 3D volumetric medial scans on the tabletop, the magic lens can visualize cut sections of that dataset [17].

4.8 Stacking of Physical Objects



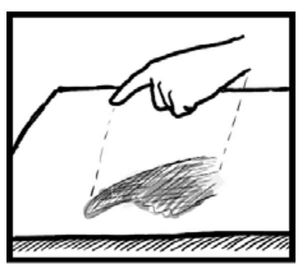
People can stack tangible objects on a tabletop, where the surface application reacts to the stacks built on the table. Baudisch et al. [2] proposed this approach with the *Lumino* project, where a Microsoft Surface tabletop recognizes stacks of small cuboid blocks. Bartindale et al. proposed a related approach detecting stacks of items on a surface, by determining the order of fiduciary markers [1]. Overall, these approaches allow a person to create physical stacks and structures on top of the usually flat tabletop surface.

4.9 6-DOF Manipulation



The free movement of people's hands in the space above the surface can be leveraged to allow full *6 degree of freedom (DOF) interaction* with digital content. These dimensions are movement along and rotation around the three axes. In previous work, Hancock contributed 2- and 3-touch techniques that allow 6-DOF manipulations of objects directly on the surface (e.g., [13, 15]). Our technique differs: instead of using only multi-touch techniques on the surface, a person can 'pick up' any digital object, and modify the object's position and orientation by the way they move their hand above the surface. Our example displays a digital object (a 3D cube) that can be picked up; the cube's position and orientation are then directly mapped to the position of the person's hand (3D location: yaw, pitch, roll). Of course, the movement along this 6-DOF is again limited by the person's reach of the hand, and the rotation is limited by the movement around the wrist joint.

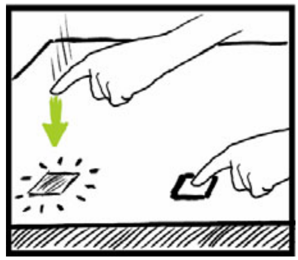
4.10 Feedback of Hand and Object Actions by Shadows



While it is relatively easy for a person to know what they are manipulating when they are directly touching it, this becomes harder to do when one's hand is lifted away from the surface as the physical connection is broken. To ease this, the system can provide *feedback* of how it is recognizing a person's gestures via *hand and arm shadows* (such as in [16] or [26]), or by visualizing abstracted hand shapes atop the objects it is manipulating. Examples of such shapes are a pointing hand, a pinching gesture, or the flat hand to reflect user actions atop the object(s) being manipulated. To avoid occlusions of user interface elements, transparency can be used to overlay feedback visualizations.

Digital objects can also provide additional continuous feedback about their status. For instance, objects that are picked up from the table (with a grab gesture) and moved around by the user can render shadows onto the table surface, where the shadow size depends on the current distance of the hand to the surface – or maybe even 'merges' with the shadow of a person's hand as proposed in [16]. This is a very natural mapping of the position in the 3D space to the displayed content on a 2D surface. Of course, all other object behaviours should reflect fine-grained actions corresponding to gestural movement: rotation, transparency, and so on. If done well, a person should be able to understand, self-correct, and fine-tune their gestures to control the object in a meaningful way.

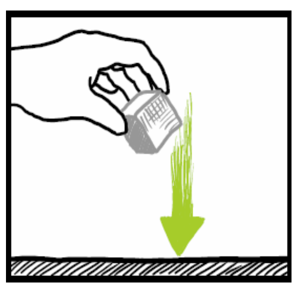
4.11 Feedback of Possible Actions by Hovering



Related to the technique above is feedback about what *actions are possible* as a person moves and *hovers* above objects. For example, by sensing the position of people's hands above the surface interface, widgets on the touch screen (such as buttons, sliders) can give people visual or audible feedback about possible actions (e.g., [9]). Similar to the way GUI buttons often change their color when the mouse cursor hovers over the button, the interactive elements of our tabletop application interface could change their visual appearance (e.g., glowing border, different background color) once a person's hand approaches the widget. This technique would extend the common two-state interaction with touchscreens (i.e., touching the surface or not) to the three-state interaction (i.e., away, hovering, touching) familiar from GUI mouse interfaces or pen-based interaction [6], and possibly even to continuous states where the feedback is very specific to the possibilities allowed at a particular Z-coordinate.

As well, the hovering feedback does not necessarily have to be binary (*on* or *off*). Instead, widgets and other on-screen elements can adjust the feedback as a function of a person's hand to the surface. For example, a button could begin showing a thin graphical border when the hand is around 20cm away, and continue increasing the thickness of this border the more the hand approaches the button. This kind of increasing amount of feedback might be in particular useful for widget actions with higher impact (e.g., deleting).

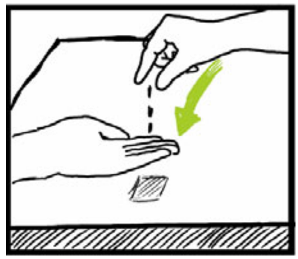
4.12 Picking and Dropping Gestures



Another interaction technique in the continuous interaction space is *picking and dropping of digital objects*, where those objects react according to a physical simulation imitating real-world behaviour. This is also called *physics-based interaction* (e.g., [16]). That is, a person can pick up a virtual object (say a 3D cube) with a 'pick' gesture, move it to another location, and drop it again by releasing the gesture. The virtual objects displayed on the screen behave according to a simulated physical model and gravity; for example, they fall back to the 'ground', push other objects away, or stop when bouncing against 'walls'. The Z distance can affect this behaviour, e.g., longer falls have larger consequences.

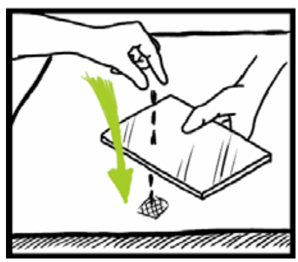
This approach of simulating the physics behaviour of digital objects has been the basis for various interactive tabletop systems (e.g., [3, 14, 16, 36, 37]). Some of the proposed interaction techniques are:

- Picking up objects through a 'virtual joint' between the hand and the object; placing objects in virtual containers (e.g., cup); piling and stacking virtual objects [16].
- Swiping objects off the side of the screen onto a person's hand; moving the virtual content on a person's hand to another surface [37].
- Picking and dropping virtual objects, augmented with 2D and 3D visualizations [3].
- Passing a virtual object from one person's hand to another person's hand (e.g., by dropping the object onto the other person's hand) [37].



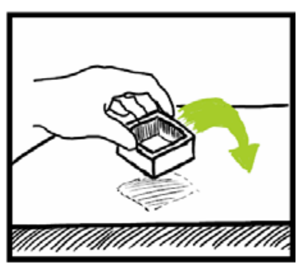
In these techniques, the distance of a person's hand to the touch surface can have impact on the physical behaviour of the objects, such as when dropping an object and it falls back onto the surface. Depending on the modelling of simulation, the distance can affect the bouncing of an object (and thus the location where the object lands), or even the possible deformation of an object (e.g., when using rendered 3D shapes).

4.13 Picking and Dropping Objects through Filters



As a novel variation of the above physics based interaction, we created another interaction category: *picking and dropping virtual objects through physical objects functioning as a filter*. In our example, a person picks a virtual object from the tabletop surface and lets the object fall though the filter (a thin acrylic panel). Once the virtual object 'falls' through the plane of the physical filter, the content of the digital object is modified according to the filter settings associated with this tangible filter object.

4.14 Pouring Gesture



When exploiting physics-based interaction, physical artefacts can function as containers for virtual objects, and afford gestures such as *pouring* to empty the container.

For example, in our prototype tabletop application a person can take a wooden box (Figure 6) and add virtual objects to this container by picking them up from the surface and dropping them from above into the container. The person can then move the box on the table or in the 3D space around the surface (the current virtual objects are displayed as shadows underneath the box). To release the content, the person turns over the box, and thus 'pours' the digital information out of the physical container (see Figure 6).

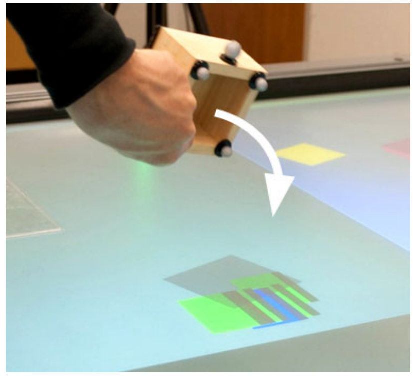


Fig. 6. *Pouring* virtual documents out of a physical wooden box

5 Related Work

Our work relates to previous research of interaction techniques out of the following three areas: multi-touch interaction with digital tabletop surfaces, 3D and physics based interaction, and tangible objects and digital devices on tabletops. In the following, we briefly review this work.

5.1 Multi-touch Tabletop Interaction

Touch-related research is thriving. While most initial work was on detecting contact points (e.g., one or two fingertips), current interest has shifted to whole hand interactions. Wu [40] and Cao et al. [7] present a touch surface that understands such whole hand touches. Epps et al. studied hand shapes use in tabletop gestures [10]; their study suggests the need for both a touch screen and computer vision-based gesture tracking and recognition, for applications that require a wide range of commands. Marquardt et al. introduced the fiduciary tagged glove on the Microsoft Surface that facilitates identifying the hand part that touched the surface [22]. Others have explored how hand gestures can control a large scale display from a distance [19, 21, 33], and the influence of virtual embodiments to increase awareness when interacting with digital surfaces [25].

Recently researchers began to use the space above the surface for interaction. Unlike the approach in this paper, most developed 'point systems' or techniques rather than consider the space holistically. Echtler tracked hand shadows to support hovering actions on a tabletop [9], mimicking the mouse hover action. Izadi applied a switchable diffuser to the tabletop, which captures hand gestures above the tabletop [17]. Parker's Tractor beam [23] studied how pointing with a pen, in addition to touching the surface, could improve interaction by enabling users to get out-of-reach objects [24]. Benko's muscle sensing technique for people's hands also allows inferring information about their movement above the tabletop surface [4]. Other techniques include dividing the space above the surface into interaction layers. Lucero [20] defines gestures that allow vertical movement in-between layers to organize piles of pictures. Subramanian et al. present a multi-layer approach to tabletops [30], where people can interact with several layers of visual content by moving their pen in the 3D space above. Kattinakere [18] further investigated user performance of steering tasks along paths in such above-the-surface layers. We were inspired by (and thus integrated) concepts out of this body of related research into our categorization of techniques.

5.2 3D Gestural and Physics Based Interaction with Tabletops

A few researchers introduced physics based and 3D gestural interaction techniques to interactive tabletops. Wilson's *Touchlight* [39] and depth sensing above tabletops [36] opened up a field for rich 3D gestural interaction with digital content. Benko introduced cross dimensional gesture interaction techniques [3] that function as a bridge between 2D and 3D interaction. Grossman investigated interaction techniques with content on 3D volumetric displays [12]. Later, Hilliges introduced physics based interaction techniques enabling natural interaction with digital content [16]. This technique was later extended with depth-sensing capabilities [38], in order to provide an even richer set of possible gestures. In [37], this approach was taken further by including tracking of multiple interactive surfaces as well as the full 3D volume of the room they are located in. This permits interaction techniques spanning across multiple interactive surfaces.

Research of tabletop systems is partially inspired by other domains. These are, for instance, computer vision techniques for reliable tracking of people's hands and arms captured by cameras (surveyed in [41]), or techniques developed in virtual reality environment research for sensing 3D hand and object motions [29]: optical, marker-based, magnetic, and other tracking approaches.

Grossman's taxonomy of 3D on tabletops [11] categorizes properties of the visualization and interaction with such systems. In this context, we focus on tabletops with 3D volumetric direct input; but are not limited regarding display properties (2D, surface constrained, heads-up projection, or volumetric; see [11]). Our contribution is the emphasis on a continuous interaction space above surfaces, and the set of interaction techniques that in particular leverage the continuous input from people using touch, gestures, and tangibles on and above tabletop surfaces.

5.3 Tangible Objects and Digital Devices on Tabletops

One restriction when interacting with digital tabletops is the limited tangible expressiveness of a flat tabletop surface: even if the display visualizes 3D objects, a person cannot touch, grab, or feel these virtual objects. Consequently, researchers began to introduce tangible objects in tabletop interaction. In early research, Ullmer's *Metadesk* introduced physical interface elements in tabletop interaction [31]. Underkoffler's *URP* [32] combined tangible blocks representing buildings with a tabletop interface to enable intuitive interaction with a digital urban design simulation.

Starner's *Perceptive Workbench* [28] later combined the recognition of people's gestures above the tabletop surface with tangible computing. Later, Weiss introduced *Madgets* [35] as physical control widgets on a tabletop (such as buttons, sliders) that can automatically move through an emitted field of an array of electro magnets underneath the table. With Baudisch's *Lumino* [2], a tabletop system can also sense the geometrical form of physical cubes stacked on a tabletop.

Tangible objects or digital devices can also facilitate the exploration of visualized digital content. Inspired by [5], *magic lenses* allow a person to get a personalized view onto a data set, or cut sections of volumetric data visualized on the table [27]. In another approach, Voida's *i-Loupe* allows a person use a mobile handheld device to view and interact with the content displayed on a large interactive surface [34].

While many techniques for tangible interaction on tabletop surfaces exist, our categorization focuses on those that in particular leverage the continuous distance of objects and their movements above the surface.

6 Summary

We proposed the concept of a continuous interaction space above a digital surface, where people can fluently move from touch interaction to gestures above the surface. The idea is simple, yet the implications are profound. Most current interaction techniques are still anchored into falling in one of two modes: either based on direct touch *or* gestures above the surface. The concept of a continuous interaction space is to remove these arbitrary restrictions, where we want people to consider both modalities working together. Our categorization of techniques is suggestive, where we use them to illustrate an (incomplete) range of new interaction techniques that merge these modalities. Through example implementations and interwoven related work, we illustrated various gesture compositions that make use of this extended space.

Of course, more advanced combinations and extensions of such gestures are not only possible but highly probable. We do not claim that our categories or examples are ideal ones or exhaustive. There is much left to do. Overall, we believe that the understanding and designing gestures that exploit this continuous space above the digital surface is beneficial for creating intuitive interactions with the digital content. Our categories and examples illustrate some of the possibilities. As a starting point they suggest future exploration, iterative refinement, and eventual evaluation.

Acknowledgments. This research is partially funded by the iCORE/NSERC/SMART Chair in Interactive Technologies, Alberta Innovates Technology Futures, NSERC, and SMART Technologies Inc. Ricardo Jota was supported by the Portuguese Foundation for Science and Technology (SFRH/BD/ 17574/2004).

References

1. Bartindale, T., Harrison, C.: Stacks on the surface: resolving physical order using fiducial markers with structured transparency. In: Proc. of ITS 2009. ACM, New York (2009)
2. Baudisch, P., Becker, T., Rudeck, F.: Lumino: tangible blocks for tabletop computers based on glass fiber bundles. In: Proc. of CHI 2010, pp. 1165–1174. ACM, New York (2010)
3. Benko, H., Ishak, E.W., Feiner, S.: Cross-dimensional gestural interaction techniques for hybrid immersive environments. In: Proc. of IEEE VR 2005, pp. 209–216 (2005)
4. Benko, H., Saponas, T.S., Morris, D., Tan, D.: Enhancing input on and above the interactive surface with muscle sensing. In: Proc. of ITS 2009, pp. 93–100. ACM, New York (2009)
5. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., DeRose, T.D.: Toolglass and magic lenses: the see-through interface. In: Proc. of SIGGRAPH 1993, pp. 73–80. ACM, New York (1993)

6. Buxton, W.: A three-state model of graphical input. In: Proc. of INTERACT 1990, pp. 449–456. North-Holland Publishing Co, Amsterdam (1990)
7. Cao, X., Wilson, A., Balakrishnan, R., Hinckley, K., Hudson, S.: ShapeTouch: Leveraging Contact Shape on Interactive Surfaces. In: Proc. of TABLETOP 2008. IEEE, Los Alamitos (2008)
8. Cockburn, A., Bryant, A.: Leogo: An Equal Opportunity User Interface for Programming. Journal of Visual Languages & Computing 8(5-6), 601–619 (1997)
9. Echtler, F., Huber, M., Klinker, G.: Shadow tracking on multi-touch tables. In: Proc. of AVI 2008, pp. 388–391. ACM, New York (2008)
10. Epps, J., Lichman, S., Wu, M.: A study of hand shape use in tabletop gesture interaction. In: CHI 2006 Extended Abstracts, pp. 748–753. ACM, New York (2006)
11. Grossman, T., Wigdor, D.: Going Deeper: a Taxonomy of 3D on the Tabletop. In: Proc. of TABLETOP 2007, pp. 137–144. IEEE, Los Alamitos (2007)
12. Grossman, T., Balakrishnan, R.: The design and evaluation of selection techniques for 3D volumetric displays. In: Proc. of UIST 2006, pp. 3–12. ACM, New York (2006)
13. Hancock, M., Carpendale, S., Cockburn, A.: Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In: Proc. of CHI 2007, pp. 1147–1156. ACM, New York (2007)
14. Hancock, M., ten Cate, T., Carpendale, S., Isenberg, T.: Supporting sandtray therapy on an interactive tabletop. In: Proc. of CHI 2010, pp. 2133–2142. ACM, New York (2010)
15. Hancock, M., ten Cate, T., Carpendale, S.: Sticky tools: full 6DOF force-based interaction for multi-touch tables. In: Proc. of ITS 2009, pp. 133–140. ACM, New York (2009)
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. of UIST 2009, pp. 139–148. ACM, New York (2009)
17. Izadi, S., Hodges, S., Taylor, S., et al.: Going beyond the display: a surface technology with an electronically switchable diffuser. In: Proc. of UIST 2008, pp. 269–278. ACM, New York (2008)
18. Kattinakere, R.S., Grossman, T., Subramanian, S.: Modeling steering within above-the-surface interaction layers. In: Proc. of CHI 2007, pp. 317–326. ACM, New York (2007)
19. Lee, J., Lee, J., Kim, H., Kim, J.: Gesture-Based Interactions on Multiple Large Displays with a Tabletop Interface. In: Universal Access in Human-Computer Interaction. Ambient Interaction, pp. 936–942 (2007)
20. Lucero, A., Aliakseyeu, D., Martens, J.: Augmenting Mood Boards: Flexible and Intuitive Interaction in the Context of the Design Studio. In: Proc. of TABLETOP 2007, pp. 147–154. IEEE, Los Alamitos (2007)
21. Malik, S., Ranjan, A., Balakrishnan, R.: Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In: Proc. of UIST 2005, pp. 43–52. ACM, New York (2005)
22. Marquardt, N., Kiemer, J., Greenberg, S.: What Caused That Touch? Expressive Interaction with a Surface through Fiduciary-Tagged Gloves. In: Proc. of ITS 2010, pp. 139–142. ACM, New York (2010)
23. Parker, J.K., Mandryk, R.L., Inkpen, K.M.: TractorBeam: seamless integration of local and remote pointing for tabletop displays. In: Proc. of Graphics Interface - GI 2005, pp. 33–40. Canadian Information Processing Society (2005)
24. Parker, J.K., Mandryk, R.L., Inkpen, K.M.: Integrating Point and Touch for Interaction with Digital Tabletop Displays. IEEE Comput. Graph. Appl. 26(5), 28–35 (2006)

25. Pinelle, D., Nacenta, M., Gutwin, C., Stach, T.: The effects of co-present embodiments on awareness and collaboration in tabletop groupware. In: Proc. of Graphics Interface - GI 2008, pp. 1–8. Canadian Information Processing Society (2008)
26. Shoemaker, G., Tsukitani, T., Kitamura, Y., Booth, K.S.: Body-centric interaction techniques for very large wall displays. In: Proc. of NordiCHI 2010, pp. 463–472. ACM, New York (2010)
27. Spindler, M., Stellmach, S., Dachselt, R.: PaperLens: advanced magic lens interaction above the tabletop. In: Proc. of ITS 2009, pp. 69–76. ACM, New York (2009)
28. Starner, T., Leibe, B., Minnen, D., Westeyn, T., Hurst, A., Weeks, J.: The perceptive workbench: Computer-vision-based gesture tracking, object tracking, and 3D reconstruction for augmented desks. Machine Vision and Applications 14(1) (2003)
29. Sturman, D.J., Zeltzer, D.: A Survey of Glove-based Input. IEEE Comput. Graph. Appl. 14(1), 30–39 (1994)
30. Subramanian, S., Aliakseyeu, D., Lucero, A.: Multi-layer interaction for digital tables. In: Proc. UIST 2006, pp. 269–272. ACM, New York (2006)
31. Ullmer, B., Ishii, H.: The metaDESK: models and prototypes for tangible user interfaces. In: Proc. of UIST 1997, pp. 223–232. ACM, New York (1997)
32. Underkoffler, J., Ishii, H.: Urp: a luminous-tangible workbench for urban planning and design. In: Proc. of CHI 1999, pp. 386–393. ACM, New York (1999)
33. Vogel, D., Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays. In: Proc. of UIST 2005, pp. 33–42. ACM, New York (2005)
34. Voida, S., Tobiasz, M., Stromer, J., Isenberg, P., Carpendale, S.: Getting practical with interactive tabletop displays: designing for dense data, “fat fingers,” diverse interactions, and face-to-face collaboration. In: Proc. of ITS 2009, pp. 109–116. ACM, New York (2009)
35. Weiss, M., Schwarz, F., Jakubowski, S., Borchers, J.: Madgets: actuating widgets on interactive tabletops. In: Proc. of UIST 2010, pp. 293–302. ACM, New York (2010)
36. Wilson, A.: Depth-Sensing Video Cameras for 3D Tangible Tabletop Interaction. In: Proc. of TABLETOP 2007, pp. 201–204. IEEE, Los Alamitos (2007)
37. Wilson, A.D., Benko, H.: Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In: Proc. of UIST 2010, pp. 273–282. ACM, New York (2010)
38. Wilson, A.D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., Kirk, D.: Bringing physics to the surface. In: Proc. of UIST 2008, pp. 67–76. ACM, New York (2008)
39. Wilson, A.D.: TouchLight: an imaging touch screen and display for gesture-based interaction. In: Proc. of ICMI, pp. 69–76. ACM Press, New York (2004)
40. Wu, M., Balakrishnan, R.: Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In: Proc. of UIST 2003, pp. 193–202. ACM, New York (2003)
41. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Computing Surveys (CSUR) 38 (2006)

AffinityTable - A Hybrid Surface for Supporting Affinity Diagramming

Florian Geyer, Ulrike Pfeil, Jochen Budzinski, Anita Höchtl, and Harald Reiterer

University of Konstanz, 78457 Konstanz, Germany
{firstname.lastname}@uni-konstanz.de

Abstract. Using affinity diagramming as an example, we investigate interaction techniques for supporting collaborative design activities. Based on an observational study, we identified design guidelines that need to be addressed to find a close fit to embodied practice. Using this knowledge, we designed and implemented AffinityTable, a hybrid surface for supporting affinity diagramming. The tool combines digital pen & paper with an interactive table and tangible tokens. An additional vertical display is used to support reflection and group coordination.

Keywords: design tools, collaborative design, affinity diagramming, reality-based interaction, digital pen & paper, hybrid interactive surfaces.

1 Introduction

Affinity diagramming is a collaborative design method applied early in the design process for analyzing a design problem or to create first design solutions. It is typically practiced with pen & paper and a shared surface such as a whiteboard or a large wall. It consists of three main phases: (i) generating (participants individually create content on sticky notes), (ii) sharing (participants present and collect notes), and (iii) structuring (participants arrange notes into meaningful categories). While teams of designers often apply similar techniques within the design process, only few existing digital tools explicitly support these without imposing limitations on the workflow, social coordination and embodiment of thought. Beyer & Holtzblatt [1] explicitly recommend traditional media for practicing affinity diagramming since using desktop-based digital tools would isolate participants, leading to a breakdown of communication that is vital for a shared understanding in the group. However, in this paper we argue that emerging technologies, like digital pen & paper, large high-resolution displays, multi-touch and tangible interaction, can be utilized to overcome these limitations and may offer additional benefits that go beyond what is possible with traditional physical-only tools. Therefore, we contribute *AffinityTable*, a hybrid tabletop system for supporting affinity diagramming. It allows users to follow their traditional workflow and embodied practice by still augmenting typical tasks in a lightweight way. Furthermore, it preserves the benefits of physical material during divergent idea generation activities, while enhancing convergent activities through artifact handling support and by providing focus and context views.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 477–484, 2011.

Our work extends related research in computing environments for co-located design work. *The Designer's Outpost* [2] introduced paper-based interaction concepts in the context of collaborative web site design. It combines physical sticky notes with an interactive whiteboard system. By using overhead image capture, notes can be combined with digital ink annotations. The *Shared Design Space* [3] combines digital pen & paper with projected images on a pen-operated interactive table for high-resolution drawing with multiple users. Optical marker tracking is utilized to recognize the location of paper on the table. *Pictionaire* [4] uses overhead image capture and projection to combine physical artifacts with digital annotations for collaborative design work on a hybrid tabletop system. *BrainStorm* [5] combines multiple pen-operated displays to support brainstorming processes. Digital notes are created on an interactive table and transferred to a wall projection for discussion and clustering. *The Designer's Environment* [6] supports the KJ creativity method with multimodal interactions for grouping and linking of digital notes on a tabletop system. Handwritten notes are sent to the table from tablet PCs. Text recognition is used on the tablets for retrieving images from the web according to the content of notes.

Our approach extends these efforts with *hybrid* interaction techniques that make use of both *physical* and *digital* representations to better integrate with the social and embodied characteristics of *existing workflows*. Hybrid surfaces are proposed in order to explore the boundaries between direct touch interfaces and tangible elements [7]. In the following, we describe design guidelines that we derived from an observational study before presenting our proposed workspace design and interaction techniques.

2 Design Guidelines

One of the main ideas of reality-based interaction [8] is to build upon characteristics of reality to make interfaces more natural. Jacob et al. recommend looking at reality first to identify critical issues. Computational power may then be used to address those issues that do not work well and can be improved by using technology. However, mechanisms that work good in reality should be preserved. We propose to use this understanding for mapping our potential design space. Therefore, we conducted an observational study of students practicing affinity diagramming. The goal was to identify issues that are critical to the success of the technique. These findings are to be used as a foundation for the design of adequate hybrid interaction techniques that combine physical and digital representations.

We observed three groups of students within a practical session in context of an interaction design course. The group sessions were videotaped and lasted between 1.5 and 2 hours. In addition, each group session was observed by a researcher who made notes of the group activities. Based on the procedures of qualitative content analysis, separate categories were inductively developed for following aspects: *interaction with physical artifacts*, *use of different workspaces* and *attention and coordination*. In the following, we will only describe the most important findings that led to the design of *AffinityTable* and relate them to supporting results from other sources. By relating our findings to that of other researches, we can improve validity of our results and may infer specific guidelines for the design of an adequate digital workspace.

Fig. 1. An observation revealed important characteristics of embodied practice

Based on our analysis we found that paper artifacts are especially useful during divergent activities, like individual ideation, where a large number of different artifacts are created in rapid cycles. Participants also used paper notes for individual reflection, like flipping through created content to search for inspiration. This is in line with results from a study reported by Cook & Bailey [9] on the use of paper in professional design practice. The authors conclude that paper is highly appreciated due to its unique affordances that support both fluent access and social interactions. Klemmer et al. [2] also emphasize the benefits of paper material for informal design activities. Therefore, we conclude that paper may be used to support individual ideation and reflection activities.

The table was mainly used for individual work. Personal workspaces (see Fig. 1, b) were used for reflection and for preparing content for presentation in the group. However, collaborative work like presenting and discussing took place at the whiteboard. This finding is supported by Vyas [10], who observed the appropriation of physical surfaces in design studios. The author reports that horizontal surfaces mainly serve as a space for action, while vertical surfaces are useful for reflection during convergent activities. We argue that this is due to the differences in accessibility of artifacts and display space. However, we observed that when engaging in actions during convergent phases, the vertical display was frequently blocked by some members of the group (see Fig. 1, a,d), thereby limiting access and visibility to the other participants, which may lead to communication problems. This issue however correlates with the number of participants. Yet, the transfer of notes between personal and shared workspace, which was required for presentation, was accompanied with physical restrictions (see Fig. 1, c) due to the distance between these work surfaces. Like other researchers [5,6], we also consider personal workspaces as essential for the efficiency of creative group work as they moderate social factors like evaluation apprehension by providing a semi-private retreat from the group. Other social factors like production blocking and free riding are partly addressed by the design technique itself. Therefore, the basic workflow of a design technique should not be changed.

During phases of discussion, we observed rapid changes of attention between personal and shared workspaces and between detailed discussions of specific artifacts (see Fig. 1, e) and coordination tasks that required an overview over all artifacts.

However, we also found that this practice may lead to coordination and communication issues when dealing with a large number of artifacts during discussion. At some points, participants had problems following discussion due to rapid shifts of attention or because some artifacts on the shared work surface were blocked by other participants.

We therefore summarize our findings with following design guidelines: support individual *ideation and reflection with paper* (DG1); support *personal workspaces* (DG2); provide support for *switching between action and reflection* (DG3); *ease the transfer* of artifacts between workspaces (DG4); *preserve the workflow* of the original technique (DG5).

3 A Hybrid Surface for Supporting Affinity Diagramming

AffinityTable was designed in an iterative design process. In a first step, a workspace was developed to address the findings from our study. In a second step, specific hybrid interaction techniques were designed to support and augment the workflow and tasks within affinity diagramming.

Fig. 2. AffinityTable combines different workspaces by using digital pen & paper, an interactive table with tangible tools and a coupled, high-resolution vertical display

3.1 Workspace Design

Our goal when designing *AffinityTable* was to replicate the benefits of reality and to improve identified issues by using interactive technology. The physical workspace used in practice was composed of one vertical and one horizontal work surface (see Fig. 1, a). We found that these work surfaces are important for preserving the workflow of the technique. Therefore, we decided to keep the basic layout, but to provide smoother transitions between action and a reflection by the means of a virtual workspace across two interactive displays (DG3). Fig. 2 shows the final workspace design, which is based on an interactive table for collaboratively interacting and reflecting on a focused region of the workspace (*shared action space,* see Fig. 2, a) and a high-resolution wall display (*process reflection space,* see Fig. 2, b) for reflecting on the overall progress of the process. Thereby, the table displays a magnified region of the whole workspace, which is shown on the wall display. The workspace can be navigated on the table with panning gestures. A focus box on the vertical display visualizes the region that is currently displayed on the table. Digital pen & paper is used on the non-interactive rim of the table for supporting individual ideation and reflection (*personal spaces,* see Fig. 2, c). Personalized interactive areas at the corners of the table (*transfer spaces,* see Fig. 2, d) provide a zone for additional individual actions. The coupled vertical display can also be used for dynamic views on selected artifacts during phases of discussion (see Fig. 2, right).

3.2 Interaction Techniques

AffinityTable supports the workflow of affinity diagramming with hybrid interaction techniques for copying, clustering, piling, and collecting. Additional functionalities that augment physical practice are integrated with interaction techniques for highlighting, focusing, searching, and image retrieval. We aimed at incorporating tangible elements for improving visibility of actions and group awareness.

Copying. In physical practice, notes are created and reflected individually in personal workspaces. *AffinityTable* preserves this ability by providing small paper notes and digital pens. Handwriting and sketches are instantly transferred to the system via a wireless connection. Physical notes can be copied to the shared workspace by placing them on the interactive table (see Fig. 2, center). A digital copy of the physical note appears within the virtual workspace, while the physical copy remains in the personal workspace. This technique allows sharing and presenting of artifacts without physical restrictions while still preserving the ability to reflect upon all individual contents in the personal workspace throughout the process (DG1,DG2). It also enables rapid transfer between the personal and the shared workspace (DG3,DG4).

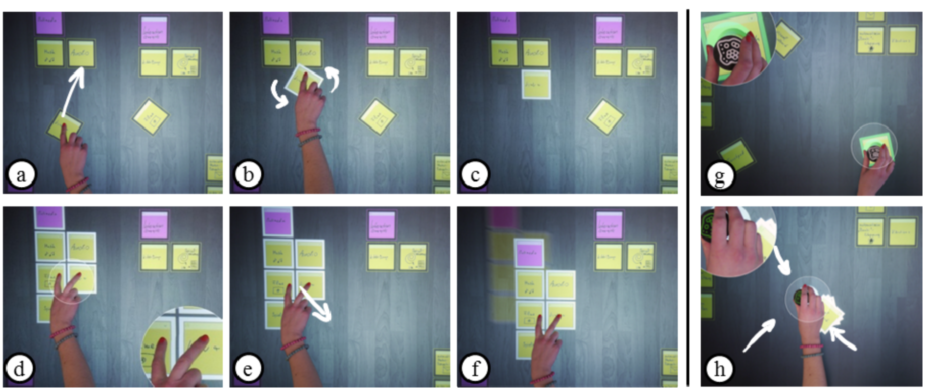


Fig. 3. Clustering notes with touch manipulations (a,b,c); moving clusters of notes by using multiple fingers (d,e,f); collecting notes by using colored tokens (g,h)

Clustering and Piling. During convergent activities, notes are discussed and organized into meaningful arrangements. Once copied into the shared workspace, digital notes can be dragged, rotated and flicked by using multi-touch manipulations on the interactive table. It is also possible to change the color of notes or to delete notes, which are basic requirements for clustering. To facilitate the collaborative organization of notes into groups, we included a simple clustering algorithm that automatically aligns and associates notes when released close to each other (see Fig. 3, a,b,c). Piles of notes can be created by releasing notes on top other notes. Clusters and piles can be moved by dragging them with multiple fingers (see Fig. 3, d,e,f). These interaction techniques closely resemble collaborative sorting actions at the whiteboard (DG5). However, since all user action is focused on a horizontal surface, which is accessible to all participants, this might lead to more equal access and increased awareness of actions within the group.

Collecting. When collecting notes or clusters from arbitrary locations within the shared workspace, it is a tedious task to move them individually via touch manipulations across longer distances in the virtual workspace. In physical practice, notes can be collected into mobile piles (see Fig. 1, c) and thus can be easily transferred to remote locations. *AffinityTable* replicates this ability by providing multiple two-sided tokens. Users can select digital notes by using the colored side of the tokens (see Fig. 3, g). The notes are then marked in that color. By turning over the tokens and placing them on the table with the blank side, selected notes are moved to this new location (see Fig. 3, h). Because multiple tokens can be used at the same time, this technique also enables loosely coupled parallel work. The tangibility of the tokens and the colored marking used as a proxy for mobile piles may increase the visibility of these parallel actions.

Highlighting and Focusing. During discussion, deictic references on notes are often used as a form of communication and for coordination in the group (see Fig. 1, d,e). *AffinityTable* augments these interactions by providing additional functionality. When touching digital notes, they are highlighted around their border with a glowing effect. Each highlight fades out after five seconds. When clustering notes, the fading glow implicitly communicates a history of actions. In addition, a token can be used to focus on a specific region within the workspace during reflection (DG3). By turning the token (see Fig. 4, a), the users can change the zoom-factor of the magnified region displayed on the vertical display (see Fig. 2, right). We chose a single tangible element as focusing tool because we believe its limited accessibility may stimulate negotiations about the focus of discussion.

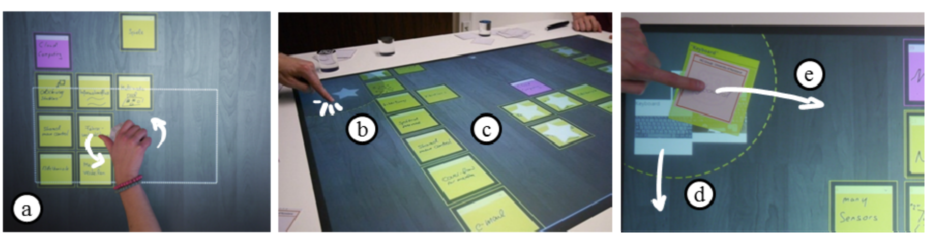


Fig. 4. Focusing on a specific region in the workspace (a); searching for notes (b,c); retrieving and adding images based on the content of notes (d,e).

Searching. *AffinityTable* provides a lightweight search function based on the author of notes. By tapping personal symbols that are available in the *transfer spaces* at the corners of the table (see Fig. 4, b), the notes created by the corresponding user are marked with that symbol (see Fig. 4, c). This allows participants to look up the responsible author of a particular issue or idea during discussion. It also allows reflecting on participants contributions to certain clusters and therefore may provide information about consensus in the group (DG3). This search function may introduce positive effects by reducing free riding, but may also have negative effects by increasing evaluation apprehension. However, since notes are anonymous throughout the session and identity is only revealed on demand, this technique may stimulate individual contributions without putting too much pressure on participants.

Image Retrieval. When presenting ideas and issues it can be helpful to add additional information to make a clear point. Therefore, *AffinityTable* provides image retrieval based on the content of notes. When placing notes into *transfer spaces* at the corners of the table, five related images from the web are displayed around the physical note. Users may then drag digital images along with a copy of the physical note to the shared workspace (see Fig. 4, d,e). Images in the workspace can be clustered along with digital notes. Hence, they can also be used to visually point out important clusters by adding representative images. This technique may have positive influence on fixation effects because the selection of images can be used to further open up divergent activities. Instead of using images from the web, this technique may also be used to augment notes with other data from custom design knowledge repositories.

3.3 Implementation

AffinityTable runs on a custom-built interactive table that measures 128cm x 157cm x 105cm. The non-interactive rim (20cm) was designed for supporting leaning, personal workspaces and for avoiding involuntary interaction. The table provides adequate space for up to five users. The graphics are rear-projected with a WUXGA projector (1900px x 1200px). Multi-touch and object recognition is provided with IR illumination and three tiled XGA cameras (1024px x 768px). The vertical display measures 76cm x 162cm and features 4K resolution (4096px x 2160px). The software runs on two networked workstations with MS Windows 7. The interface is implemented using the ZOIL framework [11] that integrates WPF multi-touch controls with distributed user interface synchronization and zooming functionality. We use ReacTIVision[1] for marker tracking and Squidy [12] for finger tracking. Data from Anoto's ADP-301 digital pens is received via Bluetooth by using the Anoto streaming API[2]. Text recognition is implemented using MS Windows 7 SDK. We use the Yahoo image search service for retrieving digital images.

4 Conclusion

In this paper we presented *AffinityTable*, a hybrid surface for supporting affinity diagramming. Based on an observational study, we identified design guidelines for a close fit to embodied practice. This knowledge was used for the design of a

---

[1] http://reactivision.sourceforge.net
[2] http://www.anoto.com

workspace that combines digital pen & paper with an interactive table and tangible tools for supporting the basic workflow of the design technique and for providing additional functionalities. A vertical display is used for supporting reflection-in-action and for enhancing discussion and coordination. The described interaction techniques can be adapted for similar design techniques that share the phases of idea-generation, idea-presentation and idea-discussion. However, this paper does not provide empirical evidence on the positive effects of the proposed interaction techniques. In future work we will improve *AffinityTable* by adding more powerful clustering techniques. Eventually, a user study is necessary to gain a deeper understanding of the tool's effects on the process of the design technique, efficiency and social factors.

Acknowledgements. This research is funded by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft) under grant number RE 1843/3-1.

References

1. Beyer, H., Holtzblatt, K.: Contextual design: defining customer-centered systems. Morgan Kaufmann, San Francisco (1998)
2. Klemmer, S.R., Newman, M.W., Farrell, R., Bilezikjian, M., Landay, J.A.: The Designers' Outpost: A tangible interface for collaborative web site design. In: Proceedings of UIST 2001, pp. 1–10. ACM Press, New York (2001)
3. Haller, M., Leithinger, D., Leitner, J., Seifried, T., Brandl, P., Zauner, J., Billinghurst, M.: Shared design space: sketching ideas using digital pens and a large augmented tabletop setup. In: Advances in Artificial Reality and Tele-Existence, pp. 185–196. Springer, Heidelberg (2006)
4. Hartmann, B., Morris, M.R., Benko, H., Wilson, A.D.: Pictionaire: supporting collaborative design work by integrating physical and digital artifacts. In: Proceedings of CSCW 2010, pp. 421–424. ACM Press, New York (2010)
5. Hilliges, O., Terrenghi, L., Boring, S., Kim, D., Ritcher, H., Butz, A.: Designing for collaborative creative problem solving. In: Proceedings of C&C 2007, pp. 137–146. ACM Press, New York (2007)
6. Tse, E., Greenberg, S., Shen, C., Forlines, C., Kodama, R.: Exploring true multi-user multimodal interaction over a digital table. In: Proceedings of DIS 2008, pp. 109–118. ACM Press, New York (2008)
7. Kirk, D., Sellen, A., Taylor, S., Villar, N., Izadi, S.: Putting the physical into the digital: issues in designing hybrid interactive surfaces. In: Proceedings of BCS HCI 2009, pp. 35-44 (2009)
8. Jacob, R.J.K., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., Zigelbaum, J.: Reality-based interaction: a framework for post-WIMP interfaces. In: Proceedings of CHI 2008, pp. 201–210. ACM Press, New York (2008)
9. Cook, D.J., Bailey, B.P.: Designers' se of paper and the implications for informal tools. In: Proceedings of OZCHI 2005. ACM Press, New York (2005)
10. Vyas, D.: Artful surfaces in design practices. In: Proceedings of CHI EA 2009, pp. 2691–2694. ACM Press, New York (2009)
11. Jetter, H.-C., Gerken, J., Zöllner, M., Reiterer, H.: Model-based design and prototyping of interactive spaces for information interaction. In: Proceedings of HCSE 2010, pp. 22–37. Springer, Heidelberg (2010)
12. König, W.A., Rädle, R., Reiterer, H.: Interactive design of multimodal user interfaces - reducing technical and visual complexity. Journal on Multimodal User Interfaces 3(3), 197–213 (2010)

Design as Intercultural Dialogue: Coupling Human-Centered Design with Requirement Engineering Methods

Chiara Leonardi, Luca Sabatucci, Angelo Susi, and Massimo Zancanaro

Fondazione Bruno Kessler (FBK-irst), via Sommarive 18,
38123 Povo, TN, Italy
{cleonardi,sabatucci,susi,zancana}@fbk.eu

Abstract. In the design of information technologies, the challenge of integrating a human-centered design approach with software engineering methods emerge in different forms. The main challenge is to set the ground for different disciplines and professional cultures communicate and work together. The orchestration of different contributions and the establishment of communication practices that facilitates the integration of the different languages and procedures are crucial steps to take full advantage of different research traditions. This paper presents a case study in which human-centered design and requirement engineering methodologies have been used within a large research projects aiming at developing innovative technologies and services to support professionals in nursing homes. The design process took the form of an intercultural dialogue that required human-centered and requirement-engineering professionals to work across borders. Starting from our case study, the paper presents the boundaries identified between the human-centered and the engineering perspective and proposes a framework to guide the integration process conceived as an intercultural dialogue between disciplines.

Keywords: Human-centered design, Requirement engineering, inter-disciplinarity.

1 Introduction

In recent years both human-centered and requirement engineering communities are facing the challenge of identifying ways to combine and integrate their respective approaches in order to cope with the growing complexity of systems and infrastructures to be designed. From one side, a human-centered approach facilitates the design of technologies that best fit users practices and therefore are less likely to be rejected; on the other side, requirement engineering approaches ensure an efficient and cost-effective way of designing these technologies.

The need of combining the two approaches is particularly crucial within the field of ubiquitous, context-aware computing because for this kind of technologies interaction of technical, organizational, social and ethical issues is posing serious challenges to requirement gathering and user-assessment of the technological solutions proposed [1]. Furthermore, all those aspects should be tackled looking at the technology from two levels, namely, the level of services and the architectural level [2].

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 485–502, 2011.

The issue of requirement gathering and management in this context is crucial and Lyytinen et al. [2] well synthesize the challenge:

"requirements capture is not just how to identify and incorporate technical requirements, but how to foresee and integrate social and organizational requirements with those capabilities. These are often hard to identify before actual deployment of the ubiquitous computing service. In order to successfully design and implement ubiquitous computing, actors who were separated in time and space in the development cycle need to establish much tighter coordination patterns. The development methods and processes for ubiquitous computing environments will need to reflect such new needs".

Both requirement engineering and human-centered design address this challenge from different points of view and the opportunities for cooperation are widely acknowledged.

Yet, several issues remain unexplored and a number of questions are still open.

The two fields have their own established methods and practices to organize knowledge and to define and communicate requirements as well as different vocabularies. Moreover, they entail different views concerning the validity of methods and hence concerning the final results of the research.

The issue of how to combine these two approaches has been recently discussed by several authors. Kaindl [3] poses the issue concerning the usefulness of common representations for requirements engineering and interaction design and the possibility and the convenience in developing a combined process. Nebe et al. [4] also argue that a fruitful cooperation is possible if we first identify integration points between disciplines.

In this paper we try to address some of the challenges posed by a joint use of the two approaches within the same design process. In particular, we discuss a case study in which a "dialogic" relationship between the disciplines was sought that allows researchers to cooperate and share knowledge taking advantages of the reciprocal strengths of the two approaches.

Our case study investigated the design of smart environment in nursing homes at support to the medical and assistance staff as part of a larger research project focused on intelligent environments. The ACube system employs a network of sensors distributed and embedded in the environment or embedded in the users' clothes to allow an efficient and unobtrusive monitoring of the nursing homes guests. The ACube project team was multi-disciplinary and included software engineers, sociologists and interaction designers. Representative end users were also directly engaged in design activities.

For the design phase, two teams were involved, one with experience with the qualitative-oriented methodologies of interaction design (in particular the use of personas and scenarios) and one with experience in the use of the semi-formal approach to requirements engineering called Tropos, a goal-oriented software engineering methodology that allows to describe the domain and the stakeholders' needs and to reason on concepts such as actors, goals and social relationships [20]. The initial attempt to define a common vocabulary for requirements engineering and human-centered design teams was the source for philosophical and methodological discussions that eventually culminated with the feeling that something different should be tried.

That confrontation was not useless: the team reached the awareness that the true problem is, first of all, epistemological and requires a serious consideration and reflection on the assumptions behind the two different approaches. It was first of all clear that a combination of the two approaches was necessary but this combination should have not to be sought in terms of reducing either one approach to the other. Our strategy, thus, was articulated around three main points:

1. to identify strengths and limits with respect to the specific design objective, in order to define integration points between human-centered and Goal-oriented requirement engineering methods and take full advantages of their reciprocal strengths;
2. to identify main barriers that make the dialogue difficult and inefficient;
3. to define communication protocols and tools to make the dialogue effective.

The design process took the form of an intercultural dialogue that required human-centered and requirements engineering professionals to work across borders.

Starting from our case study, the paper presents the boundaries identified between the human-centered and the engineering perspective and proposes a framework to guide the integration process conceived as an intercultural dialogue between disciplines.

2 Defining the Relationship between Disciplines

Several experiences are concerning the integration Requirement Engineering and human-centered design techniques. Sutcliffe and Maiden [5] propose a framework for relating scenarios to use cases: scenarios are automatically generated from use cases and are validated by rule-based frames able to detect problematic event patterns. Lamsweerde [6] and Rolland [7] claim a bi-directional relationship between goals and scenarios: goals lead to the elaboration of scenarios and scenarios may prompt the elicitation of goals or may be used for goal inference. Uchitel and Chatley [8] combined the use of goal oriented requirement elicitation and scenario and discussed some techniques for validation aims. Cockburn [9] suggests the use of goals to structure use cases by connecting every action in a scenario to a goal of another use case at a lower level of abstraction. In this sense a scenario is built each time a goal is discovered. Yu and Liu [10] propose a technique for supporting reasoning on social models by means of scenarios. They use scenarios expressed in Use Case Maps to describe elaborated business processes or workflow. Aoyama [11] uses both scenarios and provisional personas in order to identify a set of requirements for a web application in which the kind of users is wide and difficult to classify.

The problem of integrating methods coming from the human-centered and requirement engineering research areas without losing the strengths of the two approaches can be tackled, in our view, once we recognize that the main challenge is to make the setting for a dialogue between disciplines. The first step is hence the understanding of the relationships and boundaries between the different approaches and the opportunities for combining them. Traditionally, two main trends can be identified for composing research approaches of different nature: from one hand, there is the tendency to develop overarching and unifying frameworks [12], on the other hand the effort to make different disciplines speaking to each and collaborate across borders.

2.1 Limits of Unifying Frameworks

In the case of unifying framework, one disciplinary approach is usually modified to be assimilated into the other approach, or several approaches are modified to fit into an overarching meta-model. While the risk is to limit the potential of the approach itself; the advantage is to work in a situation of 'methodological purity' where a certain epistemology has been chosen to drive the process.

In software engineering, for instance, the Situational Method Engineering [13, 12] is grounded on the assimilation approach: constructing ad-hoc software engineering processes by reusing fragments of existing design processes; the basis for the assimilation technique is the method fragment [14], a self contained component that can be used as building blocks for the process composition. Techniques for fragment manipulation (extraction, selection, and composition) are still open points, and even if there is a disagreement about the level of precision, it is clear that fragment specification requires a language for describing at least the process and its products. Some recent approaches [15] make use of the SPEM notation[1] for describing the process as a workflow and meta-models as linguistic keys for bridging activities and artifacts coming from different methodologies [16]. In particular the meta-model approach demonstrated working fine in the context in which the integration concerns two engineering approaches that are based on the same epistemological nature. In the field of human-computer interaction also there have been several attempts to integrate the different disciplines and approaches composing the fields (activity theory, ethnography, distributed cognition, etc.). As stated by Rogers [17]:

"a problem with integrating such different theories and ontologies, however, is that it becomes difficult to know what frames of reference and axioms to use for a given problem space. (…) It seems that only the researchers who have developed the grand theories are able to use them".

Rogers argues that the attempt to develop overarching frameworks has the advantage of break away from the confines of specific disciplines. On the other side, the risk is to have inefficient and difficult to apply framework that can suffer from the "toothbrush" syndrome, that is relevant for the one that developed the theory but, unfortunately, for no one else [17].

2.2 Toward a "Dialogic" Perspective

By "dialogic", we mean a perspective of interaction that is grounded on communication and iterative confrontation. Differently from what happen with unifying frameworks, this

---

[1] SPEM – Software & System Process Engineering Metamodel Specification, Version 2.0 available online at http://www.omg.org/spec/SPEM/2.0/

perspective recognizes the irreducible cultural difference between the two approaches and therefore asks practitioners to work in a situation of epistemological and methodological pluralism: the goal being that not to transform or to assimilate an approach to make it fits into another one, but rather to bridge the gap between different research traditions to take advantages of their mutual strengths. In this case we have a dialogic process, where various approaches coexist and where each approach can hold more relevance depending on the specific problem to be addressed and on the phase of the research.

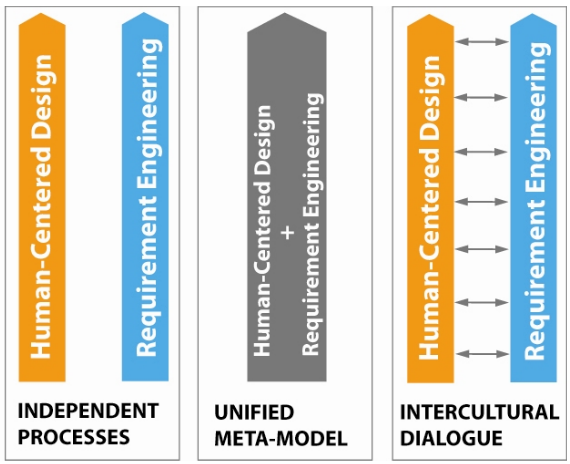


Fig. 1. Possible models of relationships between Human-Centered Design and Requirement Engineering

A dialogic perspective, therefore, does not require transforming and adapting the methods specific to each research tradition but requires that preconditions for a beneficial dialogue between the two are defined and set. The challenge here is to manage the dialectic issue concerning the concurrent usage of different research paradigms: an engineering approach such as Tropos and less formalized and qualitative approach such as the human-centered ones.

Even if the approach seems promising, it raises several issues. In particular, the different epistemological and methodological traditions have to be first understood by practitioners of both disciplines in order to be integrated.

Studies on cross-disciplinary and interdisciplinary research are helpful to understand the challenges of the collaboration between disciplines. Eigenbrode et al. [18] argue that the main challenges of cross-disciplinary research are related to the linguistic and conceptual divides and to the perceived nature of the world: researchers may have different views on what is reality and how reality can be known (e.g. constructivism vs realism). The authors posit that these are mainly philosophical challenges related to conflicting assumptions and propose a toolbox for philosophical dialogue, consisting of a set of questions designed to encourage confrontation on scientists'' philosophical views on their research.

Strober [19] also illustrates how the collaborative process is neither easy nor intuitive. Fundamental barriers are not only linguistic, but cultural. While translations are pretty an easy task, what is much more difficult is to understand and accept assumptions, methods and ways of validating the truth employed in other disciplines.

In the following, we present from our case study, the process that emerged from the confrontation of the two research teams, the challenges that rise during the process and finally propose some lessons learned from this experience.

3 Tropos: A Goal-Oriented Requirement Engineering Method

For the sake of clarity, we introduce here the Tropos method, a semi-formal methodology [20] that relies on a set of concepts, such as actors, goals, plans, resources, and dependencies to formally represent the knowledge about a domain and the system requirements. An actor represents an entity that has strategic goals and intentionality within the system or the organizational setting. An actor is used to model both human stakeholders and software and hardware systems. Goals represent states of affairs an actor wants to achieve. Executing a plan can be a means to realize a goal. Actors may depend on other actors to attain some goals or resources or for having plans executed (see Figure 3 for an example). Tropos models are visualized through actor and goal diagrams. The former are graphs whose nodes represent actors and arcs are strategic dependencies between pairs of actors. A goal diagram represents an individual actor perspective in terms of its main goals, and their decomposition into sub-goals. Furthermore, plans and resources that provide means for goal achievement are depicted through means-end relationships.

The semi-formal nature of the Goal-Oriented languages gives, in general, an effective way to communicate with the technical staff and carries some other advantages. In particular: it allows for a structured representation of the requirements; it allows to specify automatic and semi-automatic conflict analysis procedures to refine the set of requirements and to highlight, as soon as possible, requirements inconsistencies; finally, it allows to establish a clear traceability between the motivations for the existence of a given requirement and the requirement itself by linking it to the domain needs described via the goal models.

Tropos distinguishes five phases in the software development process: Early Requirements, where the organizational domain is described, Late Requirements, where the future system is introduced in the organization, System Architecture Design, System Design and System Implementation.

The use of the Tropos methodology enables structuring the domain knowledge in terms of actors, goals social dependencies and other concepts that are sharable with the UCD team. Moreover, the five phases process allows a systematic tracing of the domain knowledge towards the system implementation artifacts, with the result of establishing a bridge between requirement elicitation (in collaboration with UCD techniques) and system deployment. Specifically, in the ACube project, we applied the first two phases of the Tropos methodology, in order to sketch the strategical dependencies among the actors participating to the elderly setting, and to anticipate the impact of introducing the ACube system into the organization.

4 Case study: The Design of ACube

ACube[2] is a large research project aiming at designing a highly technological environment to be deployed in nursing homes as a support to medical and assistance staff. The project envisages a network of sensors distributed in the environment or embedded in users' clothes that should allow monitoring the nursing home guests without influencing their usual daily life. Advanced automatic reasoning algorithms allow acquiring the data through the sensor network and to promptly recognize emergency situations or prevent possible dangers for the guests themselves. An important activity in the project was the analysis of the system requirements with the objective of having a quality improvement of services in the nursing home.

The whole requirements elicitation and analysis process within ACube was articulated in seven phases all of them characterized by the use of both the approaches (see Figure 2).

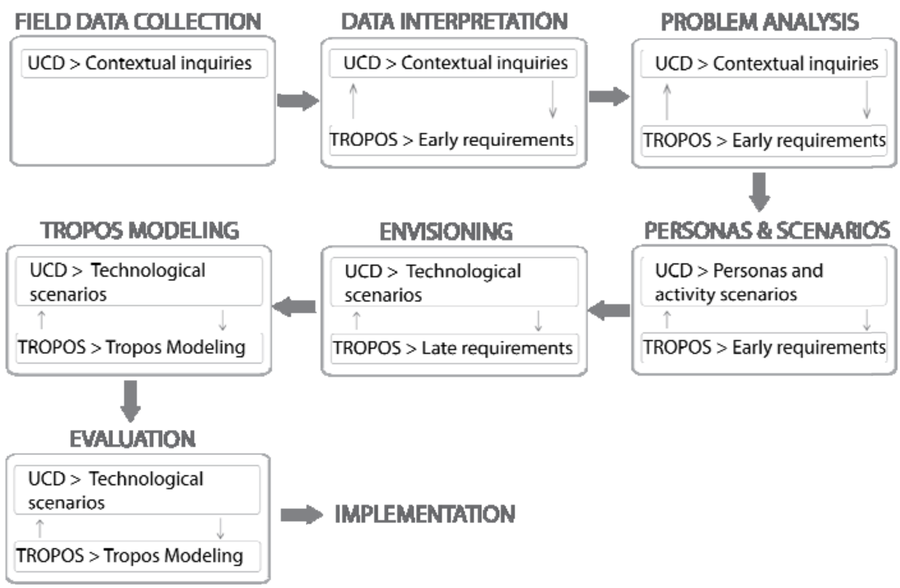


Fig. 2. The project whole cycle

Phase 1 – Field Data Collection. The process started with the investigation of the domain to understand the organizational setting in four nursing homes and to identify the needs of the involved stakeholders. First, users are classified in: primary users, which are direct users of the product (e.g., caregivers); secondary users, which are those who will occasionally use the artifact or those who use it through an intermediary (e.g., managers); and tertiary users, which are persons who will be affected by the artifact or influence its success (e.g., patients, family members). Several human-centered design

[2] ACube is funded by the local government of the Autonomous Province of Trento in Italy.

methods exist in order to get rich insights about the context to be explored. Recently contextual inquiry [21] demonstrated the capacity to satisfy the needs for a deep and, at the same time, quick understanding of complex domain. Contextual inquiry mainly consists in interviewing people in their context, preferably when performing their tasks. Contextual inquiries were performed in 4 different nursing homes and involved about 40 health professionals.

Phase 2 – Data Interpretation. The data interpretation is the step in which data coming from the domain is shared across the team and becomes knowledge. In our process, data interpretation is concurrently carried out in a twofold way: i) following contextual design approach dimension (flow model, sequence model, cultural model, artifact model, physical model) [21] and ii) exploiting the early identification phase of Tropos. The Early Requirement process is used for modeling the initial set of domain entities. It includes the Actor diagram and the Goal diagram. The Actor Diagram provides a bird-eye view over the domain, by specifying how responsibilities are distributed across roles, and providing an intuition of which interactions occur in the environment. An actor diagram is a graph whose nodes represent actors (agents, positions, or roles), while edges represent dependencies among them.

Phase 3 - Problem Analysis. The analysis of critical aspects was developed to highlight main problems that professionals of nursing homes experience in their job. The Criticality Identification is an activity that bridges the Tropos analysis with the following *personas* and scenario authoring. The aim is to highlight every possible breakdown or problem that may occur in the organization that hinders the achievement of goals, such as the situation related to the intervention for a fall of one of the patients. A criticality connects an exceptional event to front with Tropos goals and tasks that are identified to receive a negative contribution. It also encapsulates the context in which the problem may occur. A criticality could be identified in the documents produced by contextual inquiry, by considering breakdowns in users' job and procedures or in the use of artifacts, and it is documented by putting together the narrative description with goals of the model. Each criticality represents a view over the organization model that focuses on highlighting users, goals and activities when an emergent problem occurs. Criticality identification triggers the subsequent consolidation and envisioning steps, by providing motivations for the introduction of a system into the domain. Not all entries get the same importance, thus a prioritization is important for reducing the risk to focus on marginal problems of the domain, or infrequent situations. Critical aspects that may be addresses through a technological intervention, such as the problem related to the fall, were represented through Tropos Early Requirements diagrams and led to activity scenarios and personas (in which technology and services were not yet present).

Phase 4 – Personas and Activity Scenarios Design. We adopted the specific scenarios approach as developed by Rosson and Carroll [22] and subsequently enriched by Copper with the notion of *personas* [23], Personas are rich descriptions of archetype users meant to draw attention on users' goals and motivations [24].

Introducing *personas* in scenarios-based approach provides an anchor against self-referentiality in design and make scenarios more concrete. A set of activity scenario-descriptive scenarios focusing on problems and criticalities identified through user studies – were developed. The starting point for the Phase 5 is constituted by 5 personas and 5 activity scenarios.

Phase 5 – Envisioning, from Data to Design. The phase of envisioning moved from personas and activity scenarios identified in Phase 4 in order to envision how to introduce the technology in the analyzed domain.

A participative workshop was organized to identify how the system could support the critical situations identified in Phase 2 and hence develop technological scenarios to be further evaluated with end-users representatives and stakeholders. About 10 participants attended the workshop including the ID team and representative of stakeholders and technologists. The heterogeneity of the group was meant to guarantee the generation of creative but feasible ideas, to provide concrete solutions to problems identified by nursing homes professionals as well as to provide solutions that could meet engineers' expectations and their research interests. Outcomes were pursued at multiple levels: to expand the designer's prospective and to watch the problems from different points of view, to figure out how their ideas can work in a real context, to identify design criticalities and open issues, to generate requirements of the system-to-be. The workshop ended with the definition of 5 different macro-services the ACube system might provide.

As a consequence of the envisioning focus group, and the introduction of the system into the organization, the Tropos process moved from the early requirement phase to the late requirement phase. Figure 3 shows an excerpt of the Tropos model, describing a small part of the goals and the activities of the *SeniorOSS* actor (a caregiver in the nursing home). In particular the actor *Senior OSS* has the goal to [`avoid dangerous behavior of patients`] that can be AND decomposed in [`monitor patients in her visual area`] and [`coordinate interventions in the nursing home area`].

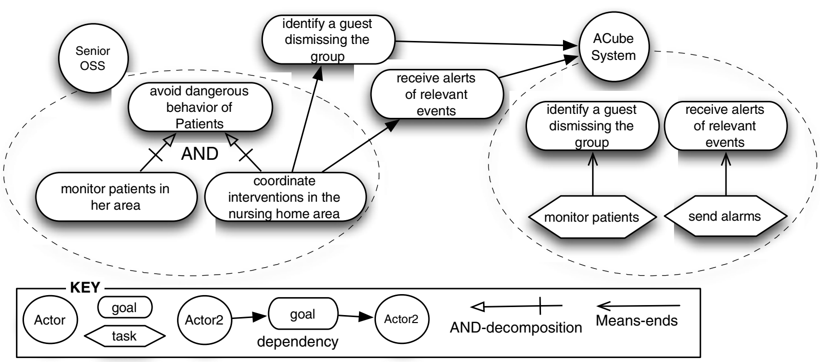


Fig. 3. An excerpt of the Tropos model for the nursing home

Phase 6 - From Design Ideas to Tropos Modeling. Tropos diagrams and scenarios were jointly used to refine the ideas emerged during the creative workshops. On one hand, technological scenarios were designed to make design ideas concrete and to trigger reflection about possible services. On the other hand, Tropos diagrams were developed to more systematically analyze how the introduction of a system impacts on the domain actors. In particular, Tropos allows for describing the relationships between the stakeholders and the system, represented as a new domain actor, in terms of actor dependencies.

This way, the envisaged services provided by the system are described within their operational environment, along with relevant functions and qualities. Modeling dependencies between actors and the system introduced in the organizational setting provides a more rigorous view on the envisioned services and helps designers in understanding critical dependencies between actors and goals. Despite scenarios and *personas* are very effective tools to show particular instance of a problem and to figure out design solution, they do not provide adequate support for abstract thinking. Tropos, on the contrary, provides an exhaustive view on the domain that is crucial to go back from specific design ideas to abstract requirements, to consider alternatives when designing novel services, and to balance the impact that each design decision have on non-functional requirements of the system. In addition the formalization of requirements supports automatic and semi-automatic procedures for the verification of completeness and conflict analysis.

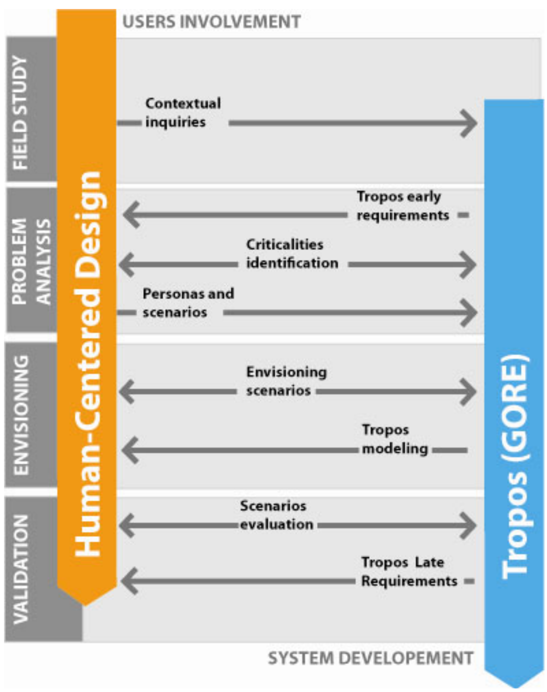


Fig. 4. The relationships between HCD and RE for our case study

Phase 7 – Evaluation of Technological Scenarios. Here visual scenarios were derived from the Tropos models and used for the validation phase, where multiple views on the domain are required to drive the negotiation and refinement of requirements with stakeholders and project partners. If envisioning scenarios provided a concrete instance of a particular design solution, that is very helpful to discuss with stakeholders, on the other hand, Tropos diagrams aided designers in reconsidering design solutions and elaborating alternatives thanks to the possibility to trace back design solution to initial abstract requirements.

A first focus group was held with the representatives of the 10 research groups involved in ACube project, 27 people attended the meeting. The second focus group was organized with the stakeholders, 3 managers of nursing homes previously involved in the early exploration phase attended the meeting. The goal of these meetings was to assess of the validity, acceptability and feasibility of requirements and to envision alternatives not considered in the scenarios.

The workshop with technological partners was focused on technical feasibility and research interest, and on the envisioning of original solutions to the critical situations identified. Acceptability and usefulness were instead the pivots of the workshop with end-users representatives. Expected outcomes of the two focus groups were the emergence of design criticalities, the resolution of open problems and the identification of new propositions and ideas, in order to collect additional elements to elaborate an organic description of the technological architecture.

The output of this phase was the agreement on early requirements and the refinement of Tropos late requirement diagrams in order to choose the main functionalities of the system, such as the need of having ACube monitoring critical events such as the fall of the patients (see Figure 5).

Fig. 5. An example of scenario used for the validation phase

In the following sections we reflect on the role of the two approaches and identify the boundaries existing between them.

5 Reflections on the Experience

After the design experience, we could assess our initial strategy based on the three main mechanisms that we employed to shape the relationships and set the dialogue for an efficient cooperation between the two teams (see section 1):

- *Strengths/limits analysis*: identify strengths and limits of the methods while achieving a given design objective in order to define integration points between HCD and RE methods and take full advantages of their reciprocal strengths;
- *Barriers identification and exploration*: identify main barriers that make the dialogue inefficient;
- *Mutual learning*: enable continuous information exchange and communication to overcome the barriers coming from the inter-cultural and multi-perspective approaches to facilitate the negotiation of the meanings of terms and the sharing of common modeling tools during the requirements elicitation and analysis activities.

In the following, we recap and discuss both what we realized during the actual work as well some new and unexpected issues in an attempt to reason about and abstract our experience to highlight problems and proposing guidelines.

5.1 Exploring Reciprocal Strengths and Limits

The first step requires identifying strengths and limits in order to define integration points between HCD and RE methods and take full advantages of their reciprocal

Table 1. Strengths/limits analysis

Tropos Modeling	Human-Centered Design
Abstract representation of the domain	Concrete representation of the domain
Semi-formal representation of requirements	Informal/narrative representation of the requirements
Static and invariant picture of the domain	Dynamic representation involving the spatio-temporal dimension
Do not provide specific tools for finer prioritizing requirements than the reasoning on alternatives and contributions	Scenarios provide a support for prioritizing requirements
Neutral representation that do not engender an emotional response	Personas-based scenarios are sort of “dramatic” representation that engenders empathy
Do not provide information about the physical context	Provide details about the physical context in which people act
Provides a general representation of invariant dependencies among actors	Provide details about how interactions occur in a given specific situation
Support requirement traceability	Do not support traceability
Support automatic requirements conflict-analysis and check of quality properties	Do not support conflict-analysis
Provide tools to efficiently communicate and validate requirements with technical staff	Provide tools to efficiently communicate and validate requirements with end-users and non-technical partners

strengths. This analysis was beneficial for the team because promoted mutual learning between practitioners and was a starting point for identifying integration points as a basis to define the research process and the methods to be exploited in the different phases of the process. The output of this analysis is synthesized in Table 1. The awareness of the differences is a first step to further define and overcome existing divides between approaches.

5.2 Making the Divides Explicit

Once reciprocal strengths and limits of the two approaches were identified, an exploration of the barriers that can prevent a synergy between the two approaches was pursued. As studies in the field of cross-disciplinary research suggest, an effective collaboration between disciplines entails that differences are explored and assumptions made explicit. We identify two main challenges that researchers should face in order to establish a dialogue: the epistemological and conceptual barrier.

Epistemological Divides. The first main issue is to consider epistemological foundations and validity criteria of both the approaches, to manage differences without weakening and distorting the two research paradigms. While requirement engineering – and hence Tropos - is grounded on a positivist research tradition [25], several methods employed in HCD origin within a constructivist and interpretative tradition (ethnography, ethnomethodology, etc.).

Positivism is an epistemological perspective that claims the knowledge is based on sense experience and positive verification. One of the key features of positivism is the ability of demonstrating the logical structure and coherence of a concern by axiomatization. Tropos is classified as a positivist approach (even if the debate on the positivist nature of many RE methods has recently been criticized [26]) by providing a precise frame for the modeling activity and the reasoning process. For instance, typical instruments for modeling the domain are: (i) the abstraction is used to reduce the complexity of the representation of the domain instance to the eyes of an observer, by focusing only on specific domain properties that are considered central to the analysis; (ii) decomposition that is used to break the complexity of an entity thus to separately study its components when moving from generic towards specific; (iii) refinement is used to derive specifications, by using argumentation and logic deduction, based on the analysis of existing data.

Constructivism recognizes that our knowledge is socially constructed and it does not reflect an external "transcendent" reality. Knowledge is hence contextual and contingent on convention and social experience.

The exploration of the epistemological divide of the two research tradition is crucial of we consider, for instance, the two different perspectives on the nature of requirements: while in RE tradition requirements often "exist out there" and the role of the analyst is to capture them, manage them, and validate them; on the other side, within the HCD approach, requirements are the output of a negotiation that happens during the whole design life-cycle [27]. Requirements are constructions produced by a number of actors (users, analysts, developers, designers) each acting in specific context and each having specific motivation and perspectives on the system to be designed.

Linguistic and Conceptual Divides. RE and HCD approaches exploits very specialized dictionaries and concepts. Often, the same terms (e.g. the concepts of 'role', the concepts of 'requirement') have very different connotations and entails different methodological choices. One of the more discussed divides concerns the view and the consequent treatment of requirements [27]. If we define a requirement as something to be captured or extracted from the users certain methods should be preferred, on the other hand, if we define requirements as something to be collaboratively constructed other methodological choices may be preferred. The conceptual definition in this case has important consequences on the whole design cycle.

A linguistic boundary is due to a mismatching in the dictionary used in the two methodologies. This aspect is specifically evident in the integration between a semi-formal language (Tropos) and an informal, often verbose and intentionally ambiguous language. The identification of these linguistic boundaries is important for the reconciliation of incompatible concepts and for creating the framework for knowledge sharing.

An example of linguistic boundary is the Tropos 'task' and the human-centered 'activity' terms. A Tropos task is defined as the conceptualization of a plan that provides the means for the operationalization of a goal. An example of task is [caregivers monitor guests' behavior]. The human-centered design activity concept captures additional information about the context in which it is carried out, including the user point of view and empathy:

"… during my job it is important to continuously observe patients' behavior, but this is often an heavy activity to carry on together with other our duties. This is due to the high number of guests compared to the low number of professionals. This working overhead causes we are incapable of concentrating on the human aspect of our job as well as we would do … " (extracted from an interview with a caregiver).

It is clear that it contains some cues (for example the frustration about the difficulty of giving priority to the human aspect) that are difficult to encode in a plan-based representation but indeed represent important information for the design.

Another example is represented by the concept of 'persona' with respect to Tropos' actor. Whereas both of them identify users of the future system, an actor is a way to abstract a role in the organization, while a persona is an archetype of user, sufficiently concrete to provide the understanding within a scenario. The cognitive and emotional dimensions are important factors which persona tries to catch for helping the designer to take decisions in the design process, characteristics that are missing in an actor.

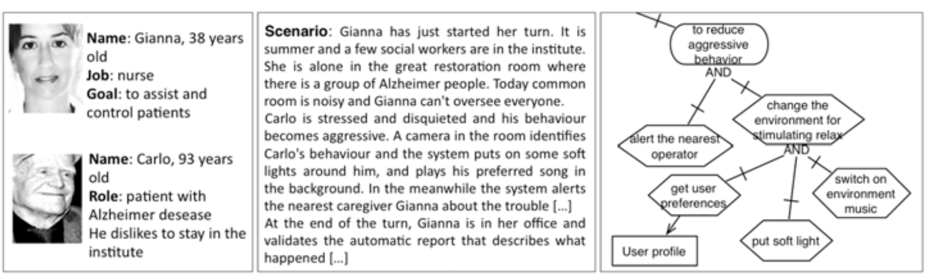


Fig. 6. Examples of artifacts: personas, technological scenario and Tropos late requirement diagram

A similar problem emerged with the definition of ‘role’. Within Tropos, a role is defined as an abstract entity that is a holder of responsibilities within the domain, owns strategic goals and may execute plans to achieve them. In Tropos roles may be bounded by dependencies, relationships that establish *depender*, *dependee* and *dependum*.

For human-centered researchers, role is a more fluid and ambiguous concept, encompassing formal and informal responsibilities. For instance, tacit responsibilities of caregivers working within a nursing home are those related to the human aspect of their work, as for instance, communicate emotional closeness to guests.

5.3 Mutual Learning

The exploration of strengths and limits as well as the analysis of the divides are in our view the prerequisites for a dialogue between the two approaches.

Mutual learning represents the crucial aspect to mediate between the different epistemologies and languages hold by requirement engineering and human-centered design. Mutual learning is here intended as something happening through action. Knowledge, following Piaget [28] can only be gained in acting upon the real and transforming it in order to understand it. This view of knowledge and mutual learning poses several challenges to research teams that begin to work together.

The first issue is to explore how Tropos and human-centered concepts are related. In our experience the mutual learning process has been achieved through the (i) definition of a shared dictionary of terms that includes all the most important concepts employed by the two methodologies; (ii) engagement in a collaborative negotiation of the definition of terms in the dictionary which leads to discover hidden relationships between terms; and (iii) the iteration and the refinement of definition until the agreement is acceptable.

The shared dictionary of terms is an important instrument for activating the communications among people with different background and professional skills. We selected the natural language as the most appropriate language to use since its expressiveness and flexibility is well suited for explaining and hence sharing the meaning of terms coming from both the methodologies. The shared dictionary has been a reference point for enabling an easily and autonomous translation among the internal knowledge of each team, coded according to a specific knowledge representation format.

The collaborative definition of terms in the dictionary represents the instrument for engaging in a mutual learning process. This is specifically useful for the appropriation of the terminology used in the other disciplines. We can say that the appropriation of the other discipline’s terminology happens when a member of the team defines, using his own terminology, a terms coming from another discipline, that may produce uncertainty or different interpretation. The iterative revision of the dictionary leads to refine the definitions in order to reach a better precision. The effort to refine definition by using other terms of the dictionary may be rewarded by discovering hidden relationships between terms.

6 Conclusions

Both human-centered and requirement engineering approaches, in particular Tropos, ground their processes in gathering knowledge about the people and their activities. In both cases the goal is to provide knowledge to design information system. However, they not only have different set of techniques and incompatible vocabularies but also they are based on two diverging epistemological foundations. Human-centered practitioners shun from formal methods and principally aim at providing rich information about the users and the context in which they carry out their activities.

On the other side, requirement engineering approaches are grounded on accountable and formal or semi-formal procedures that are exploited throughout the design life-cycle: from requirements analysis to systematic and complete system description. The "engineering perspective" to system design is well suited for managing requirements within the design life-cycle and in providing guidance for transitioning from informal to (semi-)formal knowledge and for analyzing conflicts between requirements. For instance, Tropos, - a representative of the Goal-Oriented requirement engineering approaches - plays a fundamental role in the development of enabling reasoning about the domain features with the aim of identifying conflicts and of checking for validity of functional and non-functional requirements.

In this paper we addressed some of the challenges posed by a joint use of the two approaches starting from the experience done within the ACube project. We discussed how a "dialogic" relationship between the disciplines may provide guidance for researchers from requirement engineering and human centered design field that cooperate within the same design process. In this perspective, the orchestration of different contributions, the establishment of communication practices and the engagement within a mutual learning process are presented as crucial steps to take full advantage of different research traditions.

References

1. Abowd, G.D., Mynatt, E.D., Rodden, T.: The human experience (of ubiquitous computing). In: Pervasive Computing. IEEE, Los Alamitos (January–March 2002)
2. Lyytinen, K.J., Yoo, Y., Varshney, U., Ackerman, M., Davis, G., Avital, M., Robey, D., Sawyer, S., Sorensen, C.: Surfing the Next Wave: Design and Implementation Challenges of Ubiquitous Computing. Communications of the Association for Information Systems 13, Article 40 (2004).
3. Kaindl, H., Constantine, L., Pastor, O., Sutcliffe, A., Zowghi, D.: How to Combine Requirements Engineering and Interaction Design? In: Proceedings of the 2008 16th IEEE International Requirements Engineering Conference (RE 2008), pp. 299–301. IEEE Computer Society, Washington, DC, USA (2008)
4. Nebe, K., Zimmermann, D.: Aspects of integrating user-centered design into software engineering processes. In: Jacko, J.A. (ed.) Proceedings of the 12th International Conference on Human-Computer Interaction: Interaction Design and Usability (HCI 2007), pp. 194–203. Springer, Heidelberg (2007)

5. Sutcliffe, A., Maiden, N., Minocha, S., Manuel, D.: Supporting scenario-based requirements engineering. IEEE Transactions on Software Engineering 24(12), 1072–1088 (1998)
6. Van Lamsweerde, A., Willemet, L.: Inferring declarative requirements specifications from operational scenarios. IEEE Transactions on Software Engineering 24(12), 1089–1114 (1998)
7. Rolland, C., Achour, C.B., Cauvet, C., Ralyté, J., Sutcliffe, A., Maiden, N., Jarke, M., Haumer, P., Pohl, K., Dubois, E., Heymans, P.: A proposal for a scenario classification framework. Requir. Eng. 3(1), 23–47 (1998)
8. Uchitel, S., Chatley, R., Kramer, J., Magee, J.: Goal and scenario validation: a fluent combination. Requir. Eng. 11, 123–137 (2006) ISSN: 0947-3602
9. Cockburn, A.: Structuring use cases with goals. Journal of Object-Oriented Programming 10(5), 56–62 (1997)
10. Liu, L., Yu, E.: Designing Information Systems in Social Context: A Goal and Scenario Modelling Approach Information Systems, vol. 29(2), pp. 187–203 (17) (2004)
11. Aoyama, M.: Persona-and-Scenario Based Requirements Engineering for Software Embedded in Digital Consumer Products. In: Proc. of the 13th IEEE International Conference on Requirements Engineering, p. 94 (2005)
12. Henderson-Sellers, B., Ralyte, J.: Situational Method Engineering: State-of-the- Art Review. Journal of Universal Computer Science 16(3), 424–478 (2010)
13. Brinkkemper, S.: Method engineering: engineering of information systems development methods and tools. Information and Software Technology 38(4), 275–280 (1996)
14. Cossentino, M., Gaglio, S., Garro, A., Seidita, V.: Method fragments for agent design methodologies: from standardisation to research. International Journal of Agent-Oriented Software Engineering 1(1), 91–121 (2007)
15. Seidita, V., Cossentino, M., Gaglio, S.: A repository of fragments for agent systems design. ICAR-CNR, Tech. Rep. (2006)
16. Susi, A., Perini, A., Mylopoulos, J., Giorgini, P.: The tropos metamodel and its use. Informatica 29(4), 401–408 (2005)
17. Rogers, Y.: New Theoretical approaches for Human-Computer Interaction. Annual Review of Information, Science and Technology 38, 87–143 (2004)
18. Eigenbrode, S., Wulfhorst, J.D., Althoff, D.M., Goldberg, C.S., Merrill, K., Morse, W., Nielsen-Pincus, M., Stephens, J., Winowiecki, L., Bosque-Pérez, N.A.: Employing Philosophical Dialogue in Collaborative Science. BioScience 57, 55–64 (2007)
19. Strober, M.H.: Interdisciplinary Conversations: Challenging Habits of Thought. Stanford University Press, Stanford (2010)
20. Bresciani, P., Giorgini, P., Giunchiglia, F., Mylopoulos, J., Perini, A.: Tropos: An Agent-Oriented Software Development Methodology. Autonomous Agents and Multi-Agent Systems 8(3), 203–236 (2004)
21. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco (1998)
22. Rosson, M.B., Carroll, J.M.: Scenario-based design. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook. L. Erlbaum Associates Inc., Hillsdale (2002)
23. Cooper, A., Reimann, R., Cronin, D.: About face 3: the essentials of interaction design. Wiley India Pvt. Ltd., Chichester (2007)
24. Grudin, J., Pruitt, J.: Personas, participatory design and product development: an infrastructure for engagement. In: Proceedings of Participatory Design Conference 2002 (June 2002)

25. Weber, R.: The rhetoric of positivism versus interpretivism: A personal view. MIS Quarterly 28(1), 3–12 (2004)
26. Haumer, P., Pohl, K., Weidenhaupt, K.: Requirements elicitation and validation with real world scenes. IEEE Transactions on Software Engineering 24(12), 1036–1054 (1998)
27. Robinson, M., Bannon, L.: Questioning representations. In: Bannon, L., Robinson, M., Schmidt, K. (eds.) Proceedings of the Second Conference on European Conference on Computer-Supported Cooperative Work, ECSCW 1991 (1991)
28. Piaget, J.: Piaget's theory. In: Mussen, P.H. (ed.) Manual of Child Psychology, 3rd edn., vol. 1. John Wiley, London (1970)

Predicting Selective Availability for Instant Messaging

Mirko Fetter[1], Julian Seifert[2], and Tom Gross[1]

[1] HCI Group, University of Bamberg, Germany
firstname.lastname@uni-bamberg.de
[2] Paluno, University of Duisburg-Essen, Germany
firstname.lastname@uni-due.de

Abstract. Instant messaging (IM) systems allow users to spontaneously communicate over distance, yet they bear the risk for disruption of the recipient. In order to reduce disruption, novel approaches for detecting and presenting mutual availability are needed. In this paper we show how fine-grained IM availability predictions can be made for nomadic users solely based on sensors installed on a laptop computer. Our approach provides comparable accuracies to previous work, while it eliminates the need for augmenting the offices or the users with further sensors. We performed a user study to collect sensor data. Alongside with labels collected by means of Experience Sampling, the data allow for creating probabilistic models for predicting selective availability. This way, we demonstrate how the required effort involved in proactively managing one's availability selectively towards a variety of recipients can be reduced by automatic adaptation, and give insights in the lessons learned.

Keywords: Instant Messaging, Context Inference, Sensors, Privacy.

1 Introduction

Instant messaging (IM) systems offer a great possibility for people to spontaneously communicate over distance, yet bear the risk for disruption of the recipient. In order to reduce disruptions novel approaches for detecting and presenting mutual availability are needed. Previous research examined how predictions based on sensors installed in the environment and desktop computer can help a caller to better estimate, if a recipient in an office scenario is currently interruptible [1, 5, 6, 9]. Others have looked at predicting availability for phone calls based on body-worn sensors and sensors installed on mobile phones [8, 21]. In this paper we show how fine-grained IM availability predictions can be made for nomadic users solely based on sensors installed on a laptop computer. Our approach provides comparable accuracies to previous work, while it eliminates the need for augmenting the offices or the users with sensors, allowing it to be instantly applied in diverse environments and situations. This way, we demonstrate how the required effort involved in proactively managing one's availability selectively towards a variety of recipients can be reduced by automatic adaptation.

In this paper we first motivate the need for fine-grained predictions from a user-centred perspective since they allow for lightweight selective availability (i.e., easy management of different levels of availability to different groups of social contacts).

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 503–520, 2011.

We elaborate on the technical and formal setup of the study and outline the underlying concepts for the collection of sensor data and user feedback. Further, we give detailed insight into the exploration, preparation, and analysis of the collected sensor data as well as into the machine learning performance. We identify lessons learned that can inform future research in this field and reflect our findings in comparison to related work. This paper concludes with an outlook to future work.

Accordingly, the contributions of this paper are: A demonstration of how machine learning can help to reduce the users' effort of managing well-differentiated selective availability in various environments, verified by the quality of the classification result; and a set of lessons learned that can inform future work in this field.

2 Automatic Adaptation of Selective Availability

Laptop computers give users the freedom to work at different places. Besides the office, people use them to work from home, at coffee shops or hotels, to take them to meeting rooms, presentations, or lectures, or to work on a train while commuting to the office or on a plane while flying to a business meeting. Such changing environments often come with different preferences to whom the user is currently available for communication. For communication via IM, presence and awareness information can help users to mutually present their availability for communication. However, with online status and status message most current IM systems only offer rudimentary support for their users to find an appropriate moment for initiating communication. Studies show that this deficit often leads to disruption of a recipient by an incoming message during a specific task or within an inappropriate situation resulting in communication breakdown [22]. Two fundamental challenges can be identified where current IM systems fall short:

- First, current IM systems only offer one single global online status. Therefore, users communicate their availability to all their contacts in the same way. A differentiation of their availability towards different categories of social contacts (e.g., family, colleagues, etc.) is only possible with workarounds in current systems. For example, Voida et al. [22] showed with a study that users maintain multiple IM accounts in order to achieve selective availability towards different categories of contacts. Other studies describe similar findings [2, 16].
- Second, the users need to adapt the online status manually. This leads to a self-interruption in their workflow and thus most times is simply forgotten. Hence, the resulting incorrect online status looses it significance for the users' contacts and accordingly is ignored because it is regarded as unreliable. This behaviour was also observed in the study of Voida et al. [22] were users textually manage their availability by enquiring. In their evaluation of a mobile 'Personal Presence' system Milewski and Smith [14] found that users changed their status on average only 1.4 times a day.

Based on these two premises, we conclude that managing selective availability by adapting multiple online statuses manually is not feasible for users. Therefore, we suggest automatic adaptation of the online status through predicting selective availability, which will help users to (a) maintain an up-to-date online status, (b) tailored towards different audiences, (c) while minimising their configuration effort.

In the following we demonstrate the feasibility of this approach by means of classification results with an average accuracy of 81.35% for the prediction of selective availability on the basis of real user data collected in a study.

3 Collecting the Data

We performed a user study in order to collect data and to research the predictability of fine-grained selective availability levels. The collected data, on the one hand, needs to describe the user's context. On the other hand, class labels that reflect the users inner states (i.e. availability preferences) are needed for training probabilistic models. Several related studies e.g., [8, 10] show that the Experience Sampling method in combination with sensor data logging is a suitable means for collecting such data.

In the following we illustrate the technical setup of this study and discuss in detail the selection of sensors and the Experience Sampling setup. Further, we report on the execution of the study and the collected data.

3.1 Technical Setup

Our study focuses on nomadic users where work and private life often intermix and therefore a need for selectively managing their availability exists. We used the primary tool of the study participants, their laptop computers, as platform for collecting data. On the one hand, recent laptops provide a variety of hardware sensors (e.g., camera, accelerometer), which can be used for detecting the user's context directly. Moreover, the computer usage already reflects a user's context considerably: the current used applications, the number of unread emails or the calendar entries all allow to make estimates about the users current context.

We implemented a system that runs as a daemon, drives a number of sensors, and saves lo-fi context data. A sensor is defined as a soft- or hardware component, or its combination that monitors and records the state of an artefact, an entity or the environment in the digital or physical realm. Sensors can be dynamically added as plug-ins, which are loaded at system runtime. The system provides a utility application for the investigator that allows the configuration of sensor parameters such as quantisation level or sampling rate (see Figure 1). The sampling rate for each individual sensor was balanced upfront in respect to expressiveness of the individual sensor and resource consumption and varied between 30 and 150 seconds.

By studying the related work, we came to a collection of sensors for capturing context data that worked in similar approaches [5, 6, 11]. In contrast to the related work, our approach aims at using only sensors that need no further instrumentation of the environment, in order to be applicable ad-hoc in different locations. Therefore, we discarded those sensors used in related work, which needed a stationary installation (e.g., door sensor, phone sensor) and tried to further exploit the capabilities of the mobile computers to gain more sensor information (e.g., build in accelerometer and photocell). All our sensors were implemented in Java and obtain their information either by making calls to native system libraries, by using command line calls or by running small scripts to access information from the hardware, the operating system, or specific applications.

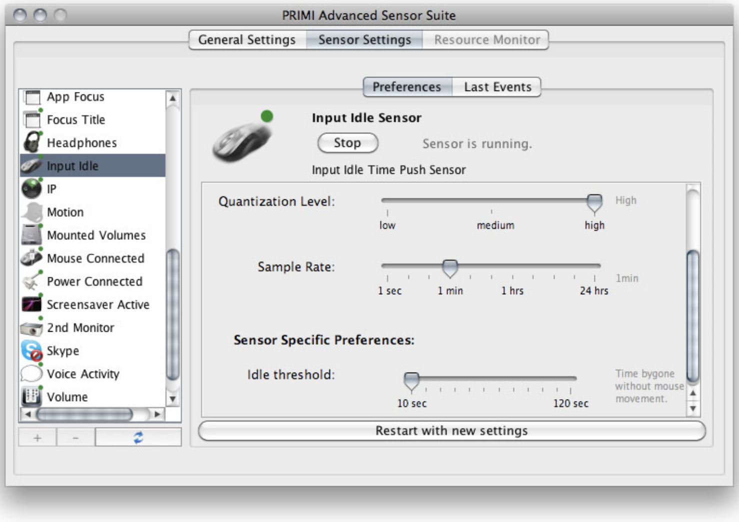


Fig. 1. Configuration dialog for adjusting the parameters of the sensors

In total we implemented 30 different sensors, each collecting one or more information chunks with each sensing, which where recorded in one of four different xml-data structures as a sensor event:

- *1×1 sensors*: return a single value per sensor event
- *1×n sensors*: return several, semantically different values of same or different types per sensor event
- *n×1 sensors*: return a list of values of the same type
- *n×m sensors*: return a matrix of values

In the following list all implemented sensors and their capabilities are described:

- Active Access Point (1×1): ID of the access point the computer is connected to
- Active Chats (1×1): number of active chat windows for all used IM applications
- Active Network Interfaces (n×1): list of active network interface IDs
- Ambient Light (1×1): intensity level of ambient light
- Applications (n×1): list of running applications
- Application Focus (1×1): name of the application in focus
- Battery (1×n): current battery capacity; charging state; being fully charged
- Bluetooth Devices (n×1): list of nearby Bluetooth devices
- Calendar (1×n): whether there are upcoming events; whether an event is ongoing
- Connected FireWire Devices (n×1): list of IDs of connected FireWire devices
- Connected USB Devices (n×1): list of IDs of connected USB devices
- CPU (1×n): user CPU load (%); system CPU load (%); idleness of CPU (%)
- Email (1×n): number of unread; number of received; and number of send email
- Ethernet Connected (1×1): is computer connected to a network by a Ethernet cable

- Face Detection (1×1): is a face detected on the image of the built-in web cam
- Focus Title (1×1): title of the front most window
- Headphones Connected (1×1): is external audio hardware connected
- Input Idle (1×1): are keyboard and mouse idle for a specific period
- IP Address (1×n): External IP address (given by the ISP); Local IP address
- Mounted Volumes (n×1): list of mounted volumes
- Motion (1×n): acceleration of X-axis; acceleration of Y-axis; acceleration of Z-axis
- Mouse Connected (1×1): whether a mouse is connected
- Power Connected (1×1): whether the power cable is plugged in
- Screensaver Active (1×1): whether the screensaver is active
- Second Monitor (1×1): number of connected monitors
- Skype (1×n): number of online contacts; mood message; status; call is ongoing
- Time (1×n): hour of the day; day of the week; weekend; part of day
- Voice Activity (1×1): is human voice detected via the built-in microphone
- Volume Settings (1×n): output volume; alert volume; audio is muted
- Wi-Fi (n×m): list of nearby Wi-Fi networks each with: SSID, BSSID, and RSSI

3.2 The Experience Sampling Dialog

In addition to this list of sensors an Experience Sampling sensor was installed. Its purpose was to prompt participants for an assessment of their current availability. These self-estimates serve as label for the succeeding training and evaluation of probabilistic models by means of machine learning. This sensor presented a dialog (see Figure 2) in which participants were prompted to give a statement about their current general availability and their selective availability as well as their current location. General availability was defined as the availability towards all of their contacts—as they would set it in their current IM system. Selective availability was measured for three *availability categories*: *Private*, *Work,* and *Public/Others*. We defined *availability category* as a group of contacts towards which the user is available in the same way.

In order to allow for comparison between the participants, we pre-specified three availability categories instead of letting the participants choose their own arbitrary number of availability categories. These are based on a pre-study in which we looked at research focussing on privacy and sharing preferences in the context of IM as well as in general awareness support (e.g., [7, 15, 17, 18]). Based on these studies we identified seven common categories for grouping contacts, towards which information is disclosed in the same way (e.g., family, team, subordinates, etc.). In order to reduce the effort for the Experience Sampling, we conducted a paper-based questionnaire survey to trim down the number of categories. We asked 30 participants to group the identified categories into optionally three or four clusters according to how they would be available for theses groups of persons. The configurations that occurred most incorporated three clusters, which make up our three availability categories and were labelled *Private*, *Work*, and *Public/Others*.

Within these categories the participants could assess their availability in terms of six *availability levels: Offline*, *Do not disturb*, *Not available*, *Away*, *Online*, and *Text Me!*. We chose these six availability levels, imitating the online statuses provided by

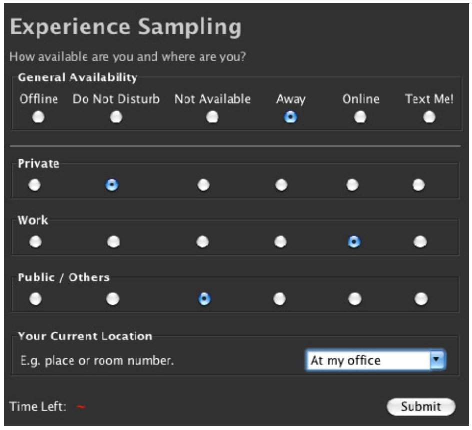


Fig. 2. Experience Sampling interface for collecting user selective availability preferences

Skype, as all our participants are frequent Skype users and thus used to these levels. We preferred this over using a five-point Likert-scale or a simple binary assessment of “Available” versus “Unavailable”.

During the study, this Experience Sampling dialog was presented every 25 to 35 minutes as the front-most window. When the participants did not start to interact with the dialog for more than 30 seconds, the dialog disappeared and reappeared 5 minutes later. In the moment participants started interacting with the dialog, this countdown was stopped, until the participants completely filled out and submitted the dialog. Afterwards the dialog disappeared and started over to appear in the normal interval. This way, we assured that we don’t miss a number of samplings when the participants are often away from screen for a short time.

3.3 Study Setup and Execution

While ESM allows collecting rich and in-depth data through repetition the disruption and the resulting effort for the individual subjects during a four-week study is very high. In order to achieve comparable results we used the same number of subjects as in related work [5, 11] in order to collect a similar amount of data. We therefore recruited four volunteers as participants (P1-P4) for a four-week long study. The participants were all males and aged between 25 and 33 (M=28.0). Three participants have a background in computer science one participant in social science: two participants were research assistants and the other two participants were student assistants. All come with several years of experience with instant messaging (M=9.5; SD=4.5) and use in average 4.5 (SD=2.0) different instant messaging accounts; all participants use Skype. Except one participant, who indicated to use instant messaging only during workdays, the others use instant messaging on a daily basis. All participants indicated that they use their laptop at different locations during a normal workday. The study was framed by a short questionnaire upfront and a semi-structured interview afterwards.

Data on the participants were collected over the period of four weeks by means of the sensor daemon installed on their laptop computers. The participants were briefed to run the sensor daemon continually and to turn it off only in exceptional situations (e.g., when holding a presentation or the computer's battery runs low). Therefore, each participant collected a different amount of sensor samples and Experience Sampling self-reports. Figure 3 shows the difference between cast and actually answered self-reports.

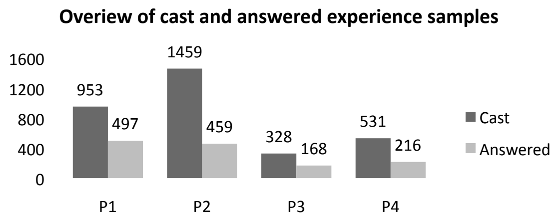


Fig. 3. Amounts of cast and actually answered self-reports for each participant

On average, the Experience Sampling dialog was presented 817.75 times (SD=500.52) and 270,726 (SD=133,612) sensor events were collected by 30 sensors. These numbers (c.f. Figure 3) show that the majority of the dialogs remained unanswered. This is an intentionally introduced bias of the study design. As described in 3.2, as long as a dialog remains unanswered the sampling interval is decreased from 30 to 5 minutes in order to miss less data when the computer is left unattended only for a short period. On the other side this leads to a high number of unanswered dialogs for longer breaks (e.g. a one-hour lunch break would result in 12 instead of 2 unanswered dialogs), which can be neglected. An approach to treat unanswered dialogs as an indicator for unavailability was discarded, as it is unclear whether the subject is unreceptive or really not available [8].

Overall participants gave their estimates concerning their availability in average 355 (SD=167.00) times. The differences between the subjects illustrated in Figure 3 are likely due to the varying periods the sensor daemon was running as well as different computer usage patterns. For instance, P4 had the software installed on a computer usually being used only during common working hours. P3, who answered the smallest amount of samples, had the sensor daemon installed on a personal laptop computer but turned it off frequently because off the performance loss it caused. P1 and P2 ran the daemon in average 336 hours during the four-week study period. The high number of unanswered self-reports of P2 could be caused by the individual usage pattern letting the computer run while being physically away from it. The samples were collected at 4.8 (SD=2.2) locations including the respective users' office, meeting rooms, lecture rooms, other people's offices, the university' cafeteria or at home. A more qualitative interpretation of the Experience Sampling data in respect to the participants' needs for selective availability is discussed in [3]. In the following we will solely focus on the collected data from the perspective of predictability of selective availability through machine learning.

4 Predicting Selective Availability

In this section we illustrate the pre-processing of the data and to what extend it is possible to predict selective availability levels based on collected sensor data. We describe the process of transforming the sensor data into features, the generation and selection of features, and the selection of the classification algorithms. This proceeds with the discussion of the evaluation of the machine learning performance.

4.1 Feature Generation, Selection, and Classification

The data provided by sensors do not allow reasoning about the users' availability right away. It is necessary to transform the collected sensor data logs from each user into meaningful separated features. For instance, the *Applications*-sensor ($1 \times n$) provides a list of all running applications with each sensor reading. Each application name contained in the sensor value refers to one feature, which needs to be extracted. Hence, when creating a probabilistic model all the lists of application names need to be processed to collect the complete list of available features.

Besides this simple transformation of sensor data into features, we generated additional features by integrating the feature values over time. This allows to take advantage of a sensor's value history which reveals trends and tendencies. In order to keep the number of generated features low, we decided to generate time integrations only for five, eight, and 14 minutes. In particular, for numeric features we calculated the mean and standard deviation for the respective history time. For Boolean values we calculated to what extend the particular feature was true or false during the respective period. This approach of feature generation increases the dimension of the feature space enormously. The next step therefore was to discard features with a minimal information gain upfront. The resulting number of features after this reduction still was 1437.3 in average.

As most machine-learning algorithms perform better in a low dimensional feature space, it was necessary to further reduce the number of features for the classification. We compared different techniques for selecting a small subset of features. In particular, we focused on Correlation-based Feature Selection (CFS), Information Gain, Gain Ratio, and Chi-square Feature Subset Selection. Since different classifiers perform significantly different when trained with differing feature subsets, it is necessary to explore the selection of the classifier in combination with the feature subset-selection. Therefore, we considered four different classifiers: Naïve Bayes, C4.5 Decision Tree, Random Trees, and Support Vector Machine (SVM). We used the machine-learning workbench WEKA [23] and its experimenter tool for determining which subset of features in combination with which classifier leads to the best classification results. The results show that WEKA's SMO Classifier—a SVM implementation utilising the Sequential Minimal Optimization (SMO) [19] algorithm for training—in combination with a feature subset selected by means of Correlation-based Feature Selection (CFS) outperforms all other combinations. Subsequently, the best parameters for the SMO algorithm were determined using WEKA's GridSearch for further optimising the classification performance.

4.2 Evaluation of Machine Learning Performance

The evaluation of the predictability of selective availability is based on the data collected in the user study. The evaluation consists of building four probabilistic models for each user using the SMO algorithm based on the (in average 23.4 (SD=10.8)) features selected by the CFS in combination with the (in average 355 (SD=167.00)) samples based on the labels given by the individual participants. For each participant, four models were built—one for *General availability* and three for the availability categories. Each model is trained to predict the six availability levels, which makes the classification task a multi-class problem. The 10-fold cross-validated classification results for the 4x4 models are summarised in the following Table 1.

Table 1. Results of 10-fold cross-validations for each of the four participants' models, and the average over all models of all participants. Each table presents the accuracy in percent for the ZeroR classifier (to show the base probability), the SMO classifier, and the improvement of the SMO classifier against ZeroR classifier (Diff.) for the different models.

Model	ZeroR	SMO	Diff.
P1 General	58.15	75.65	17.51
P1 Private	55.53	75.86	20.32
P1 Public	59.76	78.27	18.51
P1 Work	51.31	75.86	24.55
P1 Average	**56.19**	**76.41**	**20.22**

Model	ZeroR	SMO	Diff.
P3 General	74.40	91.67	17.26
P3 Private	86.90	92.26	5.36
P3 Public	69.64	86.90	17.26
P3 Work	78.57	92.26	13.69
P3 Average	**77.38**	**90.77**	**13.39**

Model	ZeroR	SMO	Diff.
P2 General	68.62	84.10	15.48
P2 Private	70.29	83.68	13.39
P2 Public	66.53	82.01	15.48
P2 Work	71.13	84.94	13.81
P2 Average	**69.14**	**83.68**	**14.54**

Model	ZeroR	SMO	Diff.
P4 General	62.50	74.54	12.04
P4 Private	70.37	76.85	6.48
P4 Public	74.54	76.85	2.31
P4 Work	64.35	69.91	5.56
P4 Average	**67.94**	**74.54**	**6.60**

Model	ZeroR	SMO	Diff.
All Average	**67.66**	**81.35**	**13.69**

In the most left column of each table the name of the respective data set is listed. In the second column, the base probabilities generated by the ZeroR algorithm are listed. In the third column the accuracy based on a stratified 10-fold cross-validation of the models that were built by the support vector machine SMO are listed. The last column contains the difference between base probability, as estimated by the ZeroR classifier, and the percentage of correctly classified instances by the SMO. The average over the four models of each participant reflects the overall experience the users will have, as all four models will work in unison to adapt the users selective availability.

The results show that over all participants the base line is quite high with 67.66%. This means, in average, one online status is selected more often then all other online status together by a user for a specific availability category. Accordingly, a classifier that always guesses this dominant class would make a correct prediction in 67.66% of the cases. Therefore, the results from the evaluation of the probabilistic model need to be regarded relative to their respective base line. The results shows that in average the probabilistic models perform considerably better (13.69%) than the base line. Overall, with an average accuracy of 81.35% our results can compete with the results of related work; as discussed in detail later.

5 Lessons Learned

In the following section we elaborate on lessons learned during the study and the explorative machine learning procedure, which can inform future research on similar approaches. We elaborate on the performance of the individual sensors for machine learning. We give guidelines on the engineering of robust sensors for this kind of studies. We motivate our argumentation for building personalised models for each individual user. We describe our findings on the usage of the Experience Sampling method. And we show patterns we found in the configuration of selective availability.

5.1 Promising Sensors and Features

The results of the feature selection mechanisms allow drawing conclusions on the relevance of single sensors for the prediction result. Such knowledge can inform future research at selecting relevant sensors. Therefore, we analysed the features of all 16 models (four participants, each with a model for General, Private, Work, and Public/Others) that were selected by the correlation-based feature selection algorithm as most relevant. CFS uses a heuristic evaluation method and has proven to be equally good at finding relevant features in machine learning tasks in comparison to wrapper-based approaches. Hence, the selected subset can give a valid estimate on the contribution of individual sensors.

The CFS algorithm reduced the number of features from 1437.3 (SD=303.6) per model to the 23.4 (SD=10.8) most relevant. First, we consolidated the reduced feature sets of all 16 models and analysed what feature generation mechanism delivered the largest number of features to this set. 71.93% of the features were generated by the time-based feature generation. The features that take into account the last 14 minutes made up the majority with 32.89%, followed by 5 minutes (21.39%) and 8 minutes (17.65%). As expected, taking into account the sensor values for a period of time instead of only a point in time can deliver richer features for the classification [4].

Further, we analysed the sensors providing one or more relevant features to one or more of the four models for each participant according to the result of the CFS. Table 2 shows that while eleven sensors generated features relevant for at least one model of each of the participants, ten sensors did not. For the category of unused sensors, we found that the users did not influence some measured parameters during the four-week study period. For example, none of the users connected FireWire devices or mounted a volume. Accordingly, these sensors can be valuable for users that use such devices or services more often. Other unused sensors, like the Mouse Connected Sensor might be substituted by the information of a second sensor, like the Connected USB Devices sensor. Further, our expectation that different combinations of sensors

Table 2. Sensors that delivered features to the final feature subset: "+" sensor delivered one or more features , "-" sensor delivered no feature for this participant's models.

Sensor	P1	P2	P3	P4	No.
Active Access Point	+	+	+	+	4
Active Chats	-	-	-	-	0
Active Network Interfaces	-	-	-	+	1
Ambient Light	-	+	+	+	3
Application Focus	+	+	+	+	4
Applications	+	+	+	+	4
Battery	-	-	-	-	0
Bluetooth Devices	+	+	+	+	4
Calendar	+	+	+	-	3
Connected FireWire Dev.	-	-	-	-	0
Connected USB Dev.	+	+	+	+	4
CPU	-	-	+	+	2
Email	-	-	-	-	0
Ethernet Connected	-	-	-	-	0
Face Detection	-	+	+	+	3

Sensor	P1	P2	P3	P4	No.
Focus Title	+	+	+	+	4
Headphones Connected	-	-	-	-	0
Input Idle	-	+	+	+	3
IP Address	+	+	+	+	4
Motion	+	+	-	+	3
Mounted Volumes	-	-	-	-	0
Mouse Connected	-	-	-	-	0
Power Connected	+	-	-	-	1
Screensaver Active	-	-	-	-	0
Second Monitor	-	-	-	-	0
Skype	-	+	+	+	3
Time	+	+	+	+	4
Voice Activity	+	+	+	+	4
Volume Settings	+	+	+	+	4
Wi-Fi	+	+	+	+	4

monitoring cables attached to the computer (Ethernet, Headphones, Power, etc.) would give a good estimate about the current location, were not met, as theses sensors practically played no role for the models.

However, most of these unused sensors are not computationally or energetically expensive, as they basically listen to system events, in comparison to more expensive operations like constantly polling for available Wi-Fi networks or performing face-detection on the camera image. As there is a chance that these sensors might be useful for other users, it might be reasonable to not discard them completely. Further, some of these sensors might be valuable for indicating transitions between different contexts. As our Experience Sampling study used a random interval of circa 30 minutes to assess the users' availability, the collected samples cannot reflect the exact moment when users' inner states change e.g., from Available to Do Not Disturb. Some of the sensors that were not used for building the predictive models could be valuable for predicting a good moment for adapting the online status. For example, disconnecting the mouse and Ethernet from a mobile computer could indicate that the user is leaving the current place, and suggest a change in availability.

5.2 Carefully Engineered Sensors

The design and implementation of the sensors and the surrounding architecture are crucial for the success of the study. In the following we list several technical requirements, which we found in small informal pre-tests and during the study.

The sensors for data collection need to be robust. During the study the aim was to continuously collect data while the system was in normal use. Accordingly, the implementation should be robust to system restarts, idle-times, crashes or reconfigurations while not limiting the users in their tasks. Crashing sensors should automatically recover without the need for user intervention. We implemented mechanisms that monitor and automatically restart sensors that stopped delivering values. At the same time, sensors should not interfere with user interactions. Sensors that exclusively allocate system resources (in our case camera and microphone) need to be designed in away that these resources are still available for the users' primary tasks (in our case for a video chat). In our setup the sensors obtained data and were released for each sampling, with small pauses between the samplings. If a resource was occupied, the sensors tried again until the resource was obtainable (e.g., after the video chat was finished). In the same way, sensors should not open applications like calendar or email to retrieve sensor values, unless the users open them.

In addition to this, collecting sensor data continuously from a mobile computer leads to a reduced battery lifetime and system performance. Continuously accessing hardware like the camera or the embedded acceleration sensors in combination with repeated polling for nearby Wi-Fi or Bluetooth networks has an unavoidable but substantial effect on the battery and CPU. In order to reduce this drawback, we balanced the sampling intervals for the different sensors between maximising the amount of valuable data that is collected and minimising usage constraints.

Further, privacy-protecting mechanisms have a major effect on the users' willingness to collect data. We implemented a hash-based privacy based on the Subtle system [4]. Each nominal value that was collected (e.g., the title of the current focussed application, the SSID of nearby Wi-Fi access points) was translated into a unique hashed representation. In this way the data is still meaningful for machine learning algorithms, but reduces its descriptiveness for humans, who prepare the data for machine learning. However, this comes with a decreased interpretability for reflecting on more qualitative aspects of the data in an analysis of the study results. For example, we are unable to make statements about availability preferences while using specific applications like Word or Firefox—which could provide interesting insights for CSCW research—due to the encryption.

Finally, the data collection process needs to be as unobtrusive to the users as possible, yet has to allow the users to stay in control. Whether for reasons of system performance or for privacy reasons, the user has to be able to pause the data collection at any time. In our system, users were able to stop and start the collection with the click on a button. However, in order to reduce the users' need to do so, and to obtain more data from the study, the above guidelines can help.

5.3 General Models vs. Individual User Models

One drawback of building an agent-based system is that the system needs to learn from scratch and so is of limited utility to the user at the beginning or in new situations. In order to overcome this issue Collaborative Interface Agents [13] were introduced that share their knowledge among user models. In the field of predicting human interruptibility Fogarty et al. [6] for example built one general model for all subjects and general models for each group of subjects (manager, researcher, intern) in order to overcome the cold-start problem. However, their individual group models performed better than the global general model. From our results we can confirm this findings and even advocate for completely personalised individual models.

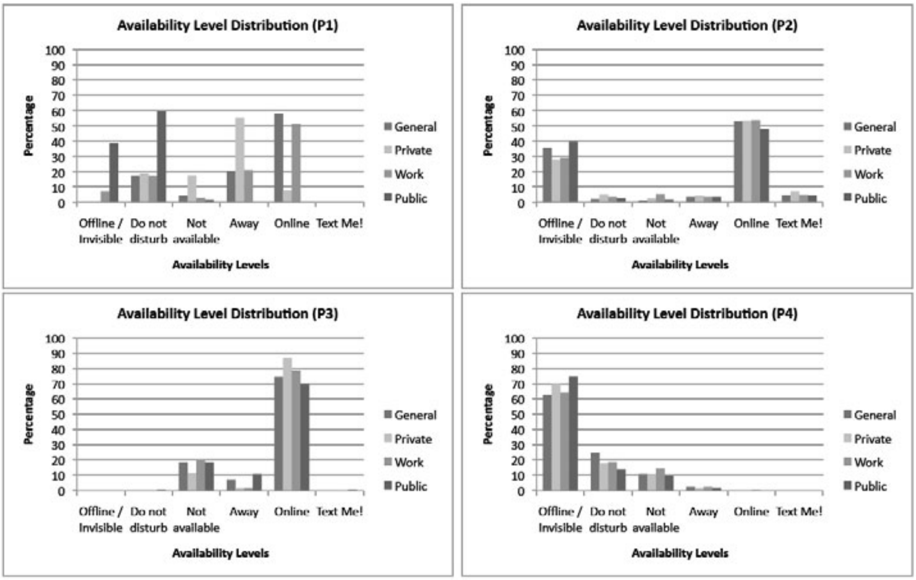


Fig. 4. Selective Availability distributions for the four subjects in percent

The first reason is that we found that even among the small number of our subjects, the individual availability expressed through the chosen availability levels strongly varied as can be seen in Figure 4. The Figure shows that while P3 had a clear tendency to being highly available, P4 has a clear tendency to be unavailable, while P1 and P2 have a broader spectrum of availability. This argues for the users' individual understanding of and a personal perspective on availability that a general model can hardly reflect.

Further, such general models need to be built solely on features that all users have in common. This bears three challenges: First, a feature with a high informative value for an individual model, like a wireless network at home, would most likely be discarded from the feature vector of a general model, as it would contain missing values for the other subjects. Second, some of the features may represent contradictory information for different users. Referring to wireless networks as an example, the wireless networks that are accessible from a certain place, may indicate the office for one user and a meeting room for another user, and accordingly different availability preferences. Third, some features' values might correlate very well for different users, but cannot be automatically merged into one feature. For example, the use of the browser or a word processor might lead to similar availability preferences for different users. However, when different users use different applications (e.g., one user uses Firefox, the other user Opera) this sensed information cannot be combined easily. Of course, all three challenges could be overcome by introducing a semantic layer between the sensed data and the generated features. But in many cases (e.g., mapping the Wi-Fi Access Points at home to the semantic representation home) requires further effort by the users for generating individual semantic abstractions.

Accordingly we found for our data: The performance of general models that were built with different classifiers on the intersecting subset of features from all the collected data performed insignificantly better than the base probability of the ZeroR classifier, but most times worse. Satisfactory classification results could not be achieved by using a general model.

5.4 Designing the Experience Sampling

The Experience Sampling dialog is the predominant interface to participants and needs special care. During the study the users are interrupted by the dialog several times a day during their daily tasks. In order to keep the study participants motivated to deliver valid self-reports, it is necessary to minimise the effort for the users on the one side and to prevent habitual, repeated responding [20] on the other side.

The design of our interface was optimised in iterative steps. It minimises the number of clicks for users, and prevents a bias towards repeated responses. We abandoned a design, where the choices from the previous sampling were pre-selected, as this resulted in a bias towards just confirming the previous choices. A design where we randomised the ordering of the items each time dramatically increased the users' cognitive effort and accordingly increased their frustration with the study—and so was also discarded.

In order to minimise the variance of free-text answers for the "Current location"-field, we implemented an auto-complete mechanism. From the subsequent interviews we learned, that this sometimes led to an automatism to accidentally enter and select the location that the users stayed at the most (e.g., their office). This came with a desire of the users for a mechanism to recall and alter or discard the last assessment.

The dialog always appeared as the front-most window in the centre of the screen and could not be discarded by the user. This rather invasive design helped to maximise the number of samples, since the fastest way to get rid of the dialog was to quickly respond to the questions. In the interviews we learned that this was perceived as acceptable in situations where the participants directly interacted with the computer, but was regarded as problematic where the computer was used more indirectly (i.e. to watch a video or to give a presentation). One participant suggested a configuration, which would allow shifting the Experience Sampling to a mobile phone for a certain period of time.

Further, the timing of the Experience Sampling plays a crucial role. We decided to show the dialog at a random interval, with one exception: Every time the users logged in, a sampling was triggered. During the study and afterwards, from the data, we learned that there are a lot of simple cues (that our sensors are able to measure) which can be easily formulated as plain rules to allow sampling at a moment, that typically marks the transition between different availability levels: when the computer is moved (Motion-Sensor), cables are attached or detached (e.g., Power Connected-Sensor, Ethernet Connected-Sensor) or someone starts to speak (Voice Activity-Sensor). These are a few examples for such moments that show how a context-aware Experience Sampling approach [12] could help to improve the quality of the data.

In combination with the sensor logging, the time between the presentation of the dialog and the response have to be taken into account. As the presentation of the dialog (Application Focus-Sensor) and the users interaction with the dialog (Input Idle-Sensor) have an influence on what the sensors measure, they introduce a systematic error. Therefore it is necessary to discard the sensor data from the moment before the dialog is shown until the response was entered. If the time between presentation and response is to long, the sensor data and the Experience Sampling data will loose its correlation. In our design the user had 30 seconds to answer to the Experience Sampling before it disappeared, and was presented again five minutes later. In these 30 seconds no sensor data was recorded.

Finally, users showed great interest in the data collected when we discussed it in the subsequent interviews. To make their data available after the study to each of them individually in a meaningful way (e.g., as charts or diagrams) could make a great incentive to keep them motivated throughout the study.

5.5 Patterns for Selective Availability Configurations

The analysis of the collected data revealed that users tended to give similar answers in reoccurring contexts. For example during work-time a user might be available for work related contacts, less for private, and offline for others. This raises the question, how many different patterns or configurations users utilise to state their selective availability in the majority of the times.

We used the expectation-maximisation-algorithm for finding the patterns that fit best to the users' answers. The following table presents the results.

Table 3. Patterns (a-d) of availability, by means of clustering the participants ESM data (1=Offline, 2=Do not disturb, 3=Not available, 4=Away, 5=Online, and 6=Text Me!)

	P1				P2				P3			P4			
	a	b	c	d	a	b	c	d	a	b	c	a	b	c	d
Private	4	5	3	2	5	6	4	1	5	4	3	1	2	3	4
Public	2	2	1	1	5	6	1	1	5	4	3	1	2	3	4
Work	5	1	4	2	5	6	4	3	5	5	3	1	3	3	4

For three participants, four patterns were identified (a, b, c, d). For P3 only three patterns (a, b, c) were identified. These patterns respectively make up 84.32% (SD=3.40%) of the used configurations. On average the participants only used 23.25 (SD=14.97) combinations of the 216 possible permutations of availability levels and availability categories.

Since the numbers of the patterns for each participant are smaller than the number of all possible permutations, the complexity of the resulting classification problem is reduced. After transforming the data, again, we created probabilistic models, which were evaluated in terms of their ability to predict the correct pattern when given an example. The results showed, that the individually trained models for each availability category significantly outperform the models for predicting patterns.

Even though the results in this case are discouraging, the insight in such patterns can inform the design of future studies and of instant messaging systems.

6 Related Work

We briefly discuss related work and outline similarities and differences to our work. To our knowledge, so far no research has focussed on the predictability of selective availability based on sensors on laptop computers. We concentrate on work that focuses on predicting the availability for phone calls based on mobile sensors and two prominent examples that deal with the predictability of general interruptibility based on sensor data in office environments, which inspired our approach.

Ter Hofte [21] used the software SocioXensor for PDAs and mobile phones to collect data from 10 participants over 7 days via ESM. The resulting predictions are not explicitly focused on selective availability. These predictions—based on answers to the ESM (in conversation, location, etc.) and not on real sensor information—only achieve an accuracy of 63,9% for a two-class problem (green light/no green light). Ho and Intille [8] use activity recognition based on wearable accelerometers and a decision tree classifier on a PDA to detect postural and ambulatory transitions. They found that transitions are a good indicator for receptiveness towards interruptions. Their approach burdens the user to wear accelerometer sensors at ankle and thigh.

Fogarty et al. [6] use a combination of computer (mouse and keyboard activity, used applications, etc.) and room sensors (door, telephone, motion, etc.) to predict the interruptibility of 10 participants. They collected 975 oral interruptibility self-reports with interruptibility on a five-point Likert-scale. They used a naïve Bayes classifier to build three models for the three different types of participants and one general model for all participants. The achieved accuracies lie between 80.1% to 87.7% for the three models and 79.5% percent for the general model. However, they reduced the complexity of the prediction to a two–class problem (1-4 and 5) to obtain these results. In comparison, our average accuracy for predicting a six-class problem based on individual models lies at 81.35%. Their room sensors limit its applicability. Our participants collected data in several different locations in work and private contexts, which also introduces a greater complexity to the prediction task.

Also the BusyBody system [11] relied on sensor data collected on computers in an office environment. Four participants collected between 789 and 2365 assessments of their interruptibility ("Busy" and "Not Busy"). The average achieved accuracy with a Bayesian Network is 78.25%. Their prediction is also only for a binary classification problem. There is no information given, in which contexts the data was collected.

Concluding, none of the related work focussed on predicting selective availability for sensors on laptop computer. However, all the works greatly informed our research. Our aim was to combine the advantages of stationary and mobile approaches and to deepen it, in order to show that even more complex problems can be addressed with this approach while minimising the setup effort for deploying additional sensors.

7 Conclusions and Future Work

We have shown that the prediction of selective availability with machine learning is a feasible approach. Based on our results we conclude that systems that rely on sensor data collected on a laptop computer can be trained by users in order to allow for the automatic adaptation of selective availability in Instant Messaging. Such systems support users in better managing their availability towards different social contacts while at the same time minimising their effort in comparison to manual adaptation.

We showed how this is possible by information that is directly accessible on the users' laptop computer. We showed that we can predict, with an accuracy of 81.35%, in which way users are available for different social contacts in different situations. This was shown for work and private contexts. And it allows users to express availability with fine-grained online statuses.

In order to evaluate how this approach gives a greater flexibility to users while minimising their effort, user tests based on a system that provides such predictions in real-time are needed. To support individual availability categories for each user such systems have to support life-long learning to train an arbitrary number of individual models for each user (i.e. continuously and actively learn from the users' inputs as well as predict to adapt the system in the background). To achieve this, such systems should automatically extract, generate, and select useful features from sensor-data streams. At the same time they should request input from the users in a much more implicit way.

Acknowledgments. We thank the members of the Cooperative Media Lab, the participants in this study, and the anonymous reviewers for their insightful comments. Part of the work has been funded by the Federal Ministry of Transport, Building, and Urban Affairs (TransKoop FKZ 03WWTH018).

References

1. Begole, J.B., Matsakis, N.E., Tang, J.C.: Lilsys: Sensing Unavailability. In: Proceedings of ACM CSCW 2004, pp. 511–514 (2004)
2. Davis, S., Gutwin, C.: Using Relationship to Control Disclosure in Awareness Servers. In: Proceedings of GI 2005, pp. 145–152 (2005)
3. Fetter, M., Seifert, J., Gross, T.: Lightweight Selective Availability in Instant Messaging. In: Extended Abstracts of ACM CHI 2010, pp. 3817–3822 (2010)
4. Fogarty, J., Hudson, S.E.: Toolkit Support for Developing and Deploying Sensor-Based Statistical Models of Human Situations. In: Proceedings of ACM CHI 2007, pp. 135–144 (2007)
5. Fogarty, J., Hudson, S.E., Akteson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., Yang, J.: Predicting Human Interruptibility with Sensors. ACM Transactions on Computer-Human Interaction (TOCHI) 12(1), 119–146 (2005)
6. Fogarty, J., Hudson, S.E., Lai, J.: Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. In: Proceedings of ACM CHI 2004, pp. 207–214 (2004)
7. Gross, T., Fetter, M.: Disclosure Templates for Selective Information Disclosure. In: Proceedings of IADIS CT 2010, pp. 101–108 (2010)
8. Ho, J., Intille, S.S.: Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In: Proceedings of ACM CHI, pp. 909–918 (2005)
9. Horvitz, E., Apacible, J.: Learning and Reasoning about Interruption. In: Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI 2003, pp. 20–27 (2003)
10. Horvitz, E., Kapoor, A.: Experience Sampling for Building Predictive User Models: A Comparative Study. In: Proceedings of ACM CHI 2008, pp. 657–666 (2008)

11. Horvitz, E., Koch, P., Apacible, J.: BusyBody: Creating and Fielding Personalized Models of the Cost of Interruption. In: Proceedings of ACM CSCW 2004, pp. 507–510 (2004)
12. Intille, S.S., Rondoni, J., Kukla, C., Ancona, I., Bao, L.: A Context-Aware Experience Sampling Tool. In: Extended Abstracts of ACM CHI 2003, pp. 972–973 (2003)
13. Lashkari, Y., Metral, M., Maes, P.: Collaborative Interface Agents. In: Proceedings of the Twelfth National Conference on Artificial Intelligence - AAAI 1994, pp. 444–449 (1994)
14. Mileswki, A.E., Smith, T.M.: Providing Presence Cues to Telephone Users. In: Proceedings of ACM CSCW 2000, pp. 89–96 (2000)
15. Olson, J., Grudin, J., Horvitz, E.: A Study of Preferences for Sharing and Privacy. In: Proceedings of ACM CHI 2005, pp. 1985–1988 (2005)
16. Patil, S., Kobsa, A.: Instant Messaging and Privacy. In: HCI 2004: 18th British HCI Group Annual Conference. pp. 85–88 (2004)
17. Patil, S., Kobsa, A.: Uncovering Privacy Attitudes and Practices in Instant Messaging. In: Proceedings ACM SIGGROUP 2005, pp. 109–112 (2005)
18. Patil, S., Lai, J.: Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application. In: Proceedings of ACM CHI 2005, pp. 101–110 (2005)
19. Platt, J.C.: Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) Advances in Kernel Methods - Support Vector Learning, pp. 185–208. MIT Press, Cambridge (1998)
20. Scollon, C.N., Kim-Prieto, C., Diener, E.: Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. Journal of Happiness Studies 4(1), 5–34 (2003)
21. Ter Hofte, G.H.: Xensible Interruptions from Your Mobile Phone. In: Proceedings of MobileHCI 2007, pp. 178–181 (2007)
22. Voida, A., Newstetter, W.C., Mynatt, E.D.: When Conventions Collide: The Tensions of Instant Messaging Attributed. In: Proceedings of ACM CHI 2002, pp. 187 – 194 (2002)
23. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2006)

Testing the Usability of a Platform for Rapid Development of Mobile Context-Aware Applications

Valentim Realinho[1,2], A. Eduardo Dias[2], and Teresa Romão[2]

[1] Escola Superior de Tecnologia e Gestão/Instituto Politécnico de Portalegre
Lugar da Abadessa, Apartado 148,
7300-901 Portalegre, Portugal
valentim.realinho@estgp.pt
[2] CITI, DI-Faculdade de Ciências e Tecnologia/Universidade Nova de Lisboa
Quinta da Torre
2829-516 Caparica, Portugal
{tir,aed}@di.fct.unl.pt

Abstract. In this paper, we present the usability evaluation of IVO (Integrated Virtual Operator), a platform that supports the rapid development of context-aware applications by users with no programming skills. Using only the tools provided by the platform (IVO Builder and IVO Outlook), users can define temporal and spatial conditions and associate them with workflows of activities available within the platform. This way, whenever the defined conditions occur, the user's smartphone will immediately produce the intended behavior, with no need for user intervention. The applications developed using IVO can easily be made available to other users through a distributed web platform. Moreover, an android client was developed, to run the IVO-developed applications, allowing the smartphone to act as the ubiquitous interaction device. The evaluation of the platform was performed through usability tests at both the end-user level (android client) and the developer-user level (builder tools).

Keywords: Ubiquitous Computing, Context-Awareness, Rapid Application Development, Usability Evaluation, Interaction Design, Mobility.

1 Introduction

In Mark Weiser's vision, we should be able to live with computers and not only to interact with them [1]. The basic idea behind the ubiquitous computing paradigm is that computing moves off the desktop into the environment, operating in the periphery of our attention. Greenfield describes this interaction paradigm as "information processing dissolving in behavior" [2].

In essence, this paradigm uses small devices, with robust computing power and network capabilities, seamlessly integrated into the environment, involving this heterogeneity of devices with a high degree of interoperability. Strides in nanotechnology are enabling more and more features that are added to small devices allowing a greater awareness of the dynamic world they are part of.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 521–536, 2011.

Pister introduced the term "*smart dust*", referring to tiny sensors, no larger than grains of rice that would monitor everything, acting like electronic nerve endings for the planet [3]. In this world the user is no longer part of the system (since the interaction between the system and the user is reduced) and becomes the supervisor of the system, acting only when strictly necessary [4]. However, these devices are not part of the actual ubiquitous computing paradigm yet, while the mobile phone is the first truly pervasive computer [5].

The mobile phone is always with people, helping them to keep in touch with each other and manage everyday tasks. In this project, the approach was to use smartphones as ubiquitous interaction devices, thus enabling a widespread use of the system. Smartphones provide a large set of possibilities for sensing, communication and user interface, and there is a great market demand for these types of devices [6], making them ideal for the development of ubiquitous systems.

Although using smartphones limits context-awareness possibilities to the use of the sensors available in this kind of devices, they also facilitate the widespread use of these type of applications with no or easy setup. However, the former drawback tends to disappear since smartphones are increasingly including a larger number of sensors. The use of Near Field Communications (NFC) [7] in mobile devices opens a new spectrum for context-aware applications, which was possible only through hardware and additional infrastructure. The latest version of Android already supports NFC and iPhone 5 is expected to feature NFC support.

An important research topic addressed in ubiquitous computing is context-awareness, where applications can dynamically adapt to changes in the user activities and environments [8]. The context is a particular situation for each user, and it can include, among others, location, activity being performed, time and nearby people or equipments. Context-aware computing involves sensing those situations to provide adequate information and services to the user.

Pioneering work on ubiquitous computing research, like [1, 9-11], addresses the topic of context-awareness and has demonstrated the potential of context-aware applications. Some research has focused on developing frameworks, toolkits and infrastructures to support programmers in building context-aware applications. Other projects use end-user programming techniques, which empower end-users to prototyping context-aware applications [8, 12, 13]. However, they still lack the end-user support for the creation of complete applications using the currently available smartphones as ubiquitous interaction devices.

To address this problem, we have developed IVO, a platform that enables end-users to build and deploy context-aware applications without the need to write any programming code. IVO provides a completely visual programming environment that allows end-users to define a number of context conditions and workflows of activities, which are later triggered when the user is in the presence of those contexts. The use of a control-flow model facilitates the visualization of how activities are interconnected, taking advantage of the humans' innate spatial reasoning skills.

Three main building tools are currently available, allowing end-users with no programming skills to create mobile context-aware applications: IVO Builder, a web application, used for the definition of location contexts and respective workflows of activities; and IVO Outlook, for temporal and proximity contexts, composed of two *add-ons*, for Outlook Calendar and Outlook Contacts. The set of available workflow activities provides a wide range of capabilities, while keeping the system accessible to both novice and expert end-users. An Android smartphone client was also developed to provide the workflow engine that uses the stored model to activate the required workflow when, and where, the activation conditions are met by the user.

2 Related Work

This section summarizes how IVO relates to context-aware computing. We divide this analysis in three categories: Middleware and Frameworks, End-user Prototyping and Authoring Tools.

Middleware and Frameworks. Projects like Context Toolkit [11], SOLAR [14], Context Fabric [15], JCAF [16] and Placelab [17] are targeted at programmers. They provide the infrastructure to enable them to create context-aware applications, but they do not provide any support for end-users. Of these, the Context Toolkit is one of the most widely referenced projects. It provides a software interface to physical sensors, separating the acquisition and representation of context from the delivery and reaction to context. The authors proposed context widgets as components to encapsulate the sensing details, providing context abstraction.

End-User Prototyping. *a CAPpella* [8], Topiary [12], Papier-Mâché [18] and iCAP [13] use end-user programming techniques in order to empower end-users to prototype context-aware applications. *a CAPpella* is a programming by demonstration context-aware prototyping environment intended for end-users, using four main components: a recording system, an event detection, a user interface and a machine learning system. Topiary allows the "demonstration" of scenarios representing local contexts, and then it uses these scenarios to create storyboards that describe interaction sequences, which can run on mobile devices. Papier-Mâché is a toolkit for building tangible interfaces using computer vision, electronic tags, and barcodes. iCap, is a system that allows end-users to describe a context and associate an action with it. It offers an interface that allows the creation of *if-then rules* and then associates actions to it.

Authoring Tools. We include in this analysis, projects that provide authoring tools, because we found some similarities with IVO, although intended for specific domains. They enable the rapid authoring of applications by domain specialists. Hull et al. [19], propose a framework that allows the rapid authoring of mediascapes. The applications developed can be downloaded and executed on handheld computers augments with sensors. UbiCicero is a multi-device, location-aware guide supporting museum visits, which also provides the possibility of enriching the museum visits through individual or collaborative games [20]. It allows museum curators to easily customize the contents for their museum guides through an authoring environment.

Besides these projects, we also found some similarities in standalone applications for the Android platform, as Tasker [21] and Locale [22], which seek to automate tasks based on context information. These applications allow users define in the smartphone, a set of tasks that are triggered when a defined condition occurs, while

IVO allows the creation of more complete applications. We allow the easy definition of contextual information associated with each situation and the activities to be performed in a workflow logic. IVO also allows the sharing of applications through a distributed web platform and the location and temporal contexts of IVO are richer.

IVO differs from these projects in many ways. IVO allows the creation of fully functional applications with a wide range of uses. Our focus was on the reaction to context and not on the acquisition. Thus it provides a workflow environment for the actions which allows developing more complex applications. IVO's user interface is web-based which facilitates its use globally. Finally, IVO uses widely available smartphones as the ubiquitous interaction device, therefore allowing for a widespread use of the applications created by the end-users.

3 The Solution

To address the lack of support for end-users to build context-aware applications based on current smartphones, our approach was to provide end-users with a toolset that should allow them to: (i) easily define spatial, temporal and proximity conditions, and associate them with workflows that are triggered when the defined conditions occur; (ii) easily make the created applications available to other users through a distributed web platform, making the system available worldwide; (iii) create applications that can be used for various purposes from leisure to a more professional or business use; and (iv) run applications in currently available smartphones like Android, iPhone and Windows Mobile devices.

To address the first requirement, we developed a web application, called IVO Builder, which allows the creation of applications based on location. Outlook Calendar and Outlook Contacts were chosen for the user definition of temporal and proximity contexts, therefore allowing users to make use of their previous experience with these kind of popular applications. For the definition of the temporal conditions, we have created an Outlook Calendar *add-on* that allows the incorporation of workflows in events, which are executed "*when the event starts*" or "*when the event ends*". For proximity contexts, we have created an *add-on* for Outlook Contacts, which facilitates the discovery of Bluetooth devices in the proximity and associates the selected device to the contact. We named both *add-ons* as IVO Outlook. The synchronization tools provided by the device manufacturer guarantee the synchronization of IVO Outlook with the smartphone.

The second requirement was solved with a REST web-service which provides the services that the ubiquitous device (the smartphone) can use for the synchronization of applications that the user has access to. When creating an application the user becomes its owner and decides who can use it. The developed applications can be set to private (only the owner can use them), public (available for all users) or shared (only the selected users can use them).

The set of activities available in workflow design enables users to produce applications targeting different uses, e.g. personal or business, which addresses the third requirement. Possible activities range from changing the phone profile, integration with social networks or GPS navigation, to forms and quizzes, web-services from third parties, text-to-speech or voice recognition. The complete set of possible activities is explained later in this paper.

Finally, an Android client was developed providing the first fully functional client of the system running on one of the most widespread mobile platforms. Furthermore, this Android client allowed for a user evaluation of the builder tools.

We can identify three core features in IVO's architecture: a rule based system, an event-driven workflow and a visual and control-flow programming environment.

Rule Based System. We base the system on *if-then rules*, meaning things like "*if I* ..." or "*when I* ..." am in a particular situation, perform this action. Most context-aware applications can be described in that way. The study conducted by Dey et al. [13], on the way users think about context-aware applications, concluded that most users express themselves in terms of *if-then rules*. The conditions of *if-then rules* are evaluated through the Java Expression Language (JEXL) [23], using context information sensed by the smartphone. The use of JEXL allows a great expressiveness of the conditions with the use of common operators like >,> =, <, <=, =,! =, AND and OR.

Event-Driven Workflow. We use an event-driven workflow paradigm, in which workflows are started by events that result from the evaluation of the *if-then rule* engine. Workflows determine the program flow, expressed as a sequence of activities, which can include alternative or conditional paths and cycles.

Visual and Control-Flow Programming. Control-flow programming is used to build workflows because it makes it easy for the user to understand how things connect, unlike with imperative programming, which focuses on how things happen. We use visual programming environments to create the complete workflow and the rest of the elements that make up the applications, since these type of environments have proven to be effective in taking advantage of the user's spatial reasoning skills [24].

4 The IVO Interfaces

Fig. 1 illustrates the main screen of the IVO Builder, loaded with an application for a Tourist Guide of Lisbon. The system gives access only to registered users, with each user developing his/her own applications in their area of the server. Each of them can, however, make the developed applications available to other users.

On the top area of the main screen (zone 1 in Fig. 1) there are options that allow the application management, including loading, creating, changing, deleting and printing. The left side of the screen (zone 2 in Fig. 1) is reserved for the geographical representation, which is overlaid with the IVO areas of interest, represented in green. The toolbar in this area gives access to basic features for creating areas of interest, namely: (i) manual creation of an area by defining the polygon that defines it; (ii) creation of an area represented by a circle; (iii) import areas from geo-referenced articles on Wikipedia; (iv) import areas from existing areas in Wikimapia. In addition to the toolbar, there are the usual controls to zoom and change the map view, satellite or hybrid, and a search box that allows users to quickly find a place through Google search APIs. The right side of the main screen (zone 3 in Fig. 1) is reserved for the representation of information on the application itself, such as areas of interest, forms, quizzes, files (including images, audio, video and scripts) and permissions to other users. The forms built with the IVO Builder are compatible with XForms [25].

For more advanced users, we have a workflow activity for executing scripts. The scripting languages available are those natively supported by the smartphone platform in use. For example, on Android we use the Scripting Layer for Android (SL4A) [26] which allows executing scripts coded on Python, Perl, JRuby, Lua, BeanShell, JavaScript, Tcl, and shell. These scripts have access to many of the Android APIs, which makes it possible to create activities that are not viable otherwise, when using only the IVO workflow activities.

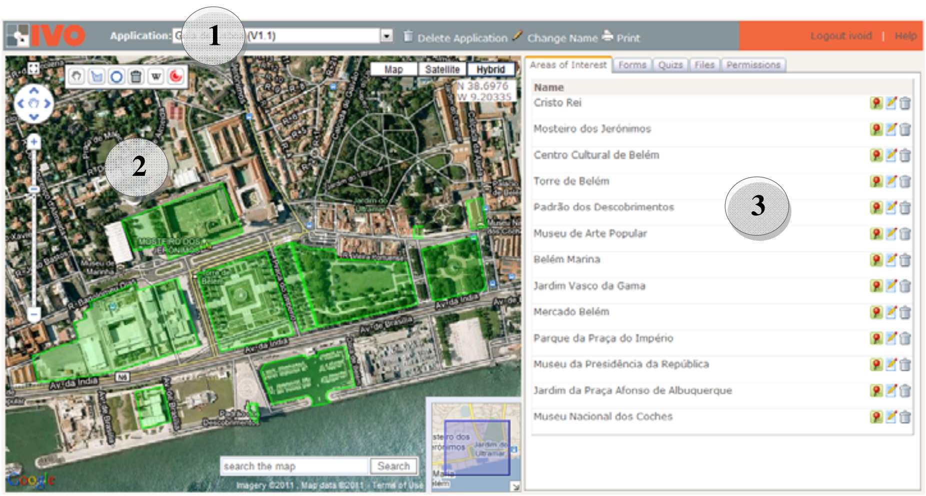


Fig. 1. Main screen of IVO Builder, loaded with an application for Tourist Guide. 1 – Application manager menu with options for loading, creating, changing, deleting and printing; 2 – The map helps the user to create location contexts. The areas of interest of IVO are highlighted in green. 3 – This area contains IVO application components, like the representation of the location contexts, forms, quizzes and files including images, audio, video and scripts.

Once created the areas of interest, the user can associate them with contextual information (Fig. 2a), using an HTML editor, or linking them to a website that will provide this information. On the same screen, the "Workflows" tab allows the creation of the entry workflows (Fig. 2b), which are executed when the user enters the area, and of the exit workflows, that are executed when the user exits the area. The example below shows (a) the information displayed to the users when entering the area of Mosteiro dos Jerónimos, Lisbon, Portugal (Wikipedia page) and (b) the entry workflow marked with green arrows (pointing right) as well as the exit workflow marked with red arrows (pointing left). In this case, when the user enters the specified area, the phone profile changes to silence mode and the user’s status on facebook is updated with the defined message. When the user exits the area, the phone profile changes to normal mode and IVO launches a quiz about the point of interest the user has just visited. After completing the quiz, the user is guided to the next point by the navigator.

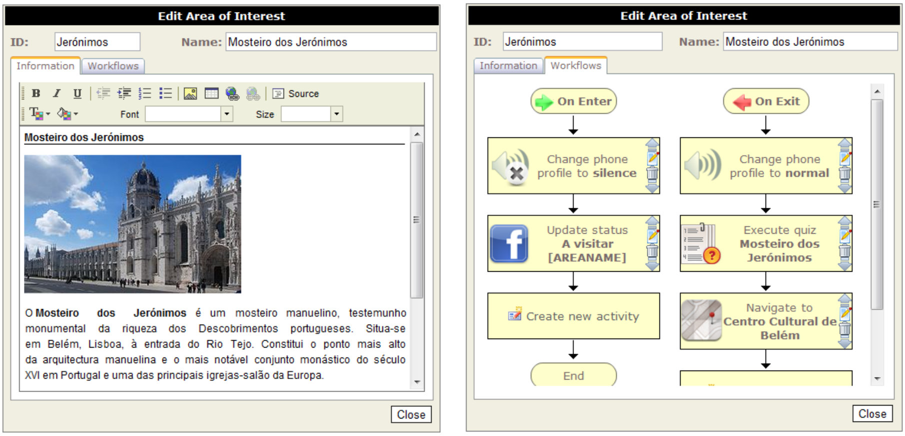


Fig. 2. Screens for the edition of areas of interest, where the user can define (a) the information about the point of interest and (b) the entry and exit workflows

Table 1. Activities currently available for the workflow construction

Icon	Activity	Icon	Activity	Icon	Activity
	Change phone profile		Update status on facebook		Execute external web-service
	Make a call		Open a web page		Read IVO code (QR-code)
	Send a SMS		Enter a value		Detect nearby people or equipment
	Send a email		Speech the text entered (text-to-speech)		Run script
	Don't show information window		Show alert message		Run external application
	Play audio file		Show toast message		Navigate to a place
	Play video file		Complete form		Take a photo
	Post message on twitter		Execute quiz		Voice recognition

Workflows are flows of activities controlled by the contextual data gathered by the smartphone as defined by the users. Presently, we only support sequential workflows, but we have already started the development for the support of more complex workflows with conditional or alternative paths and cycles, represented through an *event-driven finite state machine*. The activities that can currently be performed at each step of the workflow are illustrated in Table 1. The creation of more activities is a task that involves programming an abstract class extension.

The IVO Builder includes tools for creating quizzes and forms, which can be used as workflow activities.

A quiz can contain multiple questions with a maximum of four possible answers each, and it is possible to configure how they are displayed on the IVO client. For example, a minimum score (in percentage) can be adjusted to make the application immediately jump to the next workflow activity when the user reaches that score. The quiz can also be spoken to the user using the text-to-speech engine, when available on the client side.

The forms built with the IVO Builder are compatible with XForms [25]. The developed Android IVO Client includes an XForms processor that is able to run that type of forms. Fig. 3a) shows the screen to create a form used in an application for the agricultural sector while Fig. 3b) shows the same form running on the IVO Client. Data supplied by users while completing the form are stored on the server, allowing them to view these data as predefined reports or export them for external use.

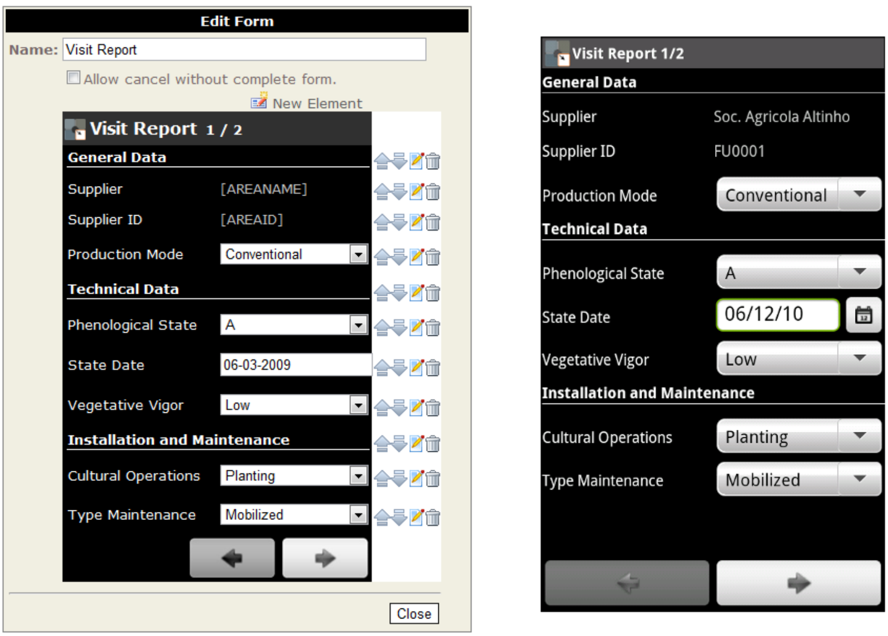


Fig. 3. a) The design of an IVO Form with a form loaded. b) The same form running on the Android IVO Client.

As mentioned before, the temporal conditions are defined in Outlook Calendar. This IVO Outlook *add-on* adds a form region at the bottom of the appointment window containing the interface that enables the users to define workflows with the same activities that are available in IVO Builder and pointed above. Fig. 4 illustrates an Outlook Appointment with a workflow that starts "*when the event starts*" and a workflow that starts "*when the event ends*". In this case, the activities defined by the user are: change phone profile and update facebook status. Users can associate the event to a location context created in IVO Builder or to the address field defined in user Contacts. This way, the system can handle more complex conditions like "*at this instant in time and when I'm at this location with these people*" do these actions.

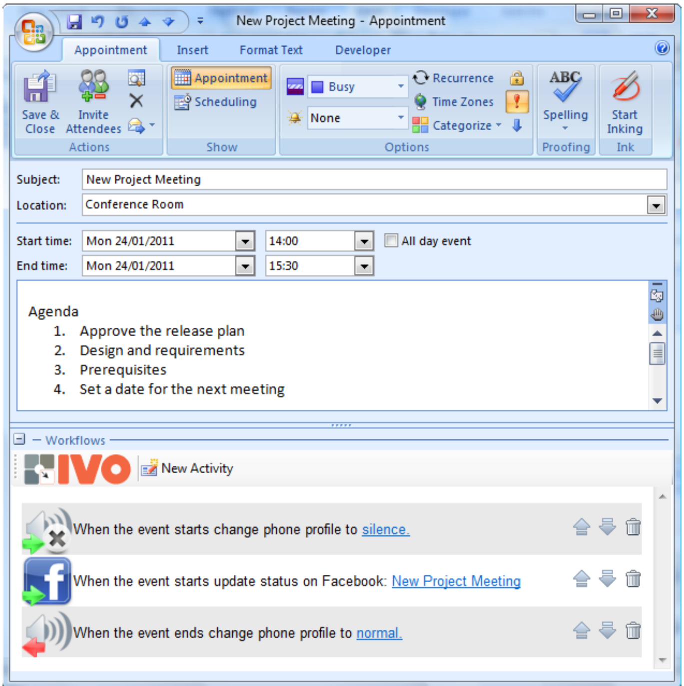


Fig. 4. Outlook Appointment with IVO workflows

5 Usability Evaluation

Usability tests played an important role in this work, because they helped us to observe different users while interacting with IVO tools. We conducted usability testing to determine and understand possible usability issues in IVO. IVO Builder and IVO Outlook were tested separately.

The IVO Builder has been tested by two groups of volunteers. A group of eleven graduate students in Computer Science with two females among them and a group of six Design college students being one of them a female. With these two groups we intended to assess the extent to which the area of expertise of the users would influence the results. The subjects for this experiment ranged in age from 20 to 43 years old with a mean age of 23.8.

IVO Outlook has been tested by a group of five graduate students in computer science (all male) and a female PhD student in Computer Science. All the participants had their first contact with the system during the test and used it under similar conditions. The subjects for this experiment ranged in age from 21 to 33 years old with a mean age of 24.5.

The test sessions were conducted individually for each user by two investigators, who played the roles of facilitator and observer. Before starting using the system, users were informed about the objectives of the test, and received a list describing the tasks they should perform (no reference on how to do it was made). The users were encouraged to "think aloud" and tell what was going through their minds while using the system. This way, users could freely explore the interface and we could clearly identify the usability issues. During the tests, the observer took notes on the user's behavior and verbalizations while performing the tasks, and the duration of each test was measured.

5.1 Test Scenarios

The test scenarios used in this evaluation were designed in order to allow for the testing of most of the features available within the system. The details of the test scenarios are described below.

Tourist Guide. In this scenario, the user was challenged to create an application using the IVO Builder to guide tourists through points of interest in the city of Lisbon. It was intended that when tourists arrive at points of interest, contextual information about them is shown using Wikipedia as a source. The phone should also be placed in silent mode and a wall post should be posted on facebook, informing the user's friends that she/he is visiting the spot. At the end of the visit to each location, the phone's profile should be switched to normal mode and the GPS Navigator would guide the tourist to the next point of interest. At certain points, the tourist would be challenged to answer some questions (quiz) about the location just visited. At the end, the user had to respond to an electronic questionnaire to evaluate the roadmap experience.

Using IVO Outlook. In this scenario the user was asked to use IVO Outlook to create contacts and associate them with a Bluetooth device (the smartphone) in order to allow activities of type "nearby people", and to create appointments to which the user then adds workflows. The user must create a new appointment, and create workflows that will run at the beginning and at the end of the appointment. When the appointment starts, the phone should be placed in silent mode and a message should be posted on facebook. The phone must also speak a message notifying the user for the start of the appointment. When the appointment ends, the phone should be switched to normal mode.

5.2 Questionnaire

After performing the usability tests, users were asked to answer a questionnaire which captured personal data, experimental feedback, as well as users' suggestion and comments. Personal data included the participant's age, gender, education level, familiarity with new technologies and frequency of use of Internet, computer, mobile phone, and game console. Experimental feedback was evaluated through four different sections of the questionnaire. The first section included three statements regarding the application easy of learning (Table 2). Users indicated their level of agreement with each statement by circling a value on a 5-point Likert-type scale, with a response of 1 (one) meaning "strongly disagree" and a response of 5 (five) meaning "strongly agree".

The second section included seven statements concerning the application ease of use criteria and one question on how easy it was to perform the tasks available in the application (Table 2). To classify the seven statements users followed the same procedure used for the first section as described above. To answer the last question, a scale from 1 (one) meaning "very difficult" to 5 (five) meaning "very easy" was used.

Table 2. Statements that composed the experimental feedback part of the questionnaire of IVO Builder. IVO Outlook questionnaire differs only in the tasks of question 11 which are: a) Create an IVO Contact; b) Create an IVO appointment and c) View IVO contacts and appointments.

Statement	
Ease of learning	
Q1	It is easy to learn to use the application
Q2	I quickly became able to use the application
Q3	I easily remember how to use the application
Ease of use	
Q4	It is easy to use
Q5	It requires few steps to accomplish what is intended
Q6	It can be easily used to build other applications, scenarios or situations
Q7	I can use it without written instructions
Q8	I don't notice any inconsistencies as I use the interface
Q9	It was easy to recover from an unexpected situation
Q10	It reacted quickly to user actions
Q11	How easy it was to:
	a) Create application
	b) Create an area of interest by importing data from Wikipedia
	c) Create an area of interest by importing data from Wikimapia
	d) Manually create an area of interest (polygon)
	e) Change the polygon that defines an area of interest
	f) Build a quiz
	g) Build a form
	h) Define workflows

In the third section users were asked to indicate the three activities they considered most useful and the three ones they considered less useful. An open question allowed users to suggest other activities they consider useful to be available in IVO builder.

The fourth section included one question based on the Microsoft "Product Reaction Cards". These aimed at capturing the user's feelings when using the system, since they facilitate the measuring of intangible aspects of the user's experience [27]. Users were asked to choose the words that best described their experience while using the system from a list of 24 words, consisting of about 60% of words considered positive and 40% considered negative. Users could choose how many words they wanted.

Finally, the questionnaire also included an open question in order to gather comments and recommendations regarding future developments of features and to obtain a more general evaluation of the system.

6 Results

The users' answers to the questions presented in Table 2 were analyzed and the results are shown and discussed below. We examined the average scores of the users' responses to see if there were general trends in their opinions (strong feelings one way or the other showing up as mean scores closer to 1 or 5). We used the standard deviation of the mean score to determine how broad the consensus about the issue was. Fig. 5 and 6 summarize the results for the questions presented in Table 2 concerning the IVO Builder and the IVO Outlook respectively.

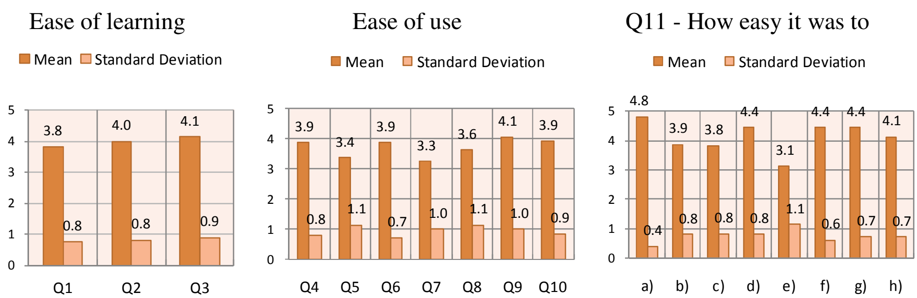


Fig. 5. IVO Builder results

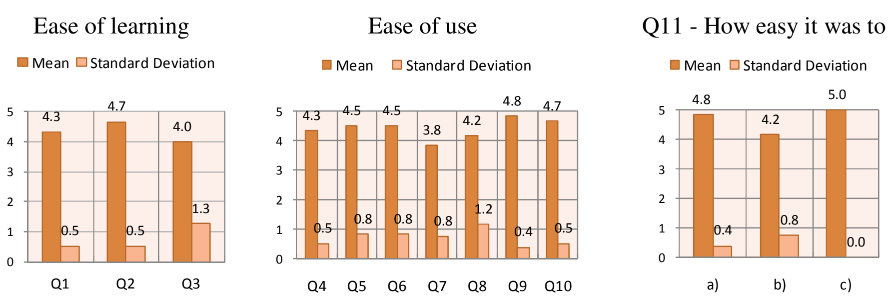


Fig. 6. IVO Outlook results

In general, the user's feedback was quite positive. Question 7, regarding the ability to use the application without written instructions, received the lowest score (although positive) for both applications.

Overall, users rated IVO Outlook better than IVO Builder, proving that it was easier to create temporal contexts than location contexts for the testing scenario. The biggest difficulty in IVO Builder was found in the task to change the polygon that defines an area of interest (Q11e). As we have already anticipated this difficulty, we had prepared written instructions that were provided to the users only in case they were struggling. After providing the instructions, users were able to accomplish the task without any difficulty.

IVO Builder. All participants were able to accomplish all tasks, although there were some occasional difficulties, especially when they were asked to change the polygon defining an area of interest. The time to complete the test ranged from 28 minutes to 1 hour and 5 minutes. The two groups that performed the tests**,** did not show a significant difference, which means that, in this case, the users' different areas of knowledge did not influence the results. However, some curious behaviors were observed in group 2 (design group), typical of desktop applications users, like the use of CTRL-Z to undo, BACK key to delete and drag and drop in some operations.

Most participants stated that it's easy to remember how to use the tool (statement 3) and they quickly became skillful with it (statement 2). Although very positive, the users' feelings about the statement "It is easy to learn to use the application" (statement1) was not so strong.

The majority of the participants stated that it was easy to use the IVO Builder (statement 4) and that it was easy to recover from an unexpected situation (statement 9). Although positive, the participants' opinions concerning the number of steps to accomplish what is intended (statement 5), the application use without written instructions (statement 7) and the inconsistencies found when using the interface (statement 8) show lower scores and higher standard deviation revealing weak consensus between users. The majority regarding the application's "ease of use" (Q4-Q10) results are positive, but the answers to four of the questions (revealed a standard deviation higher or equal than 1, revealing a lower consensus in users' opinions.

Question 11, assessed how easy users found each of the proposed eight tasks (described in Table 2). According to the results displayed in Fig. 5, the interactions were generally deemed easy to perform. The only exception was the task to change the polygon that defines an area of interest (statement 11e) which users found not so easy to perform. Their opinions were very heterogeneous and they conform to the users' observed behavior.

IVO Outlook. All participants were able to accomplish all tasks, without any difficulty, in a period between 9 and 16 minutes. The user's feedback was quite positive and consensual. Most participants agreed with all the statements. The lowest score was found in the statement about the application use without written instructions (statement 7). We considered it a very subjective matter, since the participants have not used any written instructions and we did not notice any difficulty when they were performing the tasks.

In the third section of the questionnaire participants were asked to indicate the three activities they considered most useful and the three they considered less useful. Analyzing the results, we conclude that it seems to be a pattern in the users' responses regarding the more useful activities. Users tend to refer those used in the test: change phone status (20 users), navigate to (12 users) and update status on facebook (7 users). The activities considered less useful were "voice recognition" (7 users) and "run a script" (4 users). We believe that, since participants haven't used these activities during the test, they may not have realized how they can use for.

From the analysis of the fourth part of the questionnaire (Fig. 7) we concluded that 90% of the participants held positive feelings when classifying their experience using the system. The most selected word was "useful", followed by the word "simple" and

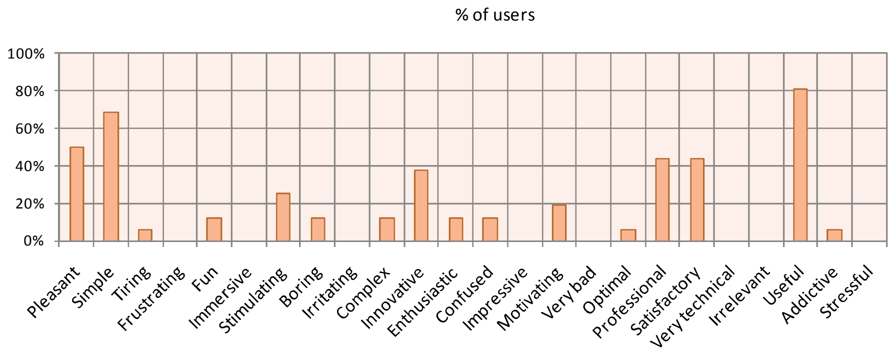


Fig. 7. User's feelings when using the system

"pleasant". 44% of the participants considered the system "professional" and "satisfactory", and 38% "innovative". Only two users reported negative feelings and the words used were "confused", "complex" and "boring".

Overall, the evaluation results were very positive, the users expressed very encouraging feelings about the tested tools and they declared themselves very willing to use it as they found it very useful.

7 Conclusions and Future Work

We have presented IVO, a platform for building context-aware applications by end-users without the need to write any code, using smartphones as the ubiquitous interaction device. It opens up the space of context-aware computing to end-users, allowing them to build fully functional context-aware applications.

The phone is considered the first truly ubiquitous device. Most top end smartphones come with a handful of sensors and it is expected that in future new sensors will be added, extending the spectrum of use for context-aware applications.

The event-driven workflow approach enables the development of a wide range of applications. An application consists of *if-then rules* and workflows that are triggered when the conditions occur. We plan to support more sophisticated workflows with alternative paths, cycles and also new activities, in order to make the system even more expressive. We also intend to improve the building tools with modules for testing and debugging.

The tests allowed us to conclude that even first-time users could build context-aware applications through the platform and revealed some minor interface issues, already solved. They also showed that visual programming environment facilitates the rapid development of such applications.

IVO is now starting to be used in a project in the area of Persuasive Computing aiming at enhancing people's behavior towards the environment. Is this context the developed applications should induce positive changes in the way users deal with daily activities such as recycling and energy or water consumption. As future work we also want to develop IVO clients for other mobile platforms, such as the iPhone, therefore reaching a larger number of users.

Acknowledgments. This work was funded in part by ILUSTRATOWN, by Agência de Inovação (ADI) under contract 70/2007/33B/00216/00178 and by FCT – Fundação para a Ciência e Tecnologia, Portugal, in the scope of project DEAP (PTDC/AAC-AMB/104834/2008).

References

1. Weiser, M.: The Computer for the 21st Century. Scientific American 265, 94–104 (1991)
2. Greenfield, A.: Everyware: The Dawning Age of Ubiquitous Computing. New Riders (2006)
3. Brett, W., Matt, L., Brian, L., Kristofer, S.J.P.: Smart Dust: Communicating with a Cubic-Millimeter Computer. Computer 34, 44–51 (2001)
4. Want, R., Borriello, G., Pering, T., Farkas, K.I.: Disappearing Hardware. IEEE Pervasive Computing 1, 36–47 (2002)
5. Rafael, B., Jan, B., Michael, R., Jennifer, G.S.: The Smart Phone: A Ubiquitous Input Device. IEEE Pervasive Computing 5, 70 (2006)
6. International Data Corporation (IDC), `http://www.idc.com/about/viewpressrelease.jsp?containerId=prUS22560610`
7. Finkenzeller, K.: RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification. John Wiley & Sons, Chichester (2010)
8. Dey, A.K., Hamid, R., Beckmann, C., Li, I., Hsu, D.: A CAPpella: programming by demonstration of context-aware applications. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 33–40. ACM, Vienna (2004)
9. Schilit, W.N.: A System Architecture for Context-Aware Mobile Computing, PhD thesis, Columbia University (1995)
10. Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The Anatomy of a Context-Aware Application. Wireless Networks 8, 187–197 (2002)
11. Dey, A.K., Salber, D., Abowd, G.D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Human-Computer Interaction 16, 97–166 (2001)
12. Li, Y., Hong, J.I., Landay, J.A.: Topiary: a tool for prototyping location-enhanced applications. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, pp. 217–226. ACM, Santa Fe (2004)
13. Dey, A., Sohn, T., Streng, S., Kodama, J.: iCAP: Interactive Prototyping of Context-Aware Applications. In: Fishkin, K.P., Schiele, B., Nixon, P., Quigley, A. (eds.) PERVASIVE 2006. LNCS, vol. 3968, pp. 254–271. Springer, Heidelberg (2006)
14. Chen, G., Kotz, D.: Solar: An open platform for context-aware mobile applications. In: Proceedings of the First International Conference on Pervasive Computing (Pervasive 2002), pp. 41-47 (2002)
15. Hong, J.I.: The context fabric: an infrastructure for context-aware computing. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems, pp. 554–555. ACM, Minneapolis (2002)
16. Bardram, J.E.: The Java Context Awareness Framework (JCAF) - A Service Infrastructure and Programming Framework for Context-Aware Applications. In: Gellersen, H.W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 98–115. Springer, Heidelberg (2005)

17. Hightower, J., LaMarca, A., Smith, I.: Practical Lessons from Place Lab. IEEE Pervasive Computing 5, 32–39 (2006)
18. Klemmer, S.R., Li, J., Lin, J., Landay, J.A.: Papier-Mâché: Toolkit Support for Tangible Input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 399–406. ACM, Vienna (2004)
19. Hull, R., Clayton, B., Melamed, T.: Rapid authoring of mediascapes. In: Davies, N., Mynatt, E., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, pp. 125–142. Springer, Heidelberg (2004)
20. Ghiani, G., Paternò, F., Santoro, C., Spano, L.D.: UbiCicero: A location-aware, multi-device museum guide. Interact. Comput. 21, 288–303 (2009)
21. Tasker for Android, `http://tasker.dinglisch.net/`
22. Locale for Android, `http://www.twofortyfouram.com/`
23. Java Expression Language (JEXL), `http://commons.apache.org/jexl`
24. Shu, N.C.: Visual programming: perspectives and approaches. IBM Systems Journal 28, 525–547 (1989)
25. The Forms Working Group, `http://www.w3.org/MarkUp/Forms`
26. Scripting Layer for Android (SL4A), `http://code.google.com/p/android-scripting`
27. Benedek, J., Miner, T.: Measuring Desirability: New methods for evaluating desirability in a usability lab setting. In: Proceedings of UPA (Usability Professional' Association). Orlando, USA (2002)

Hammering Models: Designing Usable Modeling Tools

Ko-Hsun Huang[1,2], Nuno Jardim Nunes[1], Leonel Nobrega[1], Larry Constantine[1], and Monchu Chen[1]

[1] Madeira Interactive Technologies Institute and University of Madeira, Campus Universitário da Penteada, 9000-390 Funchal Portugal
[2] Institute of Applied Arts, National Chiao Tung University, Hsinchu, Taiwan
kohsun@csie.nctu.edu.tw,
{njn,lnobrega,lconstantine,monchu}@uma.pt

Abstract. A modeling tool not only helps users express their ideas and thoughts but also serves as a communication platform among domain experts, designers, developers, and others practitioners. Existing modeling tools have shortcomings in terms of supported functionality and situated usability or do not meet the needs of users of varying levels of expertise. To facilitate improvement of such modeling tools, this research begins by identifying common problems in existing tools and proceeds by borrowing concepts from grounded theory to develop a framework of redesign guidelines. A case study illustrates how this framework can be used by applying it to MetaSketch, a metamodeling tool. The study employs multiple user experience research methods, including usability tests with paper prototypes, observations, interviews, and contextual inquiries. A set of core tasks and two significant modeling approaches were identified that directly influence interface and interaction design for modeling tools.

Keywords: Model-Based Design, Interactive Systems, Usability and Software, User Experience Design, Metamodels, Metamodeling, Participatory Design.

1 Introduction

Software developers and designers have long complained that the "tools of the trade" are unsupportive and unusable. With the advent of model-driven development, the same concerns apply to modeling tools. A good modeling tool not only helps its users express their ideas and thoughts but also serves as a communication platform among domain-experts, designers, developers, and others practitioners. These various usages must be recognized and supported in varied ways. The primary purpose of this research is to identify and understand the significant work styles of potential tool users and to identify design implications for modeling tools [1-3]. We also reconsider the functionality a metamodeling tool can support, such as an iterative modeling process, giving domain-experts greater freedom to alter metamodel definitions based on knowledge emerging from case and field studies.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 537–554, 2011.

1.1 Modeling Tools

Software development is still dominated by so-called third-generation programming language (3GLs) introduced over a half-century ago. Computer-Aided Software Engineering (CASE), the first modeling tools, arose from the so-called software crisis and the assumption that better tools would help programmers create better software [4]. Despite the partial failure of such tools [3], CASE did bring more advanced tools to the basic toolset of editor, compiler and debugger. Integration of fourth-generation programming languages (4GLs) with CASE tools enabled creation of high-end tools that boosted productivity. However, editors, compilers and debuggers, now combined into Integrated Development Environments (IDEs), are still the primary tools of software developers.

It has been argued that models might improve software engineering as they have in traditional engineering disciplines [5]. Model-driven development is no panacea, and it is not without controversy [6]. However, models and modeling languages make possible raising the level of abstraction [7]. Models can also help bridge the gap between developers and other stakeholders, in particular making application domain experts active participants in development. The growth of UML (Unified Modeling Language) and UML-based tools suggests that modeling has potential in typical software development [6]. Despite debate about models in software engineering and the role of UML [8], it is widely accepted that UML lacks the expressiveness needed to address domain specific knowledge, and, therefore, new modeling languages are required. Domain-specific languages have proved to deliver gains when compared with general-propose languages [9]. Capturing domain specificities is crucial to understanding problem contexts and, therefore, finding correct solutions. Domain-specific languages enable creation of models derived from a domain and are more easily understood by domain experts. These languages enable domain experts to contribute directly to problem definition and, by better describing needs, improving software development. Defining domain-specific modeling languages is not easy [7], and the Software Language Engineering discipline [10] has emerged to help.

Multi-language IDEs are common, but multi-language modeling editors are unusual. Benefiting from UML standardization, most popular modeling editors support only UML and its concrete syntax instead of supporting different notations as before. To support effectively the definition of models using domain-specific modeling languages, we need the capability to define these specialized languages. The Generic Modeling Environment (GME) [11] and notably the MetaEdit+ Domain-Specific Modeling environment [12] brought this capability to Domain-Specific Modeling (DSM), where new modeling languages could be defined for particular domains. Approaches like Software Factories [13] and the Eclipse Modeling Framework (EMF) [14] also open new opportunities for DSM.

With the UML 2.0 specification, the Object Management Group (OMG) concluded a long standardization process that not only improved UML as a language, but also built the basis for OMG's Model Driven Architecture (MDA) initiative to support model-driven engineering (MDE)—the concept that software development methods should be focused on creating and exploiting domain models rather than the underlying computing or algorithmic concepts. MDE is meant to increase productivity by maximizing compatibility between systems, simplifying the process of design, and promoting communication among individuals and teams developing a system [15]. The UML 2.0 language is the most popular implementation of the MDE approach, providing support for modeling the different aspects of a system using different levels of abstraction as proposed in MDA.

1.2 Structure of the MetaSketch Workbench

MetaSketch is a modeling language workbench that allows the rapid development of new modeling languages for model-driven engineering. MetaSketch is based on OMG standards including MOF 2.0, OCL 2.0, and XMI 2.1, and is specially tailored for creating new members of the UML family of languages. Currently, the workbench comprises three pre-beta release tools: the Editor, the Merger and the Designer:

– MetaSketch Editor. The editor, the heart of the workbench, is simultaneously a modeling and a metamodeling editor, meaning that the same editor can be used to create the metamodel and the models that conform to the created metamodel.
– MetaSketch Merger. This tool is only applicable at the metamodel level and basically resolves all the Package Merges used in the definition of a metamodel. In addition, the hierarchy of packages is flattened into a single package that contains all language construct definitions merged.
– MetaSketch Designer. This work-in-progress enables the graphical creation of the concrete syntax for each language. Currently, these definitions have to be done directly in the XML files.

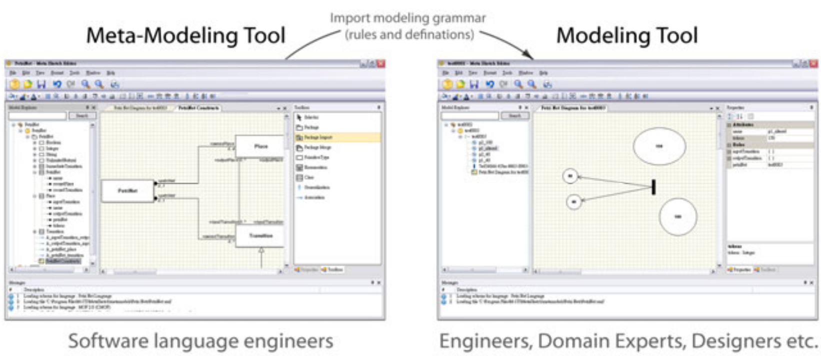


Fig. 1. Original structure of MetaSketch. On the left is the tool for software language engineers developing UML languages (metamodels). On the right is the tool enabling domain-experts, designers, developers, and others practitioners to develop models.

The original version of MetaSketch allows software language engineers to define metamodel languages and supports other users—such as designers, developers, and other practitioners—in creating models using a similar interface for metamodel definition and modeling (Fig. 1). This design does not comprehensively support users of different levels of expertise and varied work styles. Whatever kinds of models or metamodels are used, the interface is always the same.

To redesign MetaSketch, we started to reconsider who are the target users and what are their main usages. For software engineering, the current version of MetaSketch already supports quite well the standard metamodel development process. The main shortcoming is the lack of an advanced designing tool to help language engineers define the concrete syntax for new language constructs. For other users, especially field-study researchers and designers, the interface is too complicated. These users are concerned with creating models for expressing concepts and knowledge and would prefer an interaction style closer to conventional editing or drawing tools. Therefore the approach taken was to separate MetaSketch into several different parts to support different uses and to engage all potential user types.

To increase the flexibility and changeability of the metamodel and to support domain-experts in developing and revising metamodels from field cases, we reconsidered the relationship between modeling and metamodeling as an iterative process. We propose the comprehensive framework shown in Fig. 2.

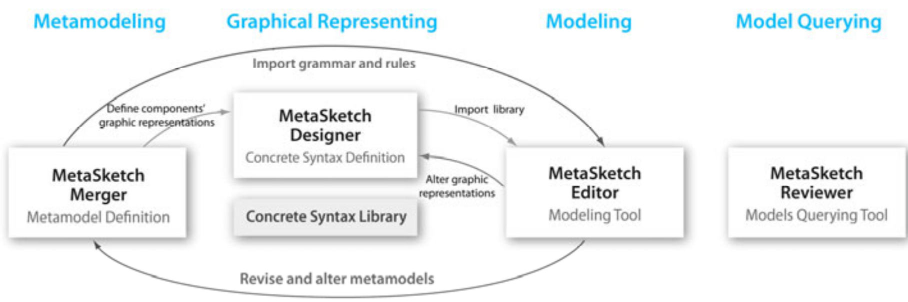


Fig. 2. Framework of modeling tool to support flexible and iterative modeling and metamodeling processes

1.3 Supporting Iterative Modeling Processes: Case-Driven Theory

We undertook a case-driven theoretical development process, borrowing its main concept from Grounded Theory [16,17], which is a systematic qualitative research method commonly used in the social sciences. Grounded Theory incorporates a variety of methods and an iterative process to build a theory from field data, including extracting concepts from cases and categorizing concepts into a theory. It provides a specific theoretical framework to build domain knowledge and produce a model for application.

Software engineers also define grammars and metamodels for domain specific situations. Consequently, before an entire grammar and rules become well defined, case collections and rough concept sketches of modelers can be considered as a part of meta-modeling development. Therefore, three concepts inspired by Grounded Theory are suggested for designing modeling tools:

- To develop a modeling language employing a field-based or case-driven process. A modeling tool can serve as the sketchpad, which gives researchers and domain experts more freedom to express their conceptualizations of field data.

- To transform cases and raw data into grammar and rules. This is the process of conceptualizing and categorizing data and then generalizing it into the most important representative form (the deductive and inductive process [18]).
- To define a metamodel and to use it to create models for cases through an iterative process. Real cases help software engineers reconsider and revise their original metamodel, and concrete syntaxes can also be improved by the evaluations of designers, researchers, and practitioners.

To support these three basic concepts, there are several functions a modeling tool like MetaSketch should provide in the future:

1. supporting field data collection through an ill-defined modeling language (e.g., a sketch on a napkin);
2. supporting different representations (concrete syntax) for different field data and concepts, therefore providing a modeler with more flexibility and expressiveness;
3. supporting the automatic categorization and contextualization of cases from raw data for future rule definition (generalizing theories from cases [19]).

2 State of the Art

Many studies have analyzed software development practices, but qualitative studies of modeling work practices are relatively rare. Some authors report a gap between how tools represent and manipulate programs and the software developers' actual experiences [20]. This work is mostly based on workstyle models for collaborative software design and shows how a tool can fail to match an intended work context. However, research on practitioners' workstyles helps in understanding how we can create new tools that better support specific workstyle transitions. A survey of how practitioners worked and used tools offered insights into how developers can build human-centered tools but did not focus on modeling tasks [20].

To understand the potential usage and meaning of a tool, it is necessary to find out the needed and desired aspects [21] and to consider the future context of use in the whole design process [22,23]. Usage of modeling tools is hard to predict because they are involved in an uncertain and complex creative process. In addition, a modeling workbench, such as MetaSketch, aims to support multiple types of users, including designers, domain experts, engineers, practitioners, and field researchers. It supports communication and coordination of perspectives among all the users, facilitating a process where these stakeholders can jointly transform their knowledge [24,25]. The research methods used in this study not only have to cover the details of usage and interactions, but also need to reflect the different workstyles and attitudes of users of different levels of expertise [20,26]. To enhance the quality and usability of the system, and to satisfy all types of users, research approaches from both user-centered design and participatory design were considered jointly [25,27,28].

To determine the knowledge required for performing modeling tasks and to understand users' mental models, several techniques for tracing and modeling cognitive and behavioral processes in system evaluation were applied, including task analysis, protocol analysis, observation, and interviews [29-31]. Techniques to describe how people actually work and to evaluate the current version of MetaSketch—such as, cognitive walkthrough, contextual inquiry, and rapid prototyping—were also considered [32-34].

There are two main challenges in this research. First, most research techniques for usability evaluation focus on interface or task-oriented planning [28,29]. However, modeling is an iterative and creative activity, which reflects the idea and thought of modelers and is an important material to help them think. Assigning users set tasks for observing and analyzing is not suitable in our case.

The second challenge is to investigate and to explore modeling behavior with low-level detailed analysis of interactions. Since the activities of modeling are not clear, to predefine an appropriate set of interactions for each task becomes very difficult. Prototyping at this stage may not reflect the richness of desired behaviors and lacks specification of low-level details of interactions [25,26].

Sketches, storyboards, and wireframes are widely used for interface and interaction design [35,36], and many varied materials and tools support designers in expressing ideas and simulating behaviors [26,37]. However, many techniques and materials are applicable to static interface design, such as the appearance, layout, or graphics for a website, but not for dynamic aspects or details of complex interactivity [38]. A recent survey of 259 practitioners found that 86% designers considered behavior more difficult to prototype than appearance [26]. The perceived and actual properties of a tool also affect possible usage [39,40].

We adapted the paper prototyping approach to make it more flexible for exploring possible functions users expect from a modeling tool. We made all components of the paper user interface independent, movable, and reusable, including system dialogs, tools, and other objects. Think-aloud and Wizard-of-Oz protocols were applied for understanding users mental models. Interactions were recorded by video and later interpreted for task analysis. After each usability test, participants were interviewed about their personal experiences and opinions about modeling tools. We also debriefed end users with the results and discussed how to refine the paper prototype and system response (see next section) [41-44]. The system responses to be simulated by the paper prototype were performed by a so-called “human computer” who was given a text description defining a series of default system responses as operational sequence diagrams covering all anticipated user actions and situations where users want to understand the system behaviors, including by trial and error interactions.

The adaptability of our paper prototype made the usability test an iterative process. In the end, all the tasks involved in modeling were elicited, and all potential interactions to achieve these goals were discussed. The results not only reveal the intentions and strategies that users have in building models but also indicate their expectations for a modeling tool.

The paper prototype is not merely an evaluations tool, but a material for designers to convey their design concepts as well as a platform for communication among domain experts, engineers and practitioners. In this research, we explore all the aspects of modeling activity, and show an integrated approach for usability testing.

3 Usability Test for MetaSketch Editor

MetaSketch Editor is the first component selected for redesign in our process. The original interface of the editor is separated into four components (Fig.3). The Explorer is a tree-view of all entities involved in the model; the Diagram is the main canvas for modeling. The Object Library and Properties share the same window and enable users to add new objects into models and perform advanced editing tasks.

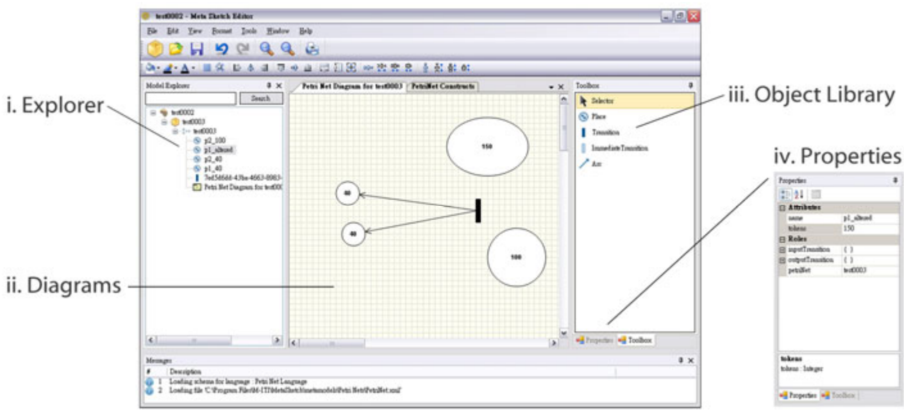


Fig. 3. The interface of MetaSketch Editor is divided into four components: Explorer, Diagrams, Object Library, and Properties

Initially we conducted a pilot study to gain better understanding of current modeling tools and their interaction design. This study was based on literature reviews and heuristic evaluations of current modeling tools, including meta-modeling tools, such as Eclipse Modeling Framework and MetaEdit+, and common modeling tools, such as Oryx, Aris Express, and Intalio. Based on the results, we discussed a diverse set of design concepts over a period of months with designers and software language engineers [10]. We designed the paper prototype to resemble well-known image-editing tools, and developed several hints and guiding functions to assist users in modeling tasks [45,46]. The prototype (Fig. 4) included basic editing tools: Object Library, Component Explorer, Navigator, Properties Window, system dialogs and hints. There are several new functions:

– Explorer and Diagram. The prototype Explorer provides additional options for users to examine the partial model and the entities of each diagram.
– Object Library. A context menu enables accessing the Object Library and adding different types of entities on the diagram directly.
– Properties. Additional information is provided for each entity, including which diagram or model an entity is involved in.
– Navigator. For editing complex models, users can easily check the undefined or unlinked entities through different filters in Navigator.
– System dialogs and Hints. Instead of traditional text-based descriptions, hints and grammar explanations are provided with graphic examples.

In redesigning MetaSketch Editor, we looked to three types of users: i) experts, who are familiar with a specific modeling language; ii) designers, who explain their concepts and ideas through a model; and iii) researchers, who reveal their knowledge of a certain field or situation through modeling. To understand the different mental models and work styles in all areas, we applied different user experience research methods for usability testing, in particular, paper prototyping [34], thinking aloud [34][47], Wizard of Oz [48,49], observation and interviews [50]. To further analyze the data, we also applied contextual inquiry [33] and task analysis [29].

Fig. 4. Based on the pilot study, a flexible paper prototype was designed to support different types of users, including domain-experts, designers, developers, and others practitioners

We defined a default set of responses to all possible user actions in order to integrate all functions of MetaSketch Editor with the new design concepts. During think-aloud tests, if users expected a different system response from the default, the "human computer" reacted in the way participants wanted, thus avoiding interrupting their current work process and helping to better understand their mental models.

For testing with engineers and designers we selected Activity Modeling and Participation Maps [51,52], since both were well understood by our potential subject group. Five participants were recruited, including four engineers and one designer. The four engineer subjects are experienced practitioners and domain experts, but only one was the software language engineer who developed Activity Modeling. All work in interaction design and are highly capable in usability tasks, such as, thinking-aloud and cognitive walkthroughs, and can give valuable feedback regarding low-level interaction details.

Two modeling strategies were distinguished in the pilot study: sequential and non-sequential. In sequential modeling, users build models one entity at a time. For instance, in flowcharting, users tend to add an entity, and then connect it immediately. In a non-sequential modeling strategy, the user first creates a number of entities and then considers potential connections among entities. Both approaches were observed in the pre-study. To understand if task order and human-computer interaction were influenced by different modeling approaches, the test was separated into two sections:

– Section I: A scenario was provided, which described a retail selling situation. Users were asked to create a Participation Map using the paper prototype. The scenario contained three paragraphs, each describing two or three entities and their relationships. In this section, users could be expected to model the situation using the sequential approach (Fig. 5).
– Section II: Users were given a Participation Map and were asked to recreate it using the paper prototype. In this section, users could be expected to use the non-sequential approach (Fig. 5).

In both sections, users were asked to think aloud to explain what they were looking for, thinking, doing, and what kind of system responses they expected. The purpose was to understand users' mental models and to evaluate our interaction design.

After each test, a brief, semi-structured interview was conducted. Users shared how they used materials or applications for modeling and gave their opinions on the design of the paper prototype [50]. All data was captured on video and later integrated into a task analysis table [41]. To represent the modeling approach applied by each participant in each section, we applied Sequence Model in Contextual Inquiry through a flowchart [53].

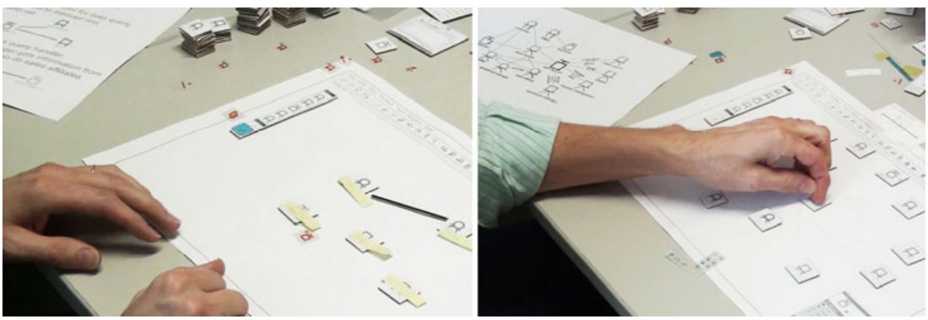


Fig. 5. The left picture shows the first section of the usability test, where the user was asked to create a model from a text-based scenario. On the right is the second section, where the user was provided with a sketch of a Participation Map to build using the paper prototype.

4 Primary and Secondary Modeling Tasks

From the pilot study and design concepts approximately a hundred potential tasks involved in modeling were identified, including adding new entities to the modeling canvas, connecting entities, editing properties or names of entities, and correcting errors. To help us order the frequencies of important interactions and make decisions for the interface design, we have to distinguish the primary and secondary tasks in the modeling activity.

Ten important modeling tasks were identified (Fig. 6). Primary tasks were: creating entities, naming entities, considering possible relationships among entities (creating links), annotating relationships (naming or editing links) and repositioning entities (moving objects). Secondary tasks occur when users need to correct errors, which are

Fig. 6. Left table: primary and secondary tasks in text-based section and in the sketch section. Right table: modeling approaches actually taken by each user in each section.

not the goal of modeling but often interrupt their current process. Secondary tasks include: changing the type of an entity, changing the type of relationship, and changing the direction of relationship (changing endpoints of links).

The primary modeling tasks are creating and connecting entities. We found out that the steps users took and their expectations of system response were highly dependent on the modeling approaches they applied.

- Creating Entities. In the text-based scenario, four of five users modeled the Participation Map using a sequential approach. A context menu enables users to choose entity type and create it on the canvas directly. However, most users still created entities from the Object Library. Users suggested making the context menu visible in the beginning or alternatively providing on-screen tips. For the non-sequential modeling approach, results show that the Object Library provides an easy and smooth way to create all the important entities on the canvas.
- Creating Links. The most diverse interactions were found in how users created links. All users selected the create-link tool from the Object Library in the beginning, but later actions and expectations of system responses depended on their modeling approaches and individual preferences. Three users wanted to create a link between entities by selecting the source and the target. While applying the sequential modeling approach, two users wanted the last clicked object to always become the source for the next action, thus allowing a link to be created by a single click: selecting the target entity. Other possible techniques for link creation discussed in interviews included clickable handles on each entity to make connection to others and applying a hot key.

- Annotating. Interviews did not indicate that task priority was influenced by the modeling approach. Users were more focused on annotating links in the non-sequential approach (Fig. 6), where editing links becomes one of the primary tasks. According to the interviews, adding a link in the non-sequential approach is one of the important parts of modeling. While users are creating a link between two entities, they are considering the type and details of this relationship. Naming an entity was considered part of creating an entity, and most users tended to insert the name immediately after adding an entity.
- Repositioning Entities. Two users mentioned that repositioning entities actually helps them think. Looking at and reorganizing the canvas helps make sense of the whole model. Hence, repositioning is an important independent activity, quite different from simple revision or editing of the model.

After considering Primary Tasks, we focused on Secondary Tasks, the most important being correcting errors. Since it is not the purpose of modeling itself, the way users corrected errors and revised models was unrelated to the modeling approaches used but was related to individual preferences in handling interruptions.

- Changing Entity Type. It is sometimes difficult to decide the type of entities while adding them to the canvas. Most users agreed that converting an entity into a different type is a frequent modeling task. They wanted to convert an entity into any other type even if that might lead, temporarily, to a grammatical error that might later be corrected. By habit and personal preferences, some users also tend to correct all errors toward the end. Modeling tools should avoid interrupting users with repeated error messages or reminders.
- Changing Relationship Type. Some users did not consider changing the direction of a link as changing the type. Adding an arrowhead was to them a part of annotation or decorating, like changing color or line weight. Instead of creating a directional relationship, they first created an undirected line and later tried to add an arrow by selecting the line and choosing an arrow from the formatting palette (Fig. 7). Whether the interaction design for this task should reflect its semantic connotation in a particular modeling language is unclear, particularly as users will want to be able to easily correct errors in direction.
- Moving Endpoints. With complex models, it is common to create links by mistake. Most users understood that changing endpoints might lead to a grammatically incorrect result but still expected the freedom to alter the model in any way they want. Instead of deleting incorrect links and creating new ones, users want to manipulate links easily, such as by dragging endpoints to new targets. As with changing entity type, the system should not limit user actions.

Tasks like editing properties, reviewing system messages, and checking guidelines and grammar rules are activities independent from modeling. Theses functions are important and should be supported by any practical modeling tool but serve very different purposes from modeling itself.

- Advanced Editing. In creating entities and building a model, setting entity properties is independent from other tasks. Users start considering properties after getting an initial picture of the model as a whole, hence properties should be considered in the second level of information architecture.

- Exploring and Querying. In both the Navigator and Explorer, filter functions provide different views of a model, allowing users to browse by types of objects, unedited objects, and unlinked objects. Users explained that the canvas can become very complicated for large models. Although the Navigator and Explorer were not used in the tests, most users appreciated having the function.
- Visual Refinement. Editing is not the primary modeling task. Positioning of entities represents information and helps the user to think, but may not be the main purpose of modeling. Common functions supported by image editing applications, like aligning and resizing objects, are not considered as core features of a modeling tool but could be of vital importance for modeling as communication.
- System Hints and Guidelines. To help first-time users learn the grammar and rules of a certain model, the system could provide a guiding tool with graphic examples. However, it is also important to avoid interrupting users. Most users did not want to be disturbed by system hints and messages. However, they also said they would like a grammar verifier to help them identify modeling errors.

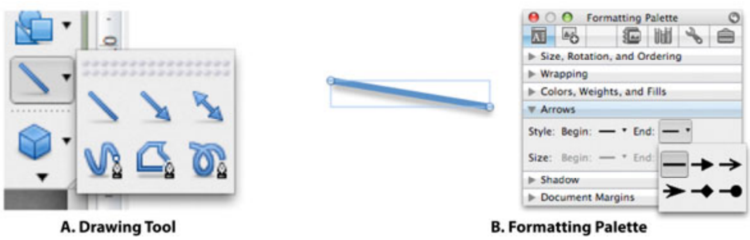


Fig. 7. Two ways to draw an arrow in common editing tools. Microsoft Word allows drawing an arrow directly with the drawing tool (A) or by drawing a simple line first, and then changing the style by adding an arrowhead from 'Formatting Palette' (B).

5 Modeling Approaches

The two modeling strategies from the pilot study, sequential and non-sequential, were both observed in the tests. As summarized in Fig. 6, three users took the sequential approach in the text-based scenario and the non-sequential approach in the sketch-based section. However, one user employed a non-sequential approach in the text-based section. Another used a sequential approach in the sketch-based test.

Fig. 8 shows the consolidated work models of sequential and non-sequential approaches. In the sequential approach, since connections are made after creating an entity, it becomes more difficult to change the type of any entity afterwards, which would require reconsidering all related connections. Moreover, it can take more steps to annotate existing relationships, not having considered meanings and direction.

In the non-sequential approach, the process is more complex. Users create a number of entities at once and consider their types and meanings afterwards. Later, they think about possible connections and the direction and specific meaning of relationships. To one user, creating links was an independent part of modeling and a well-defined task. In the non-sequential approach, the frequency of correcting links and changing their types would be lower than in the sequential approach.

According to the interviews, the rules of a modeling language and its graphic notation directly influence the way users select their modeling strategy. The pilot study indicated that models like flowcharts, sequence diagrams, and activity diagrams in UML, incline users to use a sequential approach. In contrast, use case diagrams and component diagrams in UML, which do not represent workflow or procedure, encourage non-sequential approaches. Testing revealed that personal preferences and habits also influence selection of modeling approaches. For instance, users who sketch before using modeling software and users expecting efficiency may prefer a sequential approach independent of the form of models. Users who consider modeling applications as tools for helping them think tended to use a non-sequential approach.

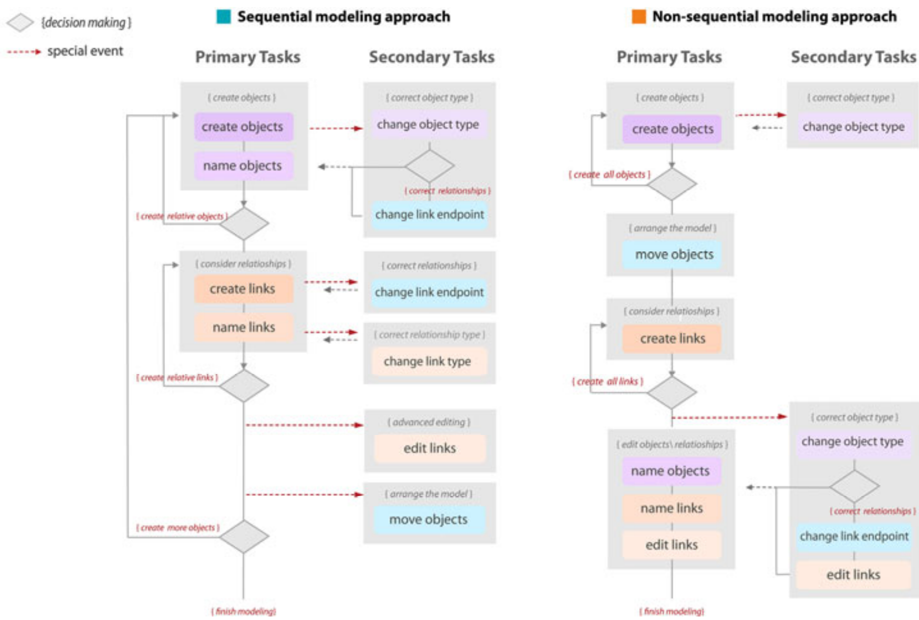


Fig. 8. Consolidated work model of modeling. In the sequential approach, the primary tasks are creating entities and links at once. In the non-sequential approach, users tend to create entities first and then consider relationships among those entities.

6 Design Implications

To design the information architecture and interface layout of a modeling tool, we argue that two different approaches and levels of user expertise should be considered. For instance, less-experienced users are drawn by strong affordance to the Object Library to create new entities rather using the hidden feature of context menus that provide the same functions. Users favoring a sequential approach will not use context menus if unaware of them. For another example, a so-called sticky tool inclines users to create all the entities of one type at once for efficiency, which might lead users favoring a sequential approach to create models by the non-sequential approach.

The interaction design for primary and secondary task is directly influenced by the modeling approaches. We propose these design implications for modeling tools:

1. Creating Entities. For both approaches, a context menu for directly creating new entities on the canvas is essential. Our tests showed that not all users are comfortable with accessing this menu through right click. We suggest two additional ways to inform users of the functionality: (i) a screen tip the pointer hovers on the blank canvas and (ii) a ghost menu while the pointer hovers over a clickable object on the canvas for a few seconds.
2. Naming Entities. For both approaches, prompting users to name entities as created can reduce work in advanced editing steps. We suggest that once an entity is created, the name text field should be highlighted and have keyboard focus to enable straightforward insertion.
3. Changing Entity Type. Changing the type of an entity is a frequent task, and grammatically incorrect links and incorrect relationships may occur as a consequence. However, the system should not automatically revise or delete them. Modeling tools should guide users and help them identify and correct errors by highlighting incorrect relationships or providing a grammar checker.
4. Creating Links. Interaction design for creating links in support of both approaches is complicated. For the sequential approach, the efficient way is to let the modeling tool automatically create a link when a user creates an entity from another. For the non-sequential approach, creating links always combines two steps: selecting the source entity and selecting the target. If a modeling tool manages to support both approaches, it could allow users to change the previous target entity into the new source entity by using a hot key, which reduces the process by one step. In addition, users argued that modeling tools should allow users to create a temporary link without selecting a target entity.
5. Annotating Links. Annotating links and naming entities should be compatible processes, such as, by highlighting the text field of a link once created. Even though annotating links is not as frequent a task in most modeling situations, it should be easily accessible without interfering with the user's primary work focus.
6. Link Direction. Editing the direction of a relationship should be straightforward but also reflect the semantics of the model. If the direction of a link is not only a part of the notation for modelers but also contains information and presents a certain relationship type with semantics, a modeling tool has to help users of all levels of expertise understand this difference.
7. Inserting Entities. Allowing users to insert an extra entity into a link between two objects is important, especially for the sequential modeling approach. The function can be included in the context menu of editing links. Advanced interactions like dragging an entity and then attaching it into a link can also be considered.
8. Hints and Guidelines. Graphic images in grammar examples help users understand the rules of the modeling language (Fig. 9). Tools should make hints and guiding windows accessible but avoid distraction. Progressive (or cascading) screen tips [54] offer simple object identification on hover, as with conventional screen tips, but more information and a link into the help system after an additional delay.

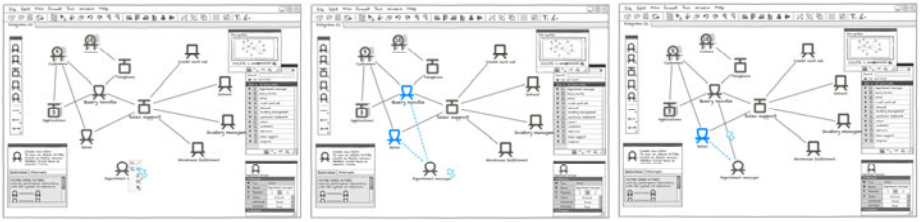


Fig. 9. Design concepts for guiding tools. To help users create links, grammar-based guiding hints (blue-dashed lines) and graphic examples (bottom left window) were proposed by users. Most agreed that graphic examples might help but that hints on the canvas interrupted work.

In modeling tools, simple operations like copy-and-paste or drag-and-drop take place in very complex situations. Since a model can be represented by more than one diagram and users may edit different models at the same time, to copy (or drag) an entity (or a group of entities) from one diagram to another might indicate two different situations. Either the user is trying to duplicate a concept for representing different instances, or the user actually wants to represent the same concept in different diagrams or to import a concept for the same instance into different models. Modeling tools need to remind users of the exact meaning and results of these interactions. An entity explorer and navigator need to clearly convey how components in diagrams and models are related. Finally, the tool should also provide users different viewpoints for merging and slicing the model.

Other important advanced functions, which should be considered for future work, are enabling users to use different graphical representations for the same instance and allowing users to merge different entities together. In our interviews, we found that these two functions are not well supported by current modeling tools, but are required when users try to clarify and reduce the complexity of models. To facilitate advanced editing, modeling tools could also allow users to group entities and components by separating them into different layers.

7 Conclusion

We present an iterative process to help improve tools for both domain-experts and software engineers define modeling languages. We consider a modeling tool as a workplace that helps users think, organize thoughts, and review concepts. We suggest several specific components should be considered in future modeling tools.

Through the contextual research described in this paper, we identified important primary and secondary modeling tasks. From this initial categorization of frequent modeling tasks, we presented and detailed the concerns and requirements of users with different levels of expertise. In addition, two modeling approaches were identified, and users' decisions to use either of these approaches were found to be based on the type of model considered as well as personal preferences. Based on this contextual analysis, we propose a different interaction design to support the primary and secondary tasks of the two modeling approaches. We argue that the interaction should support different workflows and different mental models. Design implications for modeling tools are identified that support enhanced efficiency, advanced functionality, situated usability, and varying needs of modelers.

A modeling tool serves as a communication platform among domain-experts, designers, developers, and others practitioners. Through a case study of MetaSketch Editor, we argue for better understand of the diverse purposes of using a modeling tool with different work styles. Future research will continue to evaluate remaining parts of the modeling tool, including the platform for designing concrete syntax and querying the models.

Acknowledgments. Special thanks to those who participated in this research, shared their experiences, and offered excellent opinions on the design issues.

References

1. Myers, B., Hudson, S., Pausch, R.: Past, Present and Future of User Interface Software Tools. Computer-Human Interaction 7(1), 3–28 (2000)
2. Jarzabek, S., Huang, R.: The Case for User-Centered CASE Tools. Communications of the ACM 41(8), 93–99 (2004)
3. IIvari, J.: Why are CASE Tools not Used? Communications of the ACM 39(10), 94–103 (1996)
4. Whittaker, J.A., Voas, J.M.: 50 Years of Software: Key Principles for Quality. In: IEEE IT Pro., pp, 28–35 (November / December 2002)
5. Selic, B.: The Pragmatics of Model-Driven Development. IEEE Software 20(5), 19–25 (2003)
6. Hailpern, B., Tarr, P.L.: Model-Driven Development: The Good, the Bad, and the Ugly. IBM Systems Journal 45(3), 451–462 (2006)
7. Ward, M.P.: Language-Oriented Programming. Software, Concepts and Tools 15(4), 147–161 (1994)
8. Thomas, D.A.: MDA: Revenge of the Modelers or UML Utopia? IEEE Software 21(3), 15–17 (2004)
9. Cuadrado, J.S., Molina, J.G.: Building Domain-Specific Languages for Model-Driven Development. IEEE Software 24(5), 48–55 (2007)
10. Kleppe, A.: The Field of Software Language Engineering. In: Gaševic, D., Lämmel, R., van Wyk, E. (eds.) SLE 2008. LNCS, vol. 5452, pp. 1–7. Springer, Heidelberg (2009)
11. Ledeczi, A., Maroti, M., Bakay, A., Karsai, G., Garrett, J., Thomason IV, C., Nordstrom, G., Sprinkle, J., Volgyesi, P.: The Generic Modeling Environment. In: Proc. Workshop on Intelligent Signal Processing, Budapest, Hungary (2001)
12. MetaCase (February 2011), `http://www.metacase.com`
13. Greenfield, J., Short, K., Cook, S., Kent, S.: Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. John Wiley & Sons, Chichester (2004)
14. Budinsky, F., Steinberg, D., Ellersick, R., Merks, E., Brodsky, S.A., Grose, T.J.: Eclipse Modeling Framework. Addison Wesley, Reading (2003)
15. Miller, J., Mukerji, J.: MDA Guide Version 1.0.1. OMG, Framingham (June 2003)
16. Strauss, A., Corbin, J.: Basics of Qualitative Research. Sage, Newbury Park (1990)
17. Charmaz, K.: Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Sage, Thousand Oaks (2006)
18. Eisenhardt, K.: Building Theories from Case Study Research. Academy of Management Review 14(4), 532–550 (1989)

19. Dey, I.: Qualitative Data Analysis: A User Friendly Guide for Social Scientists. Routledge, London (1993)
20. Campos, P., Nunes, N.J.: Practitioner Tools and Workstyles for User-Interface Design. IEEE Software 24(1), 73–80 (2007)
21. Löwgren, J., Stolterman, E.: Design Methodology and Design Practice. Interact, 13–20 (1999)
22. Gerdenryd, H.: How Designers Work: Making Sense of Authentic Cognitive Activities. Lund University Cognitive Studies. Lund University, Sweden (1998)
23. Schön, D.A.: The Reflective Practitioner. Basic Books, New York (1983)
24. Carlile, P.: A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. Organization Science 13, 442–455 (2002)
25. Kursat Ozenc, F., Kim, M., Zimmerman, J., Oney, S., Myers, B.: How to Support Designers in Getting Hold of the Immaterial Material of Software. In: 28th International Conference on Human Factors in Computing Systems, pp. 2513–2522. ACM press, Atlanta (2010)
26. Myers, B.A., Park, S.Y., Nakano, Y., Mueller, G., Ko, A.: How Designers Design and Program Interactive Behaviors. In: 2008 IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 177–184. IEEE press, New York (2008)
27. Norman, D.A., Draper, S. (eds.): User Centered System Design: New Perspectives on Human-Computer Interaction. Erlbaum, London (1986)
28. Kensing, F., Blomberg, J.: Participatory Design: Issues and Concerns. Computer Supported Cooperative Work 7, 167–185 (1998)
29. Nielsen, J.: Usability Engineering. Academic Press, San Diego (1993)
30. Woods, D.D.: Process-Tracing Methods for the Study of Cognition Outside of the Experimental Psychology Laboratory. In: Klein, G.A., Orasanu, J., Calderwood, R., Zsambok, C.E. (eds.) Decision Making in Action: Models and Methods, Ablex, Norwood, NJ, pp. 228–251 (1993)
31. Millen, D.: Rapid Ethnography: Time Deepening Strategies for HCI Field. In: ACM Symposium on Designing Interactive Systems, pp. 280–286. ACM Press, New York (1993)
32. Wharton, C., Bradford, J., Jeffries, R., Franzke, M.: Applying Cognitive Walkthroughs to More Complex Interfaces: Experiences, Issues, and Recommendations. In: ACM CHI 1992, pp. 381–388. ACM Press, Monterey (1992)
33. Holtzblatt, K., Jones, S.: Contextual Inquiry: A Participatory Technique for System Design. In: Schuler, D., Namioka, A. (eds.) Participatory Design: Principles and Practices, pp. 177–210. Lawrence Erlbaum, NJ (1993)
34. Mayhew, D.J.: The Usability Engineering Lifecycle: A Practioner's Handbook for User Interface Design. Academic Press, San Diego (1999)
35. Buxton, B.: Sketching User Experiences: Getting the Design Right and Right Design. Morgan Kaufmann Publishers, San Francisco (2007)
36. Wong, Y.Y.: Rough and Ready Prototypes: Lessons from Graphic Design. In: Human Factors in Computing Systems: CHI 1992, Monterey, CA, pp. 83–84 (1992)
37. Landay, J.A., Myers, B.A.: Interactive Sketching for the Early Stages of User Interface Design. In: SIGCHI Conference on Human Factors in Computing Systems, Denver, Co, pp. 43–50 (1995)
38. Bailey, B.P., Konstan, J.A., Carlis, J.V.: Supporting Multimedia Designers: Towards More Effective Design Tools. In: 8th International Conference on Multimedia Modeling, pp.267–286 (2001)

39. Gibson, J.J.: Reasons for realism: Selected essays of James J, Gibson. In: Reed, E., Jones, R. (eds.) Erlbaum, Hillsdale (1982)
40. Norman, D.A.: Psychology of Everyday Things. Basic Books, NY (1988)
41. Blomberg, J., Giacomi, J., Mosher, A., Swenton-Wall, P.: Ethnographic Field Methods and Their Relation to Design. In: Schuler, D., Namioka, A. (eds.) Participatory Design: Principles and Practices, pp. 123–156. Lawrence Erlbaum, NJ (1993)
42. Creswell, J.W.: Qualitative Inquiry and Research Design: Choosing Among Five Traditions. Sage, London (1998)
43. Lincoln, Y.S.: Emerging Criteria for Quality in Qualitative and Interpretive Research. Qualitative Inquiry 1, 275–289 (1995)
44. Lincoln, Y.S., Guba, E.G.: Naturalistic Inquiry. Sage, Beverly Hills (1985)
45. Kirwan, B., Ainsworth, L.K.: A Guide to Task Analysis. Taylor and Francis, London (1992)
46. Benyon, D.: Task Analysis and System Design: the Discipline of Data. Interacting with Computers 4(1), 246–249 (1992)
47. Ericsson, K.A., Simon, H.A.: Protocol Analysis: Revised Edition. MIT Press, Cambridge (1993)
48. Cordingley, E.S.: Knowledge Elicitation Techniques for Knowledge-Based Systems. In: Diaper, D. (ed.) Knowledge Elicitation: Principles, Techniques, and Applications, pp. 89–175. John Wiley & Sons, Chichester (1989)
49. Maulsby, D., Greenberg, S., Mander, R.: Prototyping an Intelligent Agent through Wizard of Oz. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 277–284. ACM Press, New York (1993)
50. Scott, A.A., Clayton, J.E., Gibson, E.L.: A Practical Guide to Knowledge Acquisition. Addison-Wesley Longman Publishing Co, MA (1991)
51. Constantine, L.L.: Human Activity Modeling: Toward a Pragmatic Integration. In: Seffah, A., Vanderdonckt, J., Desmarais, M.C. (eds.) Human-Centered Software Engineering: Software Engineering Models, Patterns and Architectures for HCI, pp. 27–51. Springer, London (2009)
52. Kaenampornpan, M.: A Context Model, Design Tool and Architecture for Context-Aware Systems Designs. Ph.D. thesis, University of Bath (2009)
53. Beyer, H., Holtzblatt, k.: Contextual Design: Defining Customer-Center Systems. Morgan Kaufmann Publishers, San Francisco (1998)
54. Constantine, L.L., Lockwood, L.: Instructive Interaction: Making Innovative Interfaces Self-Teaching. User Experience, pp.14–19 (Winter 2002)

Task Descriptions Using Academic Oriented Modelling Languages: A Survey of Actual Practices across the SIGCHI Community

Stanislas Couix[1,2] and Jean-Marie Burkhardt[3]

[1] EDF R&D – Management des Risques Industriels
1, avenue du Général de Gaulle, 92141 Clamart, France
[2] CNAM – Centre de Recherche sur le Travail et le Développement
41, rue Gay Lussac, 75005 Paris, France
stanislas.couix@{edf,cnam}.fr
[3] Université Paris Descartes – Laboratoire Adaptations Travail Individus
71, avenue Edouard Vaillant, 92774 Boulogne Billancourt, France
jean-marie.burkhardt@parisdescartes.fr

Abstract. There is an extensive literature on task modelling related to the design of computer systems. Task analysis and task modelling have been widely recognized as central components in human-centred approaches. The aim of this paper is to report on some results of a worldwide survey about actual practices of task descriptions languages (TDL) in SIGCHI community. Results suggest that academic TDL are not well known and not used by participants. They prefer using "home-made" TDL. This may be explained by the fact that formal TDL are not adapted to tasks analysts needs and that task modelling is an expert activity, mainly used by skilled analysts. Indeed, this study shows that task models are not only used in a productive way, *i.e.* to derive useful inputs to the design of man-machine systems. Thus, it seems that formal TDL failed to take this into account.

1 Introduction

There is an extensive literature on task modelling related to the design of computer systems. Task analysis and task modelling have been widely recognized as central components in human-centred approaches [1] to design more useful and usable interactive systems [2], as well as a main piece to develop vocational training [*e.g.*, 3, 4].

In the field of Human-Computer Interaction, task modelling (TM) can be defined as the subset of techniques employed by work analysts (ergonomists, usability specialists, etc) to analyse and to describe users' or workers' tasks in a formal way. TM relates to the phase of collecting data on work and activities during a task analysis (*e.g.* by direct observations, interviews, etc.) as well as the phase of expressing these data through one or several task models. Task models are thus the analyst's representations of the analysed task. These models are usually expressed into a graphical or a textual form. Subsequently, task description languages (TDL) can be seen as specific sets of linguistic and semantic rules used to elicit task models.

Since seminal work on Hierarchical Task Analysis (HTA) by Annett & Duncan in 1967, numerous methods and languages have been consequently proposed to support

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 555–570, 2011.

part of the modelling process (*e.g.* see the review by [2]). Surprisingly, very little literature has been devoted to the work of task analysts performing task analysis as well as specifically task modelling [5]. On the one hand, there are anecdotal reports and subsequent guidelines about the practices. For example, [6] describes the construction of task models from collected verbal report as an iterative process that involves (1) incremental definition of task items and (2) their hierarchical re-organization along the construction process as well as the collection of new data. The author also reports on some difficulties and/or specific properties of this activity, based on her experience. On the other hand, there are few empirical studies about the process of building task models with formal TDL [7, 8, 9, 10].

Finally, it is worth noting that not much is known about actual TM practices, *i.e.* what Task Description Languages are actually used, how they are used, for what intent, when during the design process, by whom, etc. Indeed, recent papers and/or workshops dedicated to task modelling (*e.g.* TAMODIA) generally focused on task description languages rather than on task modelling *per se*. These works have mainly concentrated on discussing linguistic or semantic features (proposed as desirable) for TDL, on proposing tools or articulation of approaches [11] to better support the task modelling part of the design process, on extending TDL to apply for other area like safety prevention, *e.g.* accident investigation and prevention [12], organizational development through work cultural assessment [13], cooperation and cooperative work situations [14], test case generation [15] and finally – the practical (supposed) or theoretical superiority of the specific language they are advocating. Quite no evidence has been published about the way these TDL are actually used across the task analysts' community.

This paper reports on some results of an international survey about actual practices of task descriptions using modelling languages in HCI. The purpose of the survey was to obtain a clear and situated picture of task modelling practices and task description languages actually used in the field of HCI and software ergonomics. We were interested in gathering insights about task analysis/modelling methods and task description languages produced by academics actually used across the international SIGCHI. We investigate the use of the following academic TDL : HTA (Hierarchical Task Analysis, [3]), SGT (Sub-Goal Templates, [8]), MAD (Méthode Analytique de Description, [16]), GTA (Groupware Task Analysis, [17]), CTT (ConcurTaskTrees, [18]), Diane+ [19], GOMS (Goals, Operators, Methods and Selection rules, [20]), Procope [21], TKS (Task Knowledge Structure, [22], TAG (Task-action Grammars, [23]), UAN (User Action Notation, [24]), KLM (Keystroke Level Model, [25]), TMMT (Task Method modelling Tool [26]).

The paper is organized as follows. After describing the web survey used to collect data, we will focus on the profiles on respondents, the TDL they use, and the reasons they do or do not do TM.

2 A Survey on Task Modelling Practices and Associated Task Description Languages

The questionnaire was made of 27 questions provided on a website in both French and English to have a wide access to task analysts/modellers within the international ergonomics and human-factors community.

The questions were initially organised into three main themes: (1) respondent's area, education and activity; (2) opinion, description and level of expertise related to task modelling; (3) use of task models. These questions were inspired from a pilot study [7] based on interviews with 5 professional ergonomists and on the content analysis of real task models provided by them. The initial version of the questionnaire was pre-tested with six experts in human factors and ergonomics, in terms of consistency, relevance and ease of understanding. Overall, the experts indicated that the questionnaire was relatively clear and easy to complete. A number of suggestions were made concerning the wording of several items and the overall structure of the questionnaire. These suggestions were incorporated into the revised version.

2.1 Participants' Recruitment and Survey Announcement

A call for participation was sent in January 2008 to several professional mailing lists around the world. As we were interested in the use of TDL in the academic community of SIGCHI, the call for participation was mainly sent to academic listserves. CHI-announcement[1] (international), CHI-web[2] (international), Ergoliste[3] (France), ErgoIHM[4] (France), Ergonomics[5] (United Kingdom), SIGIA-L[6] (United States), CHISIGMAIL[7] (Australia), and HCIIDC[8] (India).

2.2 Statistical Analysis

Uni- and bi-variate exploratory analyses have been performed. Bi-variate analyses were mostly based on calculating the Cramer's V2[9], a global measure of the strength of association between two variables [27], and calculating the Relative Deviations[10] (RDs) that measure local associations between specific modalities[11] of the variables [27]. We used Cramer's V and RD's as they are the equivalent of correlations for nominal variable. We did not use any reliability measure for our survey as, up to our knowledge, there is no equivalent of Cronbach's Alpha for nominal variables.

2.3 Representativeness and Limits of the Sample

Up to our knowledge, no data are currently available about the profiles of ergonomists and task/work analysts across the world. Thus, it is difficult to assess how representative

[1] http://sigchi.org/listserv
[2] http://sigchi.org/listserv
[3] https://listes.cru.fr/sympa/info/ergoliste
[4] https://listes.cru.fr/sympa/info/ergoihm
[5] https://www.jiscmail.ac.uk/lists/ergonomics.html
[6] http://www.info-arch.org/lists/sigia-l/index.php
[7] http://www.chisig.org/resources/lists.html
[8] http://tech.groups.yahoo.com/group/hdiidc
[9] V2 or Cramer's V measures the magnitude of the association between two nominal variables. Cramer's V lies between 0 and 1. It is conventionally considered to be small when $V2<0.04$, to be intermediate when $0.04<V2<0.16$, and to be strong when $V2>0.16$.
[10] RDs measures the association between two nominal variables. They are calculated on the basis of a comparison between observed and expected frequencies.
[11] There is attraction when the RD is positive and repulsion when it is negative. See [28] for details.

is our sample. A perspective is to extend this work by confronting these results to optional data originated from national ergonomics societies. Furthermore, the fact that some countries were not represented in our sample might simply mean that we failed in finding the appropriate mean to get in touch with the concerned national community during the time of the survey. It should by no mean be interpreted as an indication that no people are interested and/or doing any activity on ergonomics and in particular regarding the topics of TM.

Consequently, the following analysis will put much emphasis on qualitative interpretation of responses and emerging relationships between analysts' profiles, type of activity than on the quantified ratio of profiles and responses to be generalized to the population of ergonomic and human factors specialists.

3 Results

A total of 201 responses from all around the world were collected on the website. Respondents are professional ergonomists, interaction designers or researchers. They have various educational background (*e.g.*, psychology, engineering), are specialized in various areas (*e.g.* HCI) and have various experiences regarding task modelling. In this section, we detail all these results.

3.1 Profiles of Survey's Participants

3.1.1 Represented Countries

Europe was the most represented region (71%, n = 148), followed by Northern America (13%, n = 27), Asia (7%, n = 14) and Oceania (5%, n = 11). Africa and Southern America were not represented. See figure 1.

The French respondents were quite a bit more represented (57%, n = 115)[12] than those from other world's areas (43%). Among the latter, represented countries were by order of magnitude the US (10%, n = 21), the UK (8%, n = 16), India (6%, n = 12), Australia (5%, n = 10), Canada (3%, n = 6), the Netherlands (2%, n = 4), Germany (2%, n = 3), Italy (2%, n = 3) and Luxemburg (1%, n = 2). Finally, eight countries were represented by only one respondent: Belgium, China, Ireland, Israel, Monaco, New Zealand, Poland and Portugal.

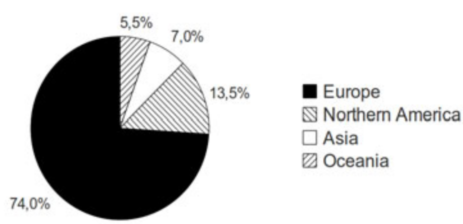


Fig. 1. Participants' geographical localisation

[12] One missing response.

3.1.2 Most Participants Have a Significant Experience in Ergonomics as Well as in Work or Task Analysis/Modelling

Globally most participants reported to be experienced in the domain of ergonomics and human factors (figure 2). They were 86% with 1 to 15 years or more of experience whereas 14% (n = 28) reported from 0 to 1 year of experience. . Participants having between 1 and 5 years of vocational experience (30%, n = 61) and between 5 and 10 years of experience (28%, n = 56) were the most represented in the sample. Participants having 15 years or more of experience were also well represented (19%, n = 38). The less represented group of participants were analysts with an experience between 10 and 15 years (8.5%, n = 17).

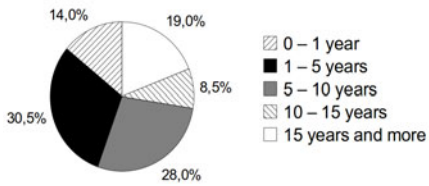


Fig. 2. Participants' number of years of experience

Among the participants, 24% (n = 49) have never done TM at the time of the survey. Actually, most of the respondents reported having a previous experience in task modelling (76%, n = 152) during their career. Among those that have an experience, most still do it in their daily activity (86%, n = 130), whereas 14% (n = 22) did not anymore (figure 3).

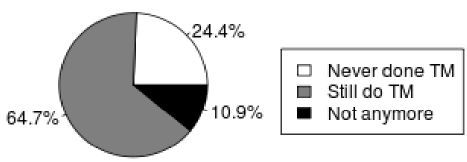


Fig. 3. Percentage of respondents having already done task modeling or not

3.1.3 Participants' Educational Background: Mainly Ergonomics, Followed by Psychology and Computer Sciences

The main proportion of respondents (59%, n = 119) had studied ergonomics (fig. 4), even if ergonomics was rarely their initial background. The remaining participants reported no formal education in ergonomics. 13% (n = 27) had a whole education in the Psychology domain, sometime with specialization close to the domain of ergonomics (*e.g.* vocational psychology); 10% (n = 22) reported having an educational background in Computer Sciences, mostly in the area of HCI or cognitive engineering; 6% (n = 13) in design or other engineering area than Computer Science, namely: aeronautics, industrial design, optometry, technical documentation. Finally, 3% (n = 6) of our

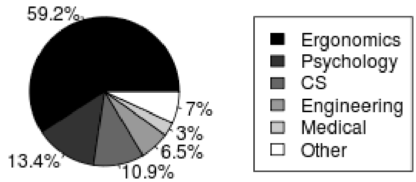


Fig. 4. Educational background of respondents. CS = computer science.

respondents were from the physiology, medical or therapy domain, and 6% (n = 13) corresponded to "other profiles" with education in sociology, health & safety, communication, business, etc.

3.1.4 Two Main Orientations in Participants' Activity: Interface Design vs. Health and Disabilities Issues

Usual domains of activity linked to task analysis and task modelling were found in our sample, although not in similar frequencies. The activities found more frequently in the sample were Interface design/specification and Interface evaluation. They were respectively performed often or most often by about 52% and about 50% of the participants. Slightly less frequent but still important, we found system's function design/specification that was sometime or often performed by 52% of respondent. On the other hand, the following activities were performed from never to rarely by more than half of the subjects: design of physical devices, health/occupational issues, system reliability, evaluation of physical devices, training program design and disability-related issues.

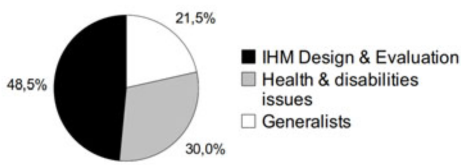


Fig. 5. Orientations in participants' activity

A hierarchical cluster analysis leads us to distinguish between three groups depending on the patterns of activity (figure 5):

- Participants specialized in interface design, specification and evaluation (49%, n = 97)
- Participants specialized in health and disabilities issues (30%, n = 60)
- Participants that have no dominant domain of activity (22%, n = 43)

3.1.5 Mostly Practitioners and a Few Researchers from Public and Private Areas

Practitioners represented about 71% (n = 142) of the participants while researchers were about 20% (n = 41). Most represented groups were threefold. 31% (n = 63) of respondents were "consultant ergonomist in a consultancy firm", 25% (n = 51) "ergonomist in a private or public company", and 15% (n = 31) "researcher in a public institution". The remaining participants were distributed across "ergonomist in freelance" (7%, n = 14), interface designer and/or analyst (7%, n = 14) and "researcher in a private or public company" (5%, n = 10) and other (7%, n = 14).

3.1.6 What Profile(s) for Participants That Reported to Do Task Modelling?

Analysis of Relative Deviations between participants profile variables and their experience in TM shows that participants currently concerned by the activity of TM are characterized as follows. They are more often researchers in a public institution (*e.g.* university, NASA), or interface designers and evaluators. They have an engineering or a psychological science background and hold a PhD. They have more than 5 years of experience in their job, and work preferentially on the design, specification or evaluation of human machine interface.

3.1.7 Participants Know Essentially the Three Following TDL: GOMS, MAD and HTA

Globally[13], three task modelling languages and methods (considering also their extensions and/or various idioms) are clearly emerging as dominating in our sample:

- GOMS -Goals, Operators, Methods and Selection rules [2]; this TDL is based on a cognitive architecture, GOMS provides task analysts with a tool to predict performance associated with the use of one particular interface. Thus, its main use is in interface evaluation.
- MAD (Méthode Analytique de Description, analytical method of description, [2]); MAD has been designed to analyse and to formalize worker's representation of their goals structure. It is mainly used to extract interface requirements.
- HTA's purpose (Hierarchical Task Analysis, [3]) is to represent how worker actually perform their tasks. It was initially intended to derive training program, but it has been adapted to derive software requirements by [8].

These TDL are respectively known by 31% (63/201), 25% (51/201) and 24% (49/201) of the respondents in the sample (table 1). Other languages are far less known: CTT, TAG, KLM, DIANE+, ETIT, TKS (see [2] for description of these TDL).

The geographical distribution differs according to the known TDL. HTA appeared to be probably the most widely known since it was present in 13 of the 18 countries represented within our sample.

GOMS is probably the second worldwide large language of description. It is quoted across 11 countries. MAD showed a specific effect since it appears to be only known in France, in Canada and in Australia in our sample.

3.1.8 Knowing a Task Description Language Does Not Mean It Is Still Used

Although widely known, the 3 methods and languages are not used as often as expected (table 1). In this view, HTA is the most TDL used since 40 respondents declare

[13] Results are based on the total number of respondents.

that they have already used it. Respectively, 31 and 26 respondents declare that they have already used MAD and GOMS. It's worth noting that, despite GOMS is the most widely known TDL, it's one of the less used TDL since only 41% of respondent that know it have actually used it.

Regarding the current use of the most known academic TDL, the number of people that still use one of the 3 widely known method is very low (table 1). Only 16 respondents still use HTA, whereas only 11 and 10 respondents still use MAD and GOMS.

Table 1. Number of responses related to knowing, having already used and still using GOMS, MAD or HTA

TDL	Known	Already used	Still used
GOMS	63	23	10
MAD	51	31	11
HTA	49	40	16

There is a weak relationship between knowing a TDL and its actual use by participants (Cramer's V2 = 0.01). Looking at Relative Deviations suggested however that some languages differ. Indeed, participants that quote GOMS tended to have actually used it less than for other languages (RD = -.20). Inversely, participants that quoted HTA still use it more than other languages (RD = +.22).

3.1.9 Academic Methods Do Not Seem Adapted to the Participants Needs

Only 23.5% (n = 38) of respondents are currently using some of the formal methods without any modification (figure 6). Most of the respondents that continue to make TM are currently using some adaptations of academic methods (39.5%, n = 64) or are using their own home-made method (37%, n = 60). Results concerning participants who stopped to perform TM are similar.

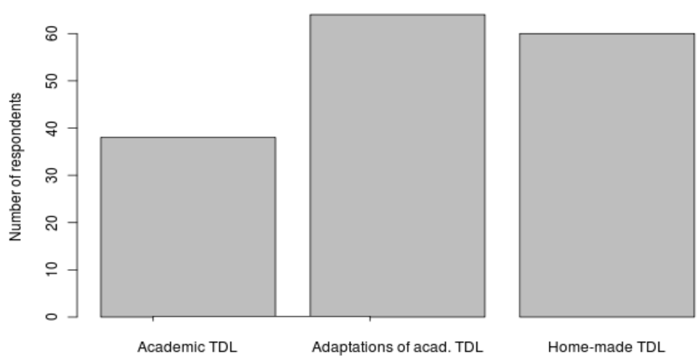


Fig. 6. Type of TDL used by respondents still performing task modeling

The main reason reported by participants is that academic methods are not actually adapted to their needs (39.5%, n = 51), either because their use is highly time consuming (13.2%, n = 17), too complex (10.8%, n = 14), or not well adapted to be shown to partners, users/operators (12.4%, n = 16) or generally speaking (3.1%, n = 4). Another reason reported is a lack of training (27%, n = 35) in these methods. It is worth noting that most of the respondents (55.87%, n = 72) didn't find any appropriate response among the suggested ones.

Participants that use an adapted method reported that their main source of inspiration was HTA followed by MAD and GOMS in terms of number of evocation. Moreover, 12 respondents reported several sources of inspirations ranging from two to three academics TDL.

3.2 What Is the Profile and Arguments of Participants That Have Never Done Task Modelling and Analysis?

3.2.1 Experience as a Crucial Factor for doing TM

In our sample, 49 participants have never done TM. These participants are over-represented among the less experienced, *i.e.* those that have less than 5 years of experience (RD = +.40). Inversely, the more experienced ones are less represented (RD = -.30). Participants that have never done TM are thus not skilled task analysts.

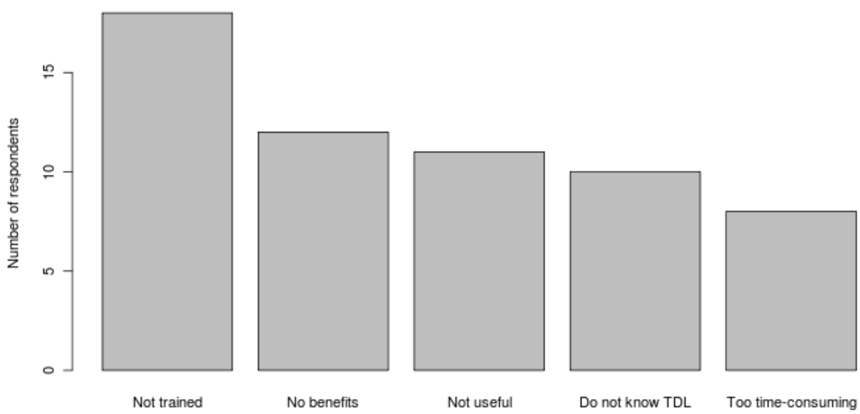


Fig. 7. Reasons for not doing task modelling

The main reason advocated by participants is that they (figure 7):

- are not trained to use them (49%, n = 18)
- see no benefits of using them in their work (24%, n = 12)
- don't think it's useful for their work (22%, n = 11)
- don't know any TDL (20%, n = 10)
- think it's a too time-consuming activity (16%, n = 8)

3.2.2 Training as a Central Determinant for Doing TM

There is an intermediate link between the fact of knowing formal methods for TM (*i.e.* HTA, MAD, GOMS) and having already performed task models (table 2), V2 = 0.05).

Table 2. Relationships between having already done TM and having reported knowing formal methods

Knowledge of formal methods	Already done TM	Never done TM	Total
Yes	88	16	104
No	64	33	97
Total	152	49	201

Participants that have never done TM tend to be highly represented among those that have no formal knowledge of formal methods (RD = +.40) whereas knowing formal methods tends to minor the fact of not having done TM (RD = -.37). These participants mention a lack of knowledge or training in TM (57%, n = 28) for not doing so. In fact, only 25% (n = 52) of respondents report to have been trained in TM. Additionally, training reported by respondents appear to vary widely in terms of duration, ranging from 2 hours to 120 hours (mean = 24h, sd = 25, median = 16h), which is quite short for training task modelling techniques.

3.3 Participants Who Reported to Have Done Task Modelling: To What Extend and for What Purpose

The advantages of doing TM were rated as follows. Globally, two assertions were rated the highest and in the same proportion (96%, n = 146, agreement vs. 4%, n = 6, disagreement). These are that TM (1) enable the modeller to better understand what operators/users are doing and (2) is a good mean of supporting the discovery of lacking information, fuzziness and uncertainties while analyzing the activity. Slightly less although still highly rated, we found the assertion that task model is a good support to communication with operators/users as well as to with other design participants (91%, n = 138, agreement vs. 9%, n = 14 disagreement). Finally, they were 77% (n = 117) vs. 33% (35/152) of participants agreeing that TM may act as a sort of external memory to store data collected in the field. Although still highly rated, the latter result shows however that such an external memory is not considered as obvious by a third of the respondent, suggesting that it could be the case only in certain conditions.

3.3.1 Starting Point for the Modelling Process

Most of the respondents begin TM before the end of the analysis on the field (78.3%, n = 119), either after the first contact with the field (38.2%, n = 58), or during systematic analysis (40.1%, n = 61). Example of “other” item is “depending on the context” or “during all the stages”.

3.3.2 Validation of the Model

Globally, most of the respondents that have already done TM reported to validate the model with operators/users (80%, n = 121). This validation was mostly done by showing the model to the operators/users and correcting it with their assistance (74%, n = 89). Validation was also carried out by comparing what the model “predicts” with what operators/users actually do (52%, n = 63) or what they do in a simulated work environment (40%, n = 49). Almost half of the respondents use only one validation method (47%, n = 57), while 35% (n = 42) use two of them and 17% (n = 20) use the three mentioned. One respondent reported not to use one of the three alternatives.

They are 47% (n = 57) to use a single method. A third of them (31%, n = 37) only interviews users or operators, 9% (n = 11) use only observation, and 8% (n = 10) use only simulation. The remaining 52% (n = 62) used between two or even three methods:

- Interviews and field observations (24%, n = 32)
- Interviews, field observations and simulation (17%, n = 20)
- Field observations and simulation (8%, n = 10)

Model's validation doesn't systematically leads to new field investigations but 38 (31%) respondents report that it's frequently the case and 57 (47%) consider that it is neither frequent nor rare. Only 2 respondents reported to never do further analysis after having carried out a validation.

3.3.3 Task Models as a Design Media to Communicate between Other Team's Members

Most respondents (81%, n = 123) hand out the task models to other members in the design team (*e.g.* during meeting, in deliverables, in requirements documents, etc.). The arguments to do this are to:

- discuss operators' or users' tasks and the appropriateness of the future tool for these tasks (74%, n = 91)
- document the design/evaluation phase of the project (64%, n = 79)
- assert design evaluation, decisions or specification choices (23%, n = 28)

4 Discussion

4.1 Formal TDL Are Not Well Known and Are Not Used

A first surprising result of this survey is that formal TDL found in the scientific literature are not known by the participants as we could have expected. Three TDL seems to be better known, namely GOMS, HTA and MAD. Other TDL (CTT, KLM, TKS, TAKD, TAG, DIANE+) are even barely known by respondents. The (lack of) promotion of TDL to students could explain this result, either because these TDL are not known by academics in charge of the training of future task analysts, or either because academics think that knowing one or two of the most popular TDL is enough for future task analysts.

Nonetheless, the fact of knowing a TDL can be questioned: indeed, it may have been interpreted by participants as "well knowing" and not just having received a short academic presentation.

A second surprising result is that even if participants are aware of one or more TDL, they don't know them well and don't use them. In our opinion, lack of training is not the only reason for this. As we have mentioned earlier, few of the people who know the academic methods have used them and fewer people still do. Lack of training or knowledge is thus not the only explanation.

4.2 Formal TDL Are Not Adapted to the Analysts' Needs

Participants also claim that the use of formal TDL is not adapted to their needs. Main difficulties quoted are that it takes to much time, it requires a lot of expertise, and it's not easy to communicate to design team members or operators with formal TDL.

Another interesting result is that although most of the respondents reported the use of only one of the academic method listed, we observed that 9 respondents mentioned the use of two or even three different methods. It suggests that features of one TDL are not completely sufficient and require some complements or "enhancements" with features borrowed from others TDL. These uncovered features are probably seen as equally important or interesting by participants. Two different TDL could provide features worth mixing when describing a single model at the same time using features from both languages and model. TDL features can also be complementary (and thus they can require articulation) in the course of the evolving phases of design.

In fact, barely no study has intended to consider how task analysts or ergonomists, in their daily activity, use task models and why, when and how they use them. This kind of study would give useful insights into the design of new TDL adapted to the needs of professional ergonomists. As already stated by [8] we need to use our user-centred design method to help ourselves design TDL fit for their users, *i.e.* our community.

Our study brings few answers to this question. But, as it's only a survey, it's not deep enough to give real inputs for the design of new methods.

4.3 Productive Use of Task Models Is Not the Only One Use

4.3.1 Task Models Are Used for a Better Understanding of Analysed Work

Respondents think that the construction of task models help them clarify the task they are analysing. Two hypotheses can be made concerning this. First, analysts need to gather several data from many different sources (*e.g.* interviews, observations, etc.). Thus, it may be difficult for them to maintain these data in working memory. Task models can be considered as an information structure to lighten mental workload associated with understanding how people perform their tasks. External representations have been recognized as one central feature in cognitive mechanism associated with problem-solving activities [30, 31, 32]. Secondly, during task analysis, ergonomists build a mental representation of users' or workers' tasks. Modelling these tasks is a way of externalizing their representations. Many researches in cognitive psychology show that externalising the representation of any situation enhances our understanding about it [33, 5, 34].

Responses suggest that task models can also be considered as an external long-term memory storing data collected during the task analysis. First, the task models allow analysts to have a retrospective view on the analysis they led and on the knowledge they acquired. Discussions in the HCI community report some similar use of task models [35, 36]. These authors advocate that task models can be used to store all the knowledge acquired on tasks in order to specify system functions that could not have been developed for the system currently designed. Second, the task model can help analysts to have a prospective view on their analysis and all the gap of knowledge that have to be filled. Construction fo the model can help analysts become aware of all the knowledge they lack about worker's or user's tasks. Thus, constructing task models can lead to a more complete task analysis.

4.3.2 Task Models Are Used to Communicate with Design Teams and Workers

Participants in the survey express loudly that they use task models to communicate the results of their analysis (the task models, not the requirements or specifications derived from their analysis). The model is used as an interface between the analysts and project members. First, the model can be shown to workers or users in order to check the validity of the model. In this case, the model is used to confront the analyst's representation of the task of the worker/user with the worker/user's own representation. Thus, the analyst can detect some mismatch between the two. Second, the model is shown to other design team members. It's used to assert some design or evaluation decision. Analysts also use the model to help other designers to understand the worker/users' tasks. Thus, their representation of it can also be closer to the reality. Nonetheless, using a task model this way implies that other designers and worker/users know how to read it. The model is thus a boundary object [37] used to encourage discussions between task analysts, worker/users and other designers.

These non-productive uses of task models, is, in our opinion, the core of why people use task models. They use them to facilitate their daily activity and to communicate with others.

4.4 Task Modelling as an Expert Activity

Generally speaking, task modelling seems to be an activity that requires high "expertise". Indeed, in our sample, only experienced ergonomists or human factors experts use TM in their daily activity.

Few studies have reported difficulties about doing TM using formal TDL [9, 10]. In most TDL, tasks are decomposed hierarchically. Authors found that decomposing a task this way is one of the biggest difficulties encountered by students in ergonomics or human factors. Our own experience in teaching TM confirms this. This may mean that students are not encouraged to do TM because of the difficulty of using formal TDL. This may explain why "home-made" TDL are the most used TDL in our study.

Nonetheless, our study suggests another explanation. As non-productive use of task models are not well documented, these are not taught. Thus only experienced task analysts have discovered them during their practice. And, as they don't know formal TDL, they use "home-made" TDL, which are perfectly tailored to their needs.

From the picture this study draws of the actual use of TDL, we aim to engage in a reflective approach for both researchers and practitioners to shorten the distance between techniques developed by researchers and their actual use by practitioners. In our opinion, shortening this distance will allow both practitioners and researchers to produce TDL well tailored for practice. Nevertheless, our study has a severe limitation. We only explore the use of descriptive task models, *i.e.* task models used to described what users or workers actually perform, either with the actual system or with the new one. However, during design, other kinds of task representations are used: prospective task models, *i.e.* task models describing what the users or workers will perform with the new device. Such TDL include UML (Unified Modelling Language,

[38]) and scenarios [39]. Maybe descriptive task models are less used than prospective task models during design. Another severe limitation of our study is the listserves used for the call for participation. We only used listserves mainly devoted to academics or to the general SIGCHI community. At the time of the study we were not aware of listserves primarily devoted to practitioners.

5 Some Implications of Our Study and Recommendations towards the HCI Community

Despite having no scientific evidence on this, we strongly believe that task modelling enhances the practice of ergonomics in design. Two reasons can be given. First, task modelling improves the understanding of task and the completeness of task descriptions. Task modelling is a more systematic way of describing tasks. Second, task modelling is an efficient way of discussing and making other designers understand the tasks of users and operators.

Based on the results of this study, a previous research [7] and our personal experience regarding task modelling, here are some recommendations to improve adoption and appropriation of task modelling.

First of all, TDL need to be more visual. Task models are used to communicate the results of task analysis. However, in most academic TDL, only tasks, their structure and their sequence are visually represented [7]. Crucial information needed to design or evaluate computer systems like triggering conditions of tasks, information needed by users or operators are hidden. This reduces the utility of academic TDL. Home-made TDL collected in our previous study [7] exhibit information needed by ergonomists to design or evaluate systems.

Most dedicated software applications (*e.g.* Euterpe for GTA) are not tailored to the actual use of task models, they lack some basic functions. Even if this is related to the drawbacks of academic TDL mentioned previously, these software applications need the support of communicating functions like printing and exporting task models.

Nonetheless, enhancement of academic TDL and task modelling software is not sufficient. We need to train students to the use of task modelling. However this training must not only focus on productive use of task models, it should increase student awareness towards their non-productive use discussed previously in the paper.

Acknowledgement. The authors would like to thank all participants. This research has been partially funded by the French National Funding Agency (ANR) in the context of PERF RV2 and V3S projects.

References

[1] Maguire, M.: Methods to support human-centred design. International Journal of Human-Computer Studies 55, 587–634 (2001)

[2] Diaper, D., Stanton, N.: The Handbook of Task Analysis for Human-Computer Interaction. Laurence Erlbaum Associates, Mahwah (2004)

[3] Annett, J., Duncan, K.D.: Task analysis and training design. Occupational Psychology 41, 211–227 (1967)
[4] Patrick, J.: Training: research and Practice. Academic Press, London (1992)
[5] Cox, R.: Representation construction, externalised cognition and individual differences. Learning and Instruction 9(4), 343–363 (1999)
[6] Shepherd, A.: An approach to information requirements specification for process control tasks. Ergonomics 36, 1425–1439 (1993)
[7] Couix, S.: Usages et construction des modèles de tâches dans la pratique de l'ergonomie: une étude exploratoire. Master's thesis, Université Paris Descartes (2007)
[8] Ormerod, T.C., Richardson, J., Shepherd, A.: Enhancing the usability of a task analysis method: a notation and environment for requirements specification. Ergonomics 41(11), 1642–1663 (1998)
[9] Ozkan, N., Paris, C., Balbo, S.: Understanding a task model: an experiment. In: Johnson, H., Nigay, L., Roast, C. (eds.) Proceedings of HCI on People and Computers XIII. Springer, Heidelberg (1998)
[10] Patrick, J., Gregov, A., Halliday, P.: Analysing and training task analysis. Instructional science 28, 51–57 (2000)
[11] Ait-Ameur, Y., Baron, M.: Formal and experimental validation approaches in hci systems design based on a shared event b model. International Journal on Software Tools for Technology Transfer 8, 547–563 (2006)
[12] Basnyat, S.: Erreur humaine, modèles de tâches et description formelle pour la conception et l'évaluation de systèmes critiques et tolérants aux erreurs. In: Actes de la seconde rencontre des jeunes chercheurs en interaction homme-machine (2004)
[13] Oedewald, P., Reiman, T.: Core task modelling in cultural assessment: a case study in nuclear power plant maintenance. Cognition, Technology & Work 5(4), 283–293 (2003)
[14] Killich, S., Luczak, H., Schlick, C., Weissenbach, M., Wiedenmaier, S., Ziegler, J.: Task modelling for cooperative work. Behaviour & Information Technology 18(5), 325–338 (1999)
[15] Barbosa, A., Paiva, A.C.R., Creissac Campos, J.: Test case generation from mutated task models. Presented at the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Pisa, Italy (June 13-16, 2011) (to appear)
[16] Scapin, D.L., Bastien, J.M.C.: Analyse des tâches et aide ergonomique à la conception: l'approche MAD*. In: Kolski, C. (ed.) Analyse et conception de l'IHM. Interaction homme-machine pour les SI, Hermès, Paris, vol. 1, pp. 85–116 (2001)
[17] Van der Veer, G., Lenting, B., Bergevoet, B.: GTA: Groupware task analysis - modeling complexity. Acta Psychologica 91, 297–322 (1996)
[18] Paternò, F.: ConcurTaskTrees: an engineered notation for task models. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction, pp. 483–501. Laurence Erlbaum Associates, Mahwah (2004)
[19] Barthet, M.F., Tarby, J.C.: The DIANE+ method. In: Vanderdonckt, J. (ed.) Computer-aided design of user interfaces, pp. 95–120. Presses Universitaires de Namur, Namur (1996)
[20] Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. Erlbaum, Mahwah (1983)
[21] Tijus, C.A., Poitrenaud, S., Richard, J.F.: Propriétés, objets, procédures: Les réseaux sémantiques d'action appliqués à la représentation des dispositifs techniques. Le Travail Humain 59(3), 209–229 (1996)

[22] Johnson, P., Johnson, H., Waddington, R., Shouls, A.: Task-related knowledge structures: Analysis, modelling and application. In: People and Computers: From Research to Implementations, pp. 35–62. Cambridge University Press, Cambridge (1988)
[23] Payne, S., Green, T.: Task-action grammars: a model of the mental representation of task languages. Human Computer Interaction 2(2) (1986)
[24] Siochi, A.C., Hartson, H.R.: Task-oriented representation of asynchronous user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Wings for the Mind, pp. 183–188 (1989)
[25] Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Communications of the ACM 23(7), 396–410 (1980)
[26] Camilleri, G., Soubie, J.L., Zalaket, J.: TMMT: Tool supporting knowledge modelling. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. LNCS, vol. 2773, pp. 45–52. Springer, Heidelberg (2003)
[27] Cramér, H.: Mathematical Methods of Statistics. Princeton University Press, Princeton (1999)
[28] Bernard, J.M.: Analysis of local dependencies in contingency tables using the imprecise dirichlet model. In: Bernard, J.M., Seidenfeld, T., Zaffalon, M. (eds.) Proceedings of the 3rd International Symposium on Imprecise Probabilities and their Applications. Carleton Scientific (2003)
[29] Sebillotte, S.: Décrire des tâches selon les objectifs des opérateurs. de l'interview à la formalisation. Le Travail humain 54(3), 193–223 (1991)
[30] Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cognitive Science 11(1), 65–99 (1987)
[31] Zhang, J.: Representations in distributed cognitive tasks. Cognitive Science 18, 87–122 (1994)
[32] Zhang, J.: The nature of external representations in problem solving. Cognitive Science 21(2), 179–217 (1997)
[33] Chi, M.T.H., De Leeuw, N., Chiu, M.H., Lavancher, C.: Eliciting self-explanations improves understanding. Cognitive Science 18, 437–477 (1994)
[34] Reisberg, D.: The detachment gain: the advantage of thinking out loud. In: Landau, B., Sabini, J., Jonides, J., Newport, E.L. (eds.) Perception, Cognition & Language, pp. 139–156. MIT Press, Cambridge (2000)
[35] Montabert, C.: McCrickard: Reuse-centric requirements analysis with task models, scenarios and critical parameters. Journal of Systemics, Cybernetics and Informatics 5(1), 72–78 (2007)
[36] Sutcliffe, A.: Symbiosis and synergy? Scenarios, task analysis and reuse of HCI knowledge. Interacting with Computers 15, 245–263 (2003)
[37] Jeantet, A.: Les objets intermédiaires dans la conception. Éléments pour une sociologie des processus de conception. Sociologie du travail 40(3), 291–316 (1998)
[38] Booch, G., Rumbaugh, J., Jacobson, I.: The Unified Modeling Language User Guide. Addison-Wesley Object Technology Series. Addison-Wesley, Reading (1998)
[39] Carroll, J.M.: Making use: scenario-based design of human-computer interactions. Massachusetts Institute of Technology (2000)

Selective Modeling to Support Task Migratability of Interactive Artifacts

Anke Dittmar and Peter Forbrig

University of Rostock, Department of Computer Science,
Albert-Einstein-Str. 21
18051 Rostock, Germany
{anke.dittmar,peter.forbrig}@uni-rostock.de

Abstract. Selective modeling is suggested as a technique that encourages designers to mix exploratory, analytical, and empirical design activities in interaction design. The co-development of models and prototypes of interactive systems is proposed to support a better balance between formal and explorative design approaches. Models serve to inform design decisions but also to analyze emerging alternatives of prototypical implementations.

Task migratability is a usability design principle that describes how control for task execution is transferred between system and user. Refined flexible task allocation is rarely achievable through pure top-down decomposition as used in many model-based approaches. The paper shows at the example of HOPS models how selective modeling can be applied to develop prototypes in a deliberated evolutionary way by using models to express different viewpoints and to explore design options at different levels of granularity.

Keywords: User-Centered Design, Model-Based Design of Interactive Systems, Exploratory Design, Tools for Design, Modeling, Prototyping, Design Rationale.

1 Introduction

Task migratability describes the ability of an interactive application to pass control for the execution of a task so that it becomes either internalized by the user or the application or shared between them [1]. The application of this usability principle to the design of interactive systems promotes the implementation of dynamic function allocation. Safety-critical systems such as the control of aircrafts, ships or chemical plants are complex and well-accepted examples for a need of dynamic function allocation. Throughout this paper a simple and mundane example is used for illustration. It allows for discussion without requiring specific domain knowledge.

In order to shape an interactive artifact that supports task migration, designers[1] must develop a clear understanding about current tasks, artifacts, habits, and situations. They must create hypotheses about appropriate grades of automation in possible future situations and how to support them by corresponding user interfaces.

[1] In this paper, the term designer refers to all active stakeholders in the design process.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 571–588, 2011.

They also need to understand the nature of transition processes: why and how do people change their current practices? One can consider task migratability as a principle that encourages designers to develop a differentiated picture of human activity and the role of artifacts. Of course, this has to be reflected in the design processes themselves. Task migratability is neither well supported by analytical approaches favoring a top-down refinement of models to implementations nor by approaches focusing mainly on the application of empirical means.

The paper suggests *selective modeling* as a technique to intertwine exploratory, analytical, and empirical design activities in a more effective way. This technique can be considered as a deliberate change to assumptions and practices in model-based design to support the integration with other user-centered design approaches.

In selective modeling, designers co-develop prototypical implementations of the interactive artifact and models describing different perspectives on it. For example, envisaged task models help to derive requirements of the interactive system under consideration. Or, exploratory prototypes are built and then empirically tested or analyzed in more depth by the creation of models. In the suggested approach, models and implementations are considered as fragmentary and "open" design representations which are used to shape interactive artifacts. Each representation evokes certain thoughts and actions. Their co-development helps designers to switch between different perspectives and to explore design alternatives in different contexts. It is acknowledged that design activities in a user-centered design process can happen in any order but must be linked by evaluation steps (e.g. [2]).

In particular, the paper investigates how the co-development of models and prototypes supports flexible task allocation. Models and prototypes have specific characteristics and serve different purposes in the suggested approach.

- Models describe selected aspects of current and envisaged tasks, artifacts and roles.
- Models inform the development of prototypes and other models.
- Models serve to analyze prototypes.
- Prototypes are implementations of design ideas and can be tested for utility, usability and user experience.
- Prototypes (and their use) inform the development of models to get a richer task understanding.
- Prototypes and models are deliberately underspecified in their behavior. They are supplemented by other design representations to explore alternative design options.

The general approach is "instantiated" in the paper by the use of Higher-Order Processes Specifications (HOPS), a universal specification formalism with high-level concepts for describing interaction from different viewpoints [3,4]. Tool support makes it possible to animate HOPS models and to map them to Java implementations. Selective modeling is supported by techniques such as model coupling, model-guided prototyping and deliberated underspecification. Throughout the paper the mastermind game serves as example to illustrate suggested ideas.

The paper is structured as follows. In the next section, task migratability is considered in depth and examples are given. Sect. 3 discusses different design approaches in HCI and their underlying understandings of tasks and human activity in general. Selective modeling is introduced in Sect. 4. It applies a more liberal task understanding on design activities themselves and suggests adaptations to current model-based design practices because flexible task allocation in systems is rarely achievable by pure top-down refinements of task models to implementations. Sect. 5 instantiates the general approach. The HOPS formalism and tool support is applied to the example problem to demonstrate a co-evolution of models and prototypes in order to explore task control options at different levels of granularity. The paper closes with a summary and future work.

2 Task Migratability

Although almost a truism, proper task allocation is still one of the most important aspects of human-computer interaction. Task migratability is a usability principle that supports dynamic task allocation. It “concerns the transfer of control for execution of task between system and user. It should be possible for the user or system to pass control of a task over to the other or promote the task from a completely internalized one to a shared and cooperative venture” [1]. Systems that allow dynamic allocation of tasks make possible more adequate responses to actual conditions. On the one hand, task migratability is based on *task conformance* describing the degree to which a system covers the users’ tasks and represents their task understanding [1]. On the other hand, designers need to acquire a deeper understanding of what the tasks of the users might be in certain situations and how to support their achievement.

Position	Specification	Amount	Unit	Unit price	Total price
01.02.0008	crushed stone	50	m^3	24,31	1215,50
01.02.0009	...				
...					

Bid:

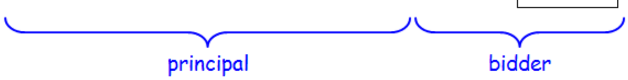


Fig. 1. A tender specifications created by the principal and priced by a bidder

User interfaces of applications that offer different task allocation options (configurations) are typically more complex. This not only concerns the representation of the functionality of each configuration and the representation of relevant information for the users to achieve their actual goals. It also concerns the representation of the actual and of possible configurations and the design of control transfer.

Sometimes a poor representation of a certain configuration may easily cause problems. Sometimes, subtle modifications to the user interface may already be sufficient to support different configurations. For example, an interactive application for managing tender specifications as depicted in Fig. 1 may support different modes for calculating bids. In one mode the application calculates the sum of the total prices of each position in the specifications automatically. In another mode the sum is calculated as well but can be modified by the user. In the first mode the bid entry may be implemented as output field, in the second as input/output field.

2.1 Current and Envisaged Worlds

"Design requires two models of the world: a current one and a future one. Design is a goal-directed activity involving deliberated changes intended to improve the current world, so the need to model the future in design is unquestionable. In practice, models of possible future worlds need to be based on models of the current world" [5].

No matter how simple design problems may appear at first glance, a deeper analysis of current practices may reveal more subtle issues. To illustrate this point, let us go back to the example of tender specifications as they are e.g. used in the building sector (see Fig. 1). In order to find a builder, a principal creates the specifications for the project. Bidders fill in unit prices, total prices and the bid. Then, the principal organizes a submission where the bids are made public to all bidders. Afterwards he checks the details of each bid for calculation errors and makes a decision. Let us assume that a computer artifact has to replace paper work. It seems to be obvious that the calculation of the total prices and the bid should be done by a computer. However, it was not uncommon that bidders miscalculate with intention in order to cover their real bids in the public submission. Against this background the design problem appears in a different light. Even if not all will be implemented, different task allocation options should be considered at least. A design process which is guided by the principle of task migratability increases the sensitivity of the designers to possible effects of allocation decisions at different levels of granularity on the resulting interactive artifact (in particular, on its user interface) and on the behavior of the users.

Dearden et al. [6] point out that allocation is not a 'zero-sum game'. It rather should be considered as a transformation of work that the human must perform and that does not necessarily result in a reduction of workload. Unintended, in some cases even undesired practices can emerge. An example is given in [7] where a captain of a modern airliner is cited with the words "You never know exactly what will be the result of flipping a switch or setting in a new parameter, so we don't interact with [the automation] during automatic landing. We simply don't know what it will do."

Fortunately, it is not possible to fully predict the future and sometimes concerns about design effects are overdone. For example, the introduction of software systems to manage tender specifications may result in a decline of intended miscalculations by bidders regardless whether these programs allow a manipulation of bids or not. However, this does not release designers from their responsibility for future practices.

It can be regarded as common ground in user-centered design that the development of interactive artifacts has to be embedded in processes of understanding current working practices and envisaging/designing future practices and that such processes are supported by the creation and use of external representations. However, Sect. 3 will show that design approaches can differ in their underlying assumptions about tasks and human activity, about models and how to use them, about modeling techniques and their combination with other design techniques.

2.2 The Example Problem

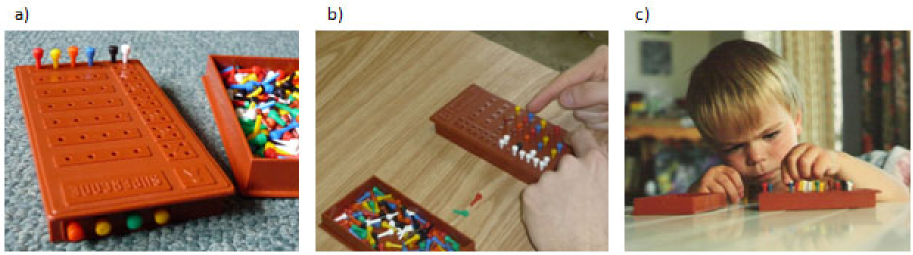


Fig. 2. a) Mastermind as board game with the hidden code in front of the picture, b) codebreaker making a guess, c) unintended use of the board

Mastermind is a logic game for two players in the roles of the codemaker and the codebreaker. At the beginning the codemaker sets a code (a hidden combination of colored pegs) which the codebreaker has to discover. Each guess of the codebreaker is marked by the codemaker. Usually, a black marking peg is used to indicate that a peg has the right color and the right position, a white peg tells the codebreaker that one peg is of a color which is in the solution, but it is in the wrong position. The game board in Fig. 2 gives the codebreaker a maximum of seven trials to solve the puzzle.

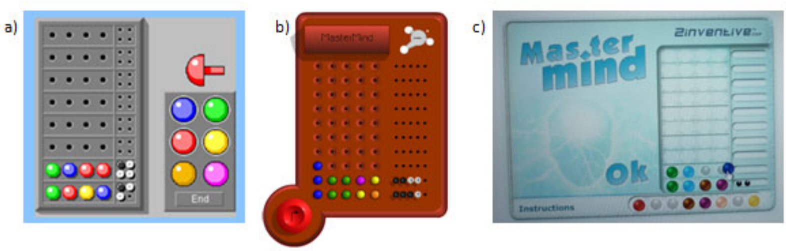


Fig. 3. Online versions of the game[2]

Mastermind is a simple but still popular game. There are a number of online versions available (see Fig. 3). In most of them, the computer acts as codemaker and the user as codebreaker. The example problem we will use throughout the paper is about how to support flexible task allocation in an interactive version of the game. It is used to play with the idea of selective modeling as a technique to combine different means such as fragmentary models and prototypical implementations in a deliberate way in order to get a refined and situated understanding about work and to shape interactive artifacts accordingly. It is also used in the next section to illustrate common ideas in HCI which form the basis for the suggested approach.

---

[2] `http://rezepte.nit.at/online_spiele/mastermind/mastermind.html`
`http://www.mathsisfun.com/games/mastermind-game.html`
`http://www.gamesbasis.com/mastermind.html` (last accessed January 2011)

3 Roots of Selective Modeling

Typical descriptions about the “history” of HCI describe the 1980s and early 1990s as a formative period with emphasis on cognitive aspects of single user’s tasks. Since the 1990s, the focus has changed to supporting cooperation and communication between users (e.g. [8]). This shift has been accompanied by changes in the understanding of goal-directed human activity. In the following, task-based and resource-based conceptions of human-computer interaction are reviewed.

3.1 Task-Based Conceptions

In task-based approaches to HCI, tasks are the unit of analysis and design to describe work in current and envisaged work systems. A task is seen as a mechanism by which intended changes in application domains are achieved. It is assumed that people possess and recruit knowledge about tasks they have previously learned and performed in a given domain. It is furthermore assumed that task knowledge can be analyzed, modeled and predicted [9]. Cognitive approaches such as HTA [10], GOMS [11] or TKS [9] describe task knowledge in terms of hierarchical task decomposition, goals, actions, plans, task domain objects etc. Task analysis refers to the process of identifying and examining the tasks that must be performed by users when they interact with systems. It is used in HCI to provide an idealized, normative model of tasks that should be supported by computer artifacts. In task design, models are used to develop hypotheses of future tasks and their execution.

To illustrate common ways to notate and use task models, Fig. 4 shows a simple analysis model of making a guess in mastermind. The task is hierarchically decomposed and temporal relations are added to describe in which order sub-tasks have to be executed. Let us assume that the model is the result of observing some games that were played with the board of Fig. 1. The codebreakers placed pegs on the board but sometimes removed them again. The codemakers were observed to set marking pegs only. The example may be sufficient to point out that analysis models can only describe small parts of behavior [5]. The presented view depends on who produced the model and for what purposes. For example, the model in Fig. 4 could have been produced to inform the design of an interactive version and the designers may have found it sufficient to look at observable behavior. The gray areas in the figure indicate possible design decisions. First, action sequences <Remove peg, Set peg> as observed in the current task situation are replaced by the operation ‘Replace peg’ (as e.g. in the version of Fig. 3b). Second, the computer takes the role of the codemaker. However, the task model in Fig. 4 could also be useful for evaluation. If it is used to reflect available user actions in the game version of Fig. 3a) it can reveal that users are not supported in removing pegs from the board.

3.2 Resource-Based Conceptions

Above mentioned approaches put much emphasis on mental representations that people develop and employ to achieve desired states in the world. However, the world should not be seen as “a largely stable collection of objects and events to be observed and manipulated according to the internal mental states of the individual” [12].

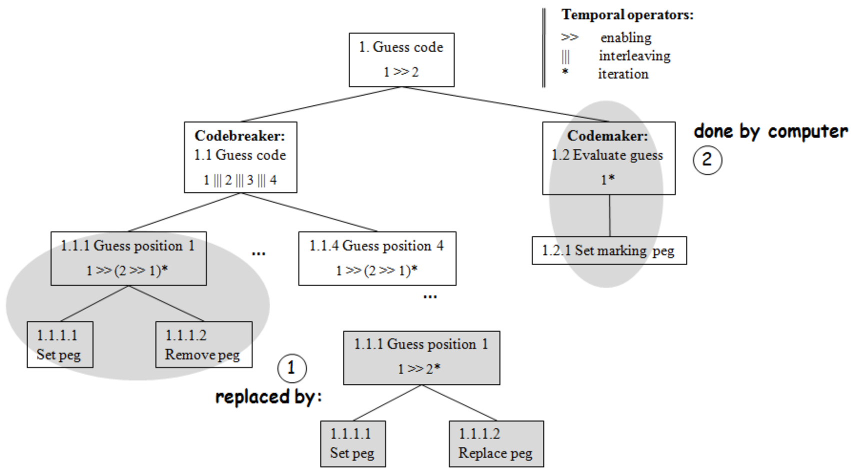


Fig. 4. A hierarchical task model describing observed behavior for a guess in mastermind. Gray areas indicate the transformation to a design model.

Resource-based conceptions of human activity such as 'situated actions' [13], Distributed Cognition [14], or Activity Theory [15] give far more attention to the interplay between actors and artifacts. It is assumed that the resources which are available in an actual situation shape the actors' behaviors. People experience their environment as *activity spaces* and change them deliberately according to their goals. The deeper their understanding of the environment the more fine-tuned are their interventions in order to trigger certain behaviors. As a consequence, plans (e.g. task models) are seen as one resource among others. They can guide but not fully control behavior.

Analysis and design activities that are guided by this attitude acknowledge the deep interplay between internal and external artifacts and the variety of 'control flows'. Of course, this also supports a deeper reflection of task migratability of interactive systems. To give an example, let us look back at Fig. 2. In picture b), the codebreaker uses his fingers to mark some pegs in order to make the next guess. Although the mastermind game was introduced as a logical game in Sect. 2.2 (and so it is perceived by many people indeed[3]), players use besides cognitive skills a variety of external means in order to succeed. They position pegs with certain colors near the board. Sometimes two or more people are the codebreaker and they exchange their thoughts about the right code. When the codebreaker is a novice to the game, the codemaker may help and so on. An interactive mastermind game could enable a single person to play or could allow people in different locations to play with each other. However, designers should also think about the fine-grained details in order to support different usage situations? In [16] human-computer interaction is analyzed as distributed cognition and six different types of information structures are identified that interactive artifacts can represent to users as possible resources for action: plans, goals, possibilities, history, action-effect relations, and states.

[3] On the webpage of the version depicted in Fig. 3b): "Try different color combinations and use your brain to figure it out".

Resource-based approaches give more plausible explanations of effects such as appropriation and unintended use of artifacts because they take a more liberal view on goal-directed behavior than task-based conceptions. The latter term refers to situations where people use artifacts in ways which were not intended by the designers. An example is given in Fig. 2c). The two years old child has no idea about the goal of the game but tries to place pegs on the holes on the board at all. It is more a matter of dexterity. What one can (re-)learn from the child is that there is no need not to use holes in 'free rows' on the board. The analysis of phenomena such unintended use can help to broaden the view on current activity spaces and prevent designers from introducing unnecessary constraints on future activity spaces. As it turned out in the analysis of the mastermind game there are in fact people who use free rows as a resource for holding pegs which they believe are part of the solution. Many available interactive versions such as those in Fig. 3 do not give the users (codebreakers) access to free rows. A more fine-grained transfer of control would be more convenient. Allocation of function is considered as a resource allocation problem in [6].

3.3 Related User-Centered Design Approaches

In User-Centered Design (UCD), many approaches follow a more task-based conception of work. For example, formal task models and top-down refinement are used in most model-based design approaches to derive user interface specifications and prototypes by applying techniques from software engineering [17, 18]. In addition, there is often no distinction between task analysis models and task design models. Consequently, modeling activities easily become specification-driven as we have shown in [3]. In contrast, Scenario-Based Design [19] or Contextual Design [20] encourage designers to ground their design in discussions of different, even conflicting perspectives on current working practices. These approaches rather support resource-based ideas.

Design approaches can furthermore differ in their acceptance and use of formal/semi-formal/informal descriptions and in their assumptions about completeness and correctness of models. They favor analytical over empirical techniques or vice versa. They support design rationale or not, and generally, put different emphasis on different types of design activities. For example, in iterative design empirical techniques such as prototyping, usability tests and iterative development in the field are preferred because it is assumed that the effects of most design choices and emerging artifacts cannot be predicted well enough by analytical means [21,22].

HCI is described as multidisciplinary but also fragmented field [8]. Of course, this fragmentation must have effects on researchers and practitioners. On the one hand, isolation "from some portion of the field's foundations" is often to be found among them as a coping strategy [8]. On the other hand, the need for amalgamating different UCD methods and techniques is well-accepted. To give just two examples: in a warning about iterative design in [1] it is written that "the ideal model of iterative design, in which a rapid prototype is designed, evaluated and modified until the best possible design is achieved…is appealing" but it is also important to be able to overcome bad initial design decisions or to understand the reasons behind usability problems and not just detect the symptoms. It is recommended to use iterative design "in conjunction with other, more principled approaches to interactive system design."

Fischer et al. [23] point out that in Design Rationale "argumentation has been considered in isolation from the activity of solution construction" but that construction and argumentation have to be unified.

Although UCD practices have been refined over the years (e.g. [24]) the question of how to feed back results of evaluation steps into design in an effective way still remains. In this paper, we explore how deliberate changes to assumptions and practices in model-based design support a move towards a resourced-based conception of human-computer interaction and the integration with other empirical and analytic design practices in order to develop more flexible interactive systems.

4 Selective Modeling – Main Ideas

We suggest the term *selective modeling* to describe design practices where designers use models as *resources* for developing interactive artifacts. In other words, we see the design process itself through the glasses of a resourced-based conception of work. Designers deliberately shape their view on the design space by creating different types of representations about tasks, actors or artifacts. They help to develop an understanding about current practices and about possible future usage situations of the system under design that emerges as a "side effect" of the co-evolution of models.

If we apply this view on model-based design the overly dominant role of (envisaged) task models on the creation of systems specifications and prototypes is challenged. Additionally, the understanding of models as selective and fragmentary descriptions of phenomena from certain points of view is more emphasized. Task models describe how to act in the world to achieve a certain goal. Artifact models describe domain objects in terms of attributes and actions serving different purposes in different contexts. Dialog models, in particular, may describe possible usage scenarios of interactive artifacts in terms of visible information and enabled action sequences. A prototype represents concrete human-computer interactions. It is represented by code in a programming language and is experienced in a different way than abstract models. Every representation can be considered as a design artifact which evokes certain responses.

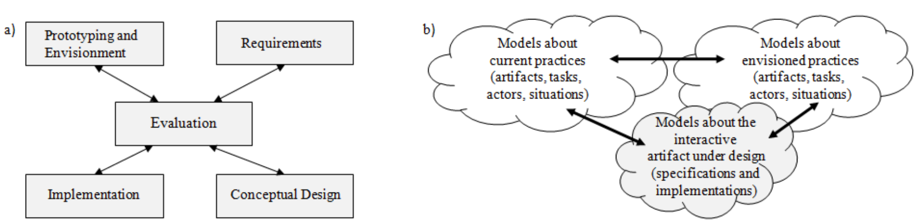


Fig. 5. a) Modified Star Life Cycle [2], b) co-development of design representations

While diagrams such as the one depicted in Fig. 5a) illustrate the interleaving of design activities and the central role of empirical and analytic evaluation steps in UCD, Fig. 5b) focuses on the co-evolution of design artifacts that must be achieved.

The selection and use of models by the designers depend on the objectives and the constraints of the actual design situation. For example, models of tasks, artifacts and roles can be coupled with dialog models in order to explore resource allocation options and to refine the dialog models. A formal systems specification can be evaluated by a corresponding prototype. Models can be used or developed for the assessment of emerging implementations. Changes to these models then can feed back into later design steps and so on. Selective modeling supports the intertwinement of different design activities by encouraging designers to use representations in a focused way.

- Design representations (e.g. formal models and prototypes) describe selected aspects of the design problem from different perspectives and at different levels of granularity.
- Representations are deliberately underspecified in order to avoid premature design commitments, e.g. a premature allocation of resources.
- Different representations are partially coupled to explore design alternatives, to inform design decisions and to enable testing and reflection of implementations. Such couplings drive the co-evolution of representations.

The next section explores these ideas further by using the HOPS formalism.

5 Selective Modeling with HOPS

HOPS (Higher-Order Processes Specification) is a universal specification formalism with tool support for describing interaction from different viewpoints and at different levels of abstraction. A short introduction to HOPS is given in Sect. 5.1. For more details see [3,4,25]. Sect. 5.2 explains how HOPS supports selective modeling techniques. In particular, model coupling, model-guided prototyping and deliberated underspecification are demonstrated at the example of designing an interactive mastermind game.

5.1 Introduction to HOPS

In HOPS, systems are considered as compositions of interacting sub-systems. They are specified by *processes*. The structure of a process is determined by components, operations and sub-processes. *Operations* refer to the smallest units of behavior that are of interest in the actual modeling context. The behavior of a process is defined by a set of sequences of operations. *Sub-processes* of a process P refer to partial behaviors of P. They are useful for creating structures (e.g. task hierarchies) or behavioral variants. Sub-processes can also be used to specify states of components. *Components* as instances of previously defined processes describe sub-systems. Processes without components are called *basic processes*. Their operations are defined by names only. Processes with components are called *higher-order processes*. An operation of a higher-order process describes either new emerging behavior or it is an abstraction of a sequence of operations of components. By using higher-level operations and sub-processes, a process can partially control the otherwise independent behavior of its components.

Basic Processes and Process Animation. The basic process *Guess_peg* given below is defined by three operations (lines 3-5) and three sub-processes *Option1-3*. Each sub-process describes a variant for placing a peg in the mastermind game. The first two options correspond to the variants in Fig.4.

```
1   PROCESS Guess_peg
2   OPS
3     setPeg(s: string),
4     removePeg,
5     replacePeg(s: string),
6   SUB PROCESSES
7     Option1 IS setPeg(?) ; (removePeg ; setPeg(?))*,
8     Option2 IS setPeg(?) ; replacePeg(?)*,
9     Option3 IS setPeg(?); ((removePeg ; setPeg(?)) [] replacePeg(?))*,
10  END PROCESS
```

The behavior of (sub-)processes is specified by partial equations. Temporal operators - as known e.g. from CCT [17] - are used to describe constraints on operation sequences (lines 7-9).

Temporal operators in HOPS:	;	sequence	#n	n-fold iteration
	[]	alternative	[…]	option
	|||	concurrency	|>	disabling
	*/+	iteration	[>	interruption

The HOPS tool offers interactive model animation. At each step of an animation run, the user can choose to execute one of the enabled HOPS operations (presented by buttons in Fig. 6 that are attached to the actual state/node in the animation tree). Each path of the animation tree represents the prefix of a valid operation sequence of the model. The user can "jump" between nodes of the tree to further explore specified behavior.

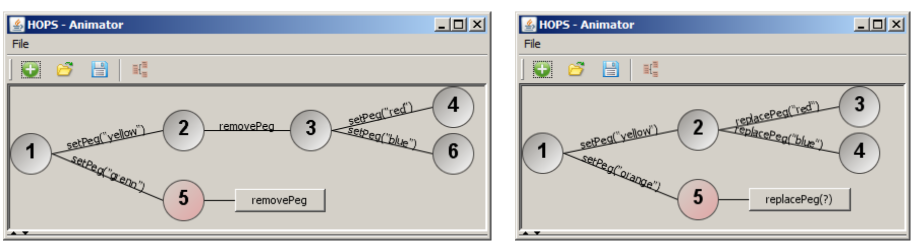


Fig. 6. Interactive animation trees for sub-processes *Option1* and *Option2* of *Guess_peg*

Mapping between HOPS Models and Object-Oriented Implementations. The HOPS tool enables an automated mapping of HOPS process instances to Java objects. If an operation is executed during model animation the corresponding Java method calls are executed as well (see Fig. 7). In the example given below, instances of process *Peg_hole* are linked to Java objects of class *PegHole*, HOPS operation *init* is mapped to the constructor method, operation *setState* to method *setState* and so on. The process describes peg holes on a mastermind board. It has a local variable (line 2) to hold the state: holes are empty or contain a peg of a specific color.

```
PROCESS Peg_hole
VAR state: string,
OPS
  init IS objId.fCall(new,[this],["PegHole","empty"]),
  setState(s: string) IS void.fCall(setState,[this],[s]),
  getState IS state.fCall(getState,[this],[ ]),
  quit,
SUB PROCESSES
  Peg_hole IS init ; (setState(?) [] getState)* [> quit,
END PROCESS
```

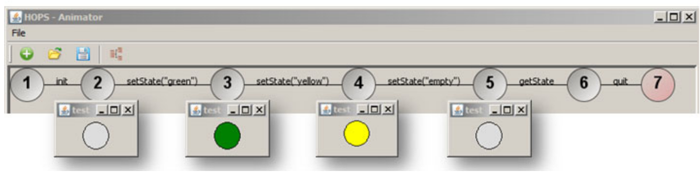


Fig. 7. The animation of operation sequence ⟨init, setState(“green”), setState(“yellow”), setState(“empty”), getState, quit⟩ controls the associated Java implementation of a peg hole

Higher-Order Processes contain components that are themselves process instances. In Fig. 8, process *Peg_dialog* contains two components: *sm* is an instance of process *Peg_hole* - embedded in a test frame, *tm* is an instance of *Guess_peg* that behaves like its sub-process *Option2* (encircled 3 in Fig. 8). The bottom panel of the animator tool in the figure visualizes the hierarchical component structure (component tree). Sub-processes and operations of higher-order processes describe the interaction of their components. Higher-level operations describe new emerging behavior or conflate sequences of components’ operations into new atomic behavioral units to increase the level of abstraction in a description. Operation *replacePeg* in Fig. 8 (encircled 1) is e.g. an abstraction of sequence *⟨tm.replacePeg, sm.setState⟩*.

The HOPS notation makes it possible to describe a system from different view-points and to couple these views by using higher-order processes. In Fig. 8, component *tm* reflects a task perspective while component *sm* describes a part of an artifact. HOPS supports both bottom-up as well as top-down thinking. On the one hand, the behavior of a process is determined by its components. On the other hand, it has partial control over the components. Partly redundant models can be used to enable description of emerging constraints and of distributed control (see e.g. [4]). For example, higher-level operation *replacePeg* in Fig. 8 is only enabled in an animation run if operations *replacePeg* and *setState* are enabled in the components *tm* and *sm*. Additionally, higher-order processes focus on and control only some operations of their components. The control of operation *getState* remains, for example, in component *sm*. This is also to be seen in the bottom left part of Fig. 8 (encircled 2) where *getState* is enabled in node *sm* of the component tree. In other words, the behavior of a higher-order process P is defined by a set of sequences of operations which are either defined by P itself or which are defined by components but not controlled by P.

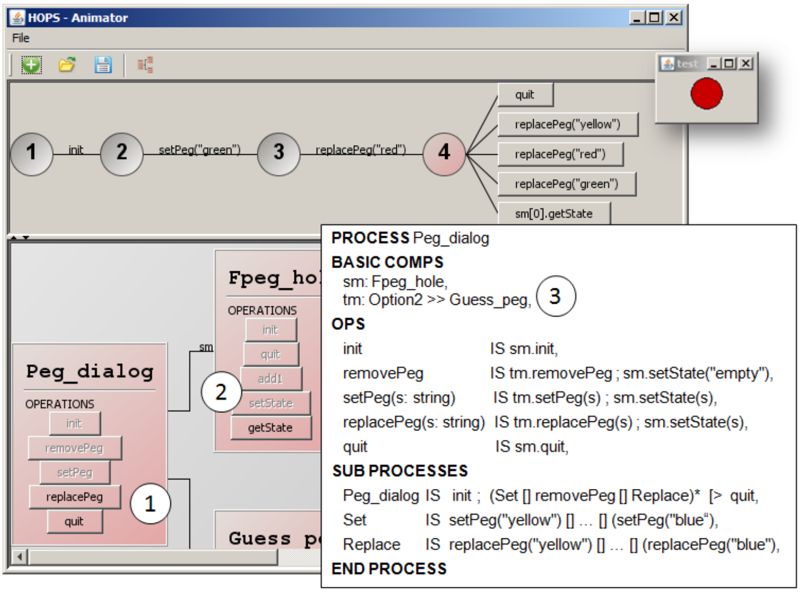


Fig. 8. Specification of higher-order process *Peg_dialog* (bottom right) and screenshot of the HOPS-animator after animating sequence ⟨init, setPeg("green"), replacePeg("red")⟩

5.2 Selective Modeling Techniques Supported by HOPS

Model coupling, model-guided prototyping and deliberated underspecification are selective modeling techniques that support resource-based design ideas, the application of analytic and empirical means, the co-evolution of design representations and the exploration of design options. Higher-order processes implement model coupling. Mappings between HOPS processes and Java classes as explained in the previous section can be used for model-guided prototyping. Deliberated underspecification is facilitated by partly redundant descriptions with distributed control in HOPS. The ideas are illustrated by considering aspects the example problem at different levels of granularity: the design of an interactive peg hole, of an interactive peg row, and of the whole mastermind board.

Peg Holes – Exploration of Design Options by Model Coupling. Higher-order process *Peg_dialog* in Fig. 8 couples a task model describing how to place a peg and a model of an interactive peg hole with few constraints on its behavior. *Peg_dialog* integrates both views but describes less constrained behavior than the task model component *tm* itself. Hence, *tm* fully controls the use of the peg hole (component *sm*): a peg can be set, and then it can be replaced arbitrary often. However, two other variants of a task model were given in the definition of process *Guess_peg* (subprocesses *Option1*, *Option3*). Fig. 9 illustrates effects on the behavior of *Peg_dialog* if the behavior of component *tm* is switched to these options. In this way, different alternatives for refining the behavior of interactive peg holes can be explored.

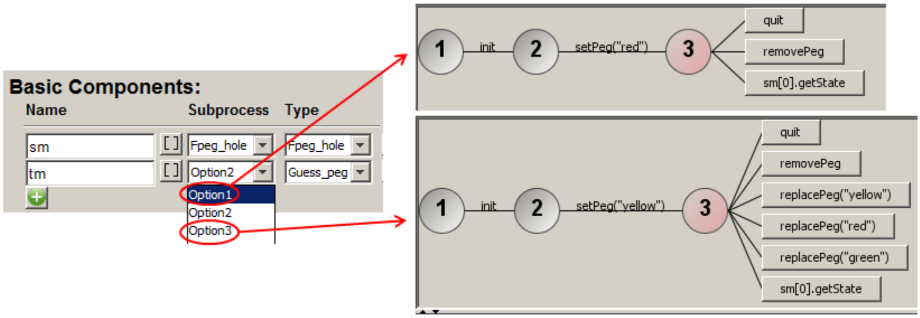


Fig. 9. Different options of the task model are coupled with the underspecified model of an interactive peg hole to explore possible refinements of its behavior in animation runs

A selected integration of different, but overlapping views on the design space supports resource-based thinking rather than task-based thinking because task models do not play the dominant role anymore. The focus of attention is reflected in the root process of the actual process composition. Its type is not restricted to task models. In a further step it is shown how prototypes and models about roles, artifacts and tasks can be used to explore different resource allocations.

Peg Rows - Model-Guided Prototyping. Model-guided prototyping in HOPS is a technique where Java implementations and HOPS models are loosely mapped to each other. Models describe only those aspects that are the actual focus of analysis. During animation they partially control the prototypes. This provides an analytical means for designers while, at the same time, prototypes can be tested empirically to a certain extent. In the example in Fig. 10, a prototypical implementation of rows of peg holes is given. In this case, the implementation of peg holes was refined according to *Option3* in Fig. 9. Users can replace pegs but they can also remove pegs from holes. This is very important for the experience of the game. The basic HOPS process does not model single peg movements anymore but is focused on enabling/disabling of a row and on setting/getting the whole code. Otherwise, users can interact directly with the prototype as indicated in the bottom of the figure.

Mastermind Board - Exploration of Resource Allocations by Deliberated Under-specifications. A co-evolution of representations is assumed to better ground design ideas in existing practices. It can drive the production of systems specifications or prototypes but it can also "slow down" the design process by re-considering current activities and describing them in more depth. It was shown in Sect. 3 that task migratability can only be achieved in systems if designers acquire a deep understanding about possible usage situations and how they could be resourced by artifacts. Task-based design approaches may easily result in artifacts that unnecessarily constrain the users' activities. For example, an interactive mastermind board may be too restrictive if the peg holes of free rows cannot be accessed or if the computer immediately starts to mark the codebreakers' guesses after they placed the last peg. In [6], (Dynamic) Task allocation is considered as task transformation and as

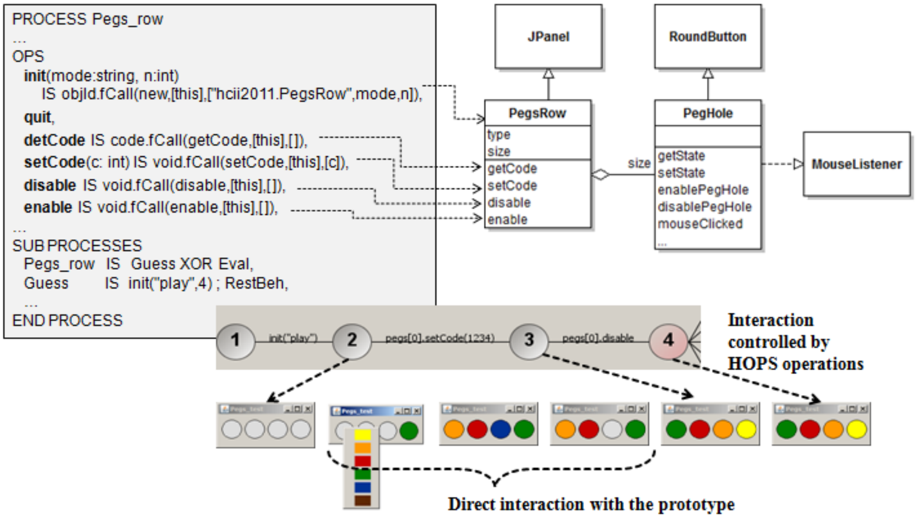


Fig. 10. Top: mapping between HOPS operations and the implementation for rows of peg holes (mastermind), bottom: test of the prototype, partially controlled by the HOPS process

resource allocation problem. It is important to understand existing transfer and control mechanisms in the interplay between different actors and (internal and external) artifacts in order to adapt them.

Fig. 11 illustrates a coupling of HOPS models that can be used to explore the design space of a mastermind board. The components describe assumptions about the actors in a mastermind game and about resources that are supplied by the interactive prototype. The computer acts as codebreaker and a human as codemaker in the specified situation (HOPS components *cb* and *cm*). The model of the activity *Break_code* is depicted in the right top corner. Take note that this model describes the codemaker signaling that the marking is finished. Although people give also such signals when they play together (e.g. by leaning back in their chair) this was not considered in the task description of Fig. 4.

An essential idea of selective modeling is that designers deliberately underspecify representations which are in the actual focus. They are coupled with complementary design representations to develop assumptions about possible usage situations and to reflect design options. In Fig. 11, the models of actors and of the code-breaking activity help to put constraints on the prototype of an interactive mastermind board and to test variants of its refined behavior in model-guided animation runs. In the interaction scenario at the bottom of the figure, the computer made a guess and the peg row was disabled. Then, the user marked the guess (indicated by HOPS operation *setPegs*) and signaled the end of this process (operation *finish*). This caused a disabling of this marking space. The prototype has to generate a new guess now.

Other models could assume e.g. a shared evaluation of guesses. There are two human players but the computer could check whether the markings are correct. This could be useful in situations where the codebreaker is a beginner and the codemaker

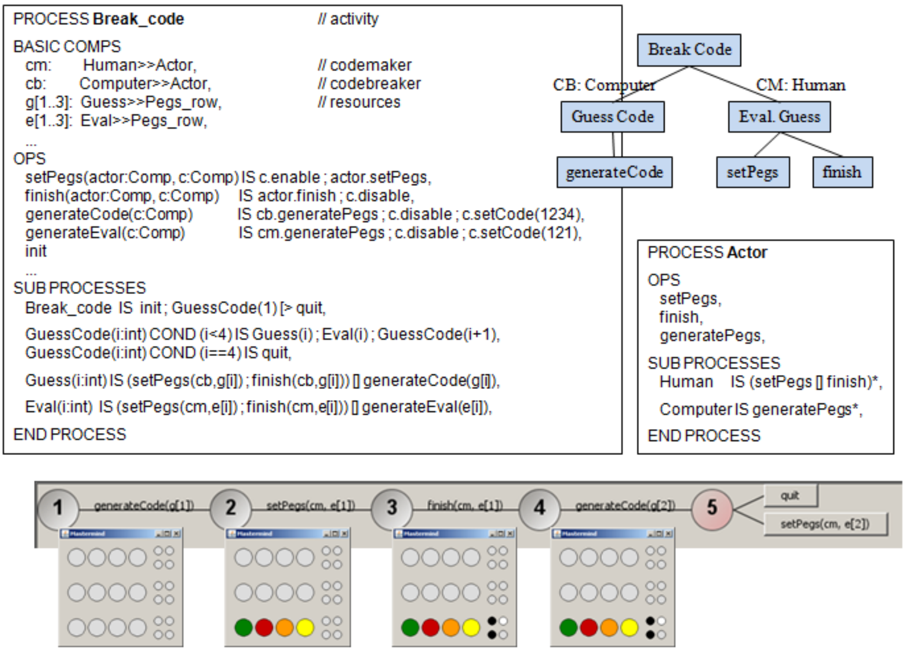


Fig. 11. The underspecified prototype of an interactive mastermind board (mainly described by components g[1..3], e[1..3], see Fig.10) is examined and explored for situations with the computer as codebreaker and the user as codemaker

is more concentrated on giving explanations than on marking. The co-evolution of models helps to develop a more fine-grained understanding of resource allocations and informs design decisions.

6 Summary

"Design are hypotheses about how artefacts shape cognition and collaboration." (D. Woods)

Task migratability is a usability principle that encourages designers to develop a deep understanding about possible usage situations of an interactive artifact. Refined flexible task allocation is rarely achievable through pure top-down decomposition of tasks. Resource-based conceptions of human activity seem to be more appropriate than task-based ones to face complex design problems.

Selective modeling has been proposed as a technique to adapt some of the assumptions and practices in model-based design. Designers are encouraged to co-evolve different representations which describe the problem space from different viewpoints. They are partially coupled. Sometimes they are deliberately under-specified to allow for reflection and exploration by other models. The interactive artifact under design "emerges" from these descriptions. Selective modeling is a model-guided but not a model-driven process. The designers need to be aware of the actual design situation.

Design reflects what we consider as important in the current world but also what we don't see or too late. The co-evolution of different design representations is often recommended in UCD in order to ground system design in current work practices. However, an effective coupling is still not easy to achieve. The general approach of selective modeling has been demonstrated in the paper at the example of the HOPS formalism and prototypes in Java. The examples given in this paper demonstrate the practicality of the method. Further work aims to improve mapping mechanisms between models, to improve tool support and to explore a broader range of examples.

Acknowledgments. We would like to thank Stefan Piehler, Toralf Hübner, and Mathias Kühn for their contributions to the HOPS tool. Our thanks also go to the reviewers of this paper for their useful comments.

References

1. Dix, A., Finlay, J.E., Abowd, G.D., Beale, B.: Human-Computer Interaction, 3rd edn. Prentice-Hall, Englewood Cliffs (2003)
2. Benyon, D., Turner, P., Turner, S.: Designing interactive systems: people, activities, contexts, technologies. Addison-Wesley, Reading (2005)
3. Dittmar, A., Forbrig, P.: Task-based design revisited. In: Proc. of ACM SIGCHI Conf. on Engineering Interactive Computing Systems (EICS 2009), pp. 111–116 (2009)
4. Dittmar, A., Harrison, M.D.: Representations for an iterative resource-based design approach. In: Proc. of ACM SIGCHI Conf. on Engineering Interactive Computing Systems (EICS 2010), pp. 135–144 (2010)
5. Diaper, D.: Understanding Task Analysis for Human-Computer Interaction. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction. Lawrence Erlbaum Associates, Inc., Mahwah (2004)
6. Dearden, A., Harrison, M., Wright, P.: Allocation of function: scenarios, context and the economics of effort. Int. Journal of Human-Computer Studies 52(2), 289–318 (2000)
7. Degani, A.: Taming HAL: Designing Interfaces Beyond 2001. St Martin's Press Inc., New York (2004)
8. Carroll, J.: HCI Models, Theories, and Frameworks: Toward a multidisciplinary science. Morgan Kaufman Publishers, San Francisco (2003)
9. Johnson, P.: Human computer interaction: psychology, task analysis, and software engineering. McGraw-Hill Book Company, New York (1992)
10. Annett, J.: Hierarchical Task Analysis. In: Diaper, D., Stanton, N. (eds.) The Handbook of Task Analysis for Human-Computer Interaction, Lawrence Erlbaum Associates, Inc., Mahwah (2004)
11. John, B.E., Kieras, D.E.: The GOMS family of user interface analysis techniques: Comparison and contrast. ACM Trans. on Computer-Human Interaction 3, 320–351 (1996)
12. Dourish, P.: Where The Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2001)
13. Suchman, L.: Plans and Situated Actions: The Problem of Human Machine Interaction. Cambridge University Press, Cambridge (1987)
14. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Comput.-Hum. Interact. 7(2), 174–196 (2000)

15. Kaptelinin, V., Nardi, B.A.: Acting with technology: activity theory and interaction design. MIT Press, Cambridge (2006)
16. Wright, P.C., Fields, R.E., Harrison, M.D.: Analyzing human-computer interaction as distributed cognition: the resources model. Human Computer Interaction 15(1), 1–42 (2000)
17. Paterno, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, Heidelberg (2000)
18. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
19. Rosson, M.B., Carroll, J.M.: Usability Engineering – Scenario-Based Development of Human-Computer Interaction. Morgan Kaufmann Publishers, San Francisco (2002)
20. Beyer, H., Holtzblatt, K.: Contextual Design – Defining Customer-Centered Systems. Morgan Kaufmann Publishers, San Francisco (1998)
21. Greene, S.L., Jones, L., Matchen, P., Thomas, J.C.: Iterative development in the field. IBM Syst. J. 42(4), 594–612 (2003)
22. Gould, J.D., Lewis, C.: Designing for usability: key principles and what users think. Communications of the ACM 28(3), 300–311 (1985)
23. Fischer, G., Lemke, A., McCall, R., Morch, A.: Making Argumentation Serve Design. In: Moran, T., Carroll, J. (eds.) Design Rationale: Concepts, Techniques and Use, pp. 267–293. Lawrence Erlbaum Associates, Inc., Mahwah (1996)
24. Barboni, E., Ladry, J.-F., Navarre, D., Palanque, P., Winckler, M.: Beyond Modelling: An Integrated Environment Supporting Co-Execution of Tasks and Systems Models. In: Proc. ACM SIGCHI Conf. on Engineering Interactive Computing Systems (EICS 2010), pp. 143–152 (2010)
25. Dittmar, A., Hübner, T., Forbrig, P.: HOPS: A Prototypical Specification Tool for Interactive Systems. In: Graham, T.C.N. (ed.) DSV-IS 2008. LNCS, vol. 5136, pp. 58–71. Springer, Heidelberg (2008)

Structuring and Composition Mechanisms to Address Scalability Issues in Task Models

Célia Martinie, Philippe Palanque, and Marco Winckler

Institute of Research in Informatics of Toulouse (IRIT), University Paul Sabatier, France
{martinie,palanque,winckler}@irit.fr

Abstract. Along tasks analysis and modeling history it has been demonstrated by experience that task modeling activities become cumbersome when performed on large, real-life systems. However, one of the main goals of task models is to provide designers with a structured and complete description of the users tasks especially when these user tasks are numerous and/or complex. Several authors proposed to handle that problem by providing tools aiming at supporting both construction and understanding (usually via simulation) of models. One of the most popular examples is CTTE environment which is dedicated to the engineering of CTT task models. The paper shows how to extend notations for task description with two kinds of mechanisms: composition and refinement/abstraction. Refinement/abstraction mechanisms make it possible to decompose a task model into several models and to interconnect them. Composition mechanisms make it possible to define communication means between task models. The paper proposes a precise definition of these mechanisms, their integration into a notation for describing task models and demonstrates that altogether, these two structuring mechanisms support the effective exploitation of task models for large scale application. The use of the mechanisms is presented on a real-life case study from the space domain describing operators' tasks to monitor a satellite and manage failures.

1 Introduction

In the field of Human-Computer Interaction (HCI), the user centered paradigm [22] has reached a popularity level where the question about it, is no longer whether it is valid or not but rather how it should be embedded in the development process of interactive applications. Many notations, processes and tools have been proposed for gathering information about the users either in formal (via formal requirements as in [17] or formal task models [27]) or informal ways (via brainstorming [7, 9] or prototyping [31]). One of the main advantages put forward by notations is that they make it possible to handle real-size applications and, if provided with a formal semantics, make it possible to reason about the models built with the notations[1] and assess the presence or absence of properties.

[1] In the rest of the paper we make the following distinction between a notation and a model: a notation provides a mean for representing information from the real world. The resulting use of the notation is called model. If a notation is defined formally it is called a formalism.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 589–609, 2011.

While exploiting a notation for building a model of the real world, usually, two main activities are carried out in order to tackle the problem of size of the resulting model:

Abstraction: details from the real world will be omitted and abstracted away if they are not of interest for the use that will be made of the model. For instance, using a finite state automata (FSA) [14] the continuous evolution of the world (such as in a flow of liquid while emptying a bottle) will be discretized in a set of states (for instance three different states: Full, Emptying and Empty) representing an abstraction of the infinite number of states of the real world.

Filtering: the notation plays the role of a filter capturing the information it is able to capture and letting through what cannot be captured. Taking again the example of a FSA, information about the time elapsed for emptying the bottle cannot be represented and thus will not appear in the model.

Of course, notations can be (and have been) extended in order to be able to represent more information than initially planned. FSA have been extended in [38] to handle data (such as for instance the size of the bottle) or even time [30]. However, the most widely used notations stay away from universality and embed strong abstraction and filtering. Examples of such notations are UML class diagram [29] or entity relationship diagrams [5] only capturing data-related information or basic Petri nets [23] or FSA [14] only capturing behavior-related information. Indeed, despite these mechanisms a notation capturing too much information would end up in oversized and unmanageable models. To represent a larger part of the world, several models have to be built using several notations. The complexity then lays in defining processes and tools making it possible to ensure conformance and consistency of the various models corresponding to different views of the same world.

In the field of HCI many notations have been proposed for capturing in models the various elements of socio-technical systems. In the last decades, several tasks notations have been developed as means to describe work carried out by users whilst interacting with a system [8]. Despite the fact that various specific task notations exist, they are mainly structured around two concepts: task decomposition (often represented as a hierarchy) and task flow (for showing the order in which tasks are executed) [16]. When adequately combined, these concepts can provide an exhaustive and complete representation of large quantity of information in a single model. However, as discussed by Paterno & Zini [28], when applied to real-life systems, tasks notations end up in very large, hard-to-manage models thus making task modeling a time-consuming and sometimes painful activity.

This paper introduces some generic structuring mechanisms to tasks notations for describing task models in order to overcome the problems identified above and to make it possible to represent users' activities in large socio-technical systems. Section 2 discusses the structuring mechanisms offered by existing tasks notations. It also includes an abstract presentation of the proposed mechanisms. Section 3 introduces informally the real-life case study from the satellite ground segments domain focusing on controllers' tasks to monitor a satellite and manage failures. It also presents an initial approach for modeling controllers' tasks with CTT [26] and highlights the need for structuring mechanisms. Section 4 gives an overview of HAMSTERS (introduced in [1]) and details how the notation has been extended to integrate the new mechanisms. Section 5 details how the various elements of the notation are used to model the entire case study. Section 6 is dedicated to the analysis in terms of efficiency of the proposed mechanisms for structuring task models. Section 7 concludes the paper and presents research directions for future work.

2 Structuring Issues for Tasks Notations

Task models are useful when designing real-size systems as they aim at representing a large quantity of information related to user goals and to the activities to be carried out in order to reach these goals.

Notation-Based Structuring: Whilst some task notations such as UAN [13] and GOMS [4] are mainly textual, CTT [26, 24], MAD [33] or AMBOSS [1] provide graphical representations, which favor legibility and understanding of complex problems [6]. Every task notation proposes hierarchical task decomposition. This hierarchical representation of tasks is grounded in psychology [32] reflecting in the models the way people structure their activities. The decomposition of tasks in subtasks enables the creation of several levels in the tree hierarchy of tasks; this has been extensively used for supporting the mechanisms of abstraction and refinement in task models.

Not all notations used for describing activities are hierarchical. Indeed, notations used in the workflow domain (such as YAWL [36] or BPMN [36]) are graphs but hierarchical representations are more prominent in task model notations as this enforces abstraction/refinement mechanisms [36]. In hierarchical representations, the order of execution of tasks is given by the navigation through the hierarchy of tasks and the temporal operators between sibling tasks (i.e. tasks in the same level of the hierarchy). By combining hierarchy and temporal operators, it is possible to have different tasks models that exhibit equivalent behavior [10].

Another important issue for structuring task models refers to the relationship between tasks and goals. In the Task Knowledge Structure (TKS) [15] for instance, a specific substructure corresponds to each goal. CTT [26, 24], MAD [33], and GTA [35] provide a more flexible organization of tasks so that a goal can be reached in different ways. Van Welie, van der Veer and Eliëns [35] argue that the higher the tasks in the hierarchy the closer they are of organizational goals, whilst low-level tasks are more likely to represent individual goals. Even if the distinction between individual's and organization's goals is sometimes difficult to settle, it has in the end an impact on how models are structured.

Structuring mechanisms can sometimes be found in unexpected places. For example, CTT [26] includes the operator *iterative tasks* (symbol next to a task*) so that repetitive tasks are represented only once even though they may occur many times. That operator is mainly used for describing iterative behavioral aspect of the task but additionally makes it possible to significantly reduce the size of a task model. CTT offers another structuring mechanism via the notion of collaborative tasks modeled using the operator *connecting tasks* (symbol ↔). In a nutshell, this notion embeds in models the fact that some tasks require the involvement of several persons (or roles) to be performed. The task of each person is modeled in a task model and the collaboration (i.e. order dependences between the tasks of the operators) is represented in another additional model. While necessary to represent collaborative activities, this notion and its related operator results in a role-based structuring of task models splitting a bigger model in several smaller models. CTTE is the unique tool currently available supporting *collaborative tasks*. However, only qualitative temporal relationships can be represented as no information (i.e. data flow) can be exchanged between the models.

An important issue that must be considered when structuring task models is its potential for reuse [3]. In fact, some tasks (such as login into systems for instance) remain structurally similar in different applications. This feature has been introduced in notations like CTT [26] so that some generic tasks can be used as building blocks that can be cut and pasted in models. However, a modification of one of these copies is not reflected to the other ones. Concerned by the reusability of tasks models, Gaffar et al. [12] have investigated structuring mechanisms around the notion of patterns to be used in task models. They propose a method and a tool to model generic tasks patterns as building blocks that can be instantiated and customized when modeling real-life socio-technical systems. One of the advantages of task patterns is the fact they provide more flexibility for reuse as they correspond to a generic problem.

Lastly, Sining et al. [34] introduce the notion of modular task models at a more generic level than the recursive tasks offered in CTT. Such modules would be structured in a task diagram describing in an exhaustive way their relationship. This concept is very similar to some one proposal in this paper. However, its implementation by means of task diagram adds unnecessary complexity (a new notation) and does not support the exchange of information between the modular models as proposed in the current paper.

Tool-Based Structuring: Some problems associated to the management of large and complex tasks models, can be overcome with the help of tool support. Currently, several of the tasks notations presented above are accompanied with an edition (and sometimes a simulation) environment. For instance, EUTERPE [35] supports the Groupware Task Analysis (GTA) notation, K-MADe supports the *Méthode Analytique de Description* (MAD) notation [33] and CTTE [25] supports CTT notation. These environments exploit the fact that task models are naturally represented as a hierarchy of sub-tasks, to implement abstraction/refinement mechanisms by supporting actions such as pruning/expanding sub-parts of the trees. More recently [28], CTTE tool has been enriched with visualization techniques (fisheye view and semantic zooming) to better support creation and management of CTT tasks models. Moreover, through collaborative tasks (previously described), CTTE is the only environment currently available supporting the decomposition of tasks in several communicating diagrams (even though the original goal was the representation of collaboration and not to support models structuring).

Abstraction/Refinement and Composition Mechanisms: This section has illustrated the mechanisms currently available for structuring task models. However, when it comes to large systems these mechanisms and tools are insufficient (see CTT model in Fig. 2). We propose two new mechanisms to handle more efficiently this complexity. The first one is based on refinement/abstraction principle and makes it possible to decompose a task model in several models and to interconnect them. A large task model can thus be decomposed into several sub-models. These sub-models can then be reused (as a "copy") in various places of the same model and even in other models. Each time one of these parts is modified, the modification is reflected in all the other "copies". The Composition mechanism makes it possible to define communication mechanisms between task models. This task model structuring mechanisms is similar to procedures calls in programming languages and parameterization is possible via input and output parameters. In order to keep the same semantics as for the single model, communication protocols have also been introduced.

3 Case Study: Ground Segment Operations for PICARD Satellite

In order to illustrate scalability problems of task models, in this section we present a real-life system belonging to a ground segment application that is currently used to monitor and to control the Picard satellite[2]. The task models presented in this section exploit standard structuring mechanisms such as those currently available in the CTT notation (and presented above). These models will then be revised in section 4 to include the structuring mechanisms of abstraction/refinement and composition that we have designed to overcome the limitations that were faced due to the large number of tasks and activities.

3.1 Informal Description of Case Study

The case study belongs to the space domain and more precisely to the command and control interactive systems of satellites. **Fig. 1** presents a schematic view of a space system as defined in the European Standard ECSS-E-70 [11]. Such space systems are made of two parts: the *space segment* (including the spacecraft) and the *ground segment* which is made up of the ground station segment (antennas for communication with the space segment) and the mission control system. Our focus is the *operation control system* (bottom-left icon in **Fig. 1**) which belongs to the mission control system. This system is in charge of maintaining the spacecraft in operation and is thus heavily related to the spacecraft structure and functioning. Next paragraphs describe the context, goals, roles, tasks and system functions of the case study that have to be represented in the task models. We only provide here the excerpt necessary to discuss the various models and the way they can be decomposed and structured. Indeed, tasks and operations of a mission control system are much more numerous than what is presented here but this excerpt is representative of the real system in terms of complexity and of operators' everyday activities.

3.1.1 Ground Segment Operations: The Context

The Operation Control System consists in a room containing several computers and various hardware equipment dedicated to: 1) receive information from the satellite via the ground station segment, 2) organize and record this information and 3) send commands to the satellite (space segment). In this room, controllers (usually called operators) are in charge of monitoring the state of the space segment platform part by examining periodically several parameters such as current voltage of the power

[2] The Picard satellite was launched by CNES in June 2010 and is dedicated to solar observation.

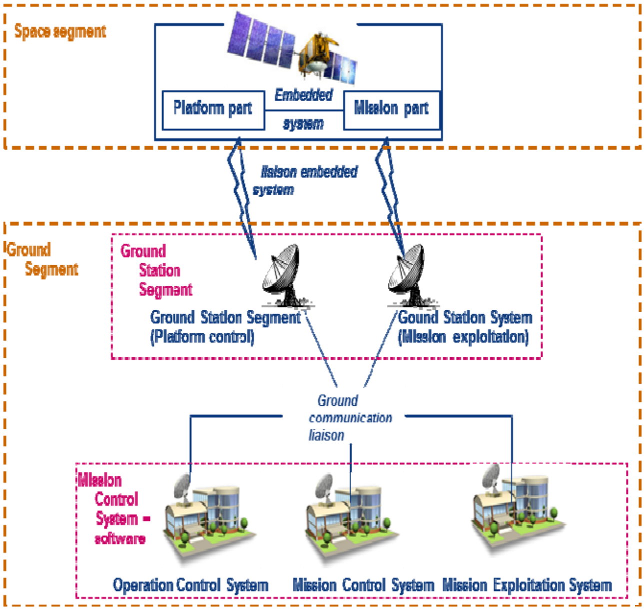


Fig. 1. The satellite application domain in a nutshell

generators, current satellite mode, and status of the communication link. The main goal of operators is to keep the satellite in a nominal mode, so that the mission (for instance observation of the sun by taking pictures) can be performed. If a severe incident occurs, the satellite can automatically switch to a mode called "survival mode" where all functions are disabled and the mission is stopped (i.e. data gathering is stopped and already collected data might be lost). In case of the occurrence of an adverse event, the team has to avoid that the satellite switches to this mode, except if it is to prevent a satellite loss. When a controller detects an adverse event (usually a failure) and understands the issue, he/she has to apply a *procedure*, selected from a list of referenced *procedures*, to recover from that failure. If the failure is more complicated to understand, he/she has to inform the entire team (one or more satellite engineer and experienced controllers can collaborate) to solve the issue and select the adequate *procedure*. If such procedure is not available they might need to design a new one that in turn has to be entered in the ground segment and sent to the space segment). The operator often collaborate with other controllers or with dedicated engineers such as, for instance, Radio Frequency engineers when special operations on the communication link are required.

3.1.2 Task Analysis of a Satellite Controller's Activities

Controllers are in charge of two main activities: observing periodically (i.e. **monitoring**) the vital parameters of the satellite and performing **maintenance operations** when a failure occurs. Depending on the satellite between a couple of thousands and tens of thousands parameters have to be monitored. The more frequent and relevant monitoring activities include observing: *satellite mode*, T*elemetry* (measures coming from the satellite), *Sun array drivers statuses*, *error parameters for* *the platform*, *error parameters for the mission*, *power voltage* (energy for the satellite), *ground station communication status*, and *on board computer main parameters*.

The number of *procedures* for maintenance operations goes beyond the hundred. Due to page limits, we only present a selection of three sub-routines concerning failure cases that are critical for the mission. However, these sub-routines allow us to exhibit all the problems that we faced and that were related to the structuring of task models. The sub-routines are:

- **Recovering from a power voltage issue:** a wrong voltage parameter value is detected and the controller has to reset the satellite flight software.
- **Re-establishing the communication link:** the ground operation control system may not be receiving Telemetry from the satellite. The first activity of the controller is also to reset the flight software in this case. The other activities are not presented.
- **Investigating why an automatic switch to survival mode occurred:** Most satellites (including Picard) are not always all the time in 'visibility' of a ground station (only geostationary ones are). For such satellites the parameters are updated and the telecommands (TCs) sent when the satellite is visible for one of the ground station. Meanwhile, it evolves in an autonomous mode (self-triggering On Board Control Procedure (OBCP) if needed). During a non-visibility period if vital parameters' values go beyond or below a given threshold, the satellite flight software (SW) switches itself to survival mode. In the next visibility period, controllers will have to understand what happened, find a solution and then send TCs to set the satellite back to its nominal mode.

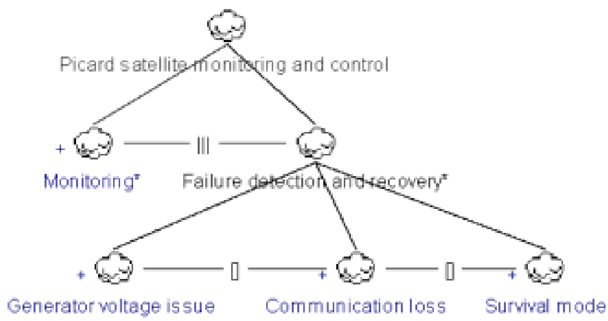
Furthermore, after each failure detection and recovery sub-routine, the controller has to record the failure that happened in a dedicated application.

Lastly, the case study exhibits operators' activities related to the management of the communication link between the Ground Segment and the Space Segment which also has to be monitored and possibly repaired.

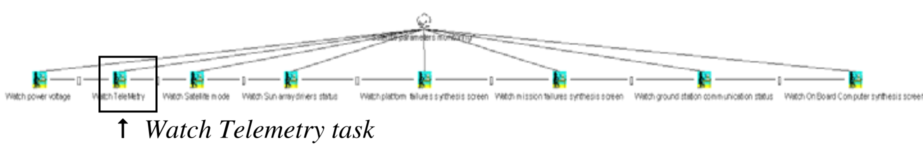
3.2 Task Model Using CTT

Our first attempt of modeling operator's tasks into sub-goals ended up with unreadable models due to their size. We present here one of the original models using CTT notation [27] (and CTTe tool [25]) as this notation is widely used in the HCI community. The point is not here to read the model but to show first that real-life tasks can be modeled and then that the resulting model can be too large to be handled in an efficient way. The tasks have been decomposed into several models, all presented in Fig. 2. Fig. 2a provides a high level view on the two main operators' tasks namely *Monitoring* and *Failure detection and recovery*. In this model we have used CTTe abstraction mechanism (represented in the model by the symbol + corresponding to abstracting away further details about these tasks). These two tasks are detailed in Fig. 2b and Fig. 2c, respectively. The subtask *Watch Telemetry* is highlighted to show where it appears in both models. This is a typical example of multiple appearance of the same task (which could be an entire sub-tree) in various models. Modifying information about it (name, duration, precondition, temporal operators …) is really cumbersome as these modifications have to be repeated manually. In addition, it is not possible to express in CTT the fact that these two sub-trees represent the same activity.

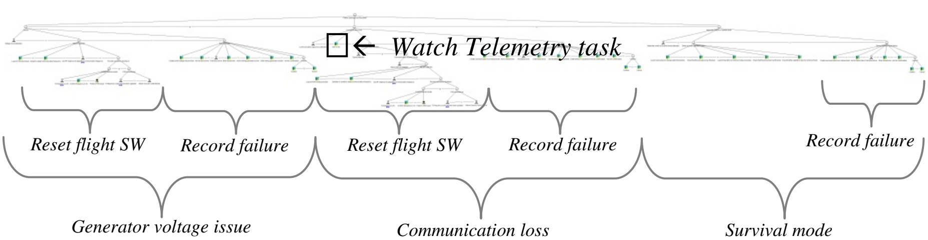
Fig. 2c provides the decomposition for the tasks *Failure detection and recovery* including subtasks *Generator voltage issue, Communication loss* and *Survival mode*. It is important to note that these three tasks only represent a limited subset of the hundreds of procedures and related tasks needed to control the satellite, but despite that, the model is already large and complex. The model in Fig. 2c contains 74 tasks, for 3 types of failure (described in section 3.1.2), and then 3 types of procedures to set the satellite back to its nominal mode. For each type of failure, a special task tree is performed by the controller, who then additionally records the failure in a dedicated application. This is the reason why the "Failure detection and recovery" sub-goal has several parts in common (both "*Reset flight software*" which appears twice and "*Record failure*" which appears three times). By lack of structuring mechanisms the corresponding tasks are duplicated. Furthermore, most of the tasks in the operation of recording a failure are also duplicated, even though there are small differences related to the fact that they have to be customized according to the type of failure.



a) Top-view of controllers' tasks, including *Monitoring* and *Failure detection and recovery*



b) Decomposition of sub-task *Monitoring* highlighting the atomic task *Watch Telemetry* that also appear in other models (see figure below)



c) Decomposition of task *Failure detection and recovery* highlighting the reoccurrence of tasks *Reset flight SW* and *Record failure* into the hierarchy of subtask *Generator voltage issue, Communication loss* and *Survival mode*

Fig. 2. Models of operators' task using CTT notation

3.3 Issues Raised by the Legibility and Scalability of the Tasks Models of the Case Study

The modeling of the case study using standard structuring mechanisms resulted in a complex and cumbersome activity.

One of the problems is that some tasks (in particular monitoring tasks) appeared several times (without specific temporal constraints) in the same model or even in several procedures. To model this kind of behavior it is required to duplicate monitoring tasks in various places of the task models.

Additionally; currently available structuring mechanisms do not support accurate representation of data flow. To describe complex conditions involving data and/or objects associated to tasks, the designer must combine hierarchies and operators between siblings' tasks in order to represent the expected behavior. As a consequence, the available solution leads to the multiplication of tasks in the model, thus making them illegible as demonstrated in Fig. 2. It is important to note that we have only modeled 3 types of failures, but at operation time in the Command and Control Centre, the number of possible failures is much higher. This demonstrated the fact that tasks notations have to be extended[3] in order to handle such problems if they are to be used for representing operators' activities in real-life complex systems. It is important to note that this aspect is critical for task modeling. Indeed, task modeling is precisely meant to be used for large scale systems. Indeed, small scale systems usually do not require the use of task models for understanding users' activities that are usually rather simple to describe.

4 HAMSTERS and Its Extensions for Structuring Models

In order to demonstrate the use of the proposed structuring mechanisms for task models, we used a tool-supported notation called HAMSTERS (Human-centered Assessment and Modeling to Support Task Engineering for Resilient Systems). The rational for the choice on HAMSTERS, is that we have the possibility to extend both the notation and the tool (as the tool has been developed by us). However, as stated in section 2 we believe the mechanisms are generic and thus could be integrated in many other notations and tools.

4.1 Short Introduction to HAMSTERS

HAMSTERS notation is inspired by existing notations, especially CTT [26] and has been designed to remain compatible with CTT (from the point of view of people building the models) as models are hierarchical and graphically represented featuring operators between the tasks. However, HAMSTERS includes extensions such as pre-conditions associated to task executions, data flow across task models, more detailed interactive tasks… HAMSTERS models can be edited and simulated in a dedicated environment which also provides a dedicated API for observing editing and simulation events making it possible to connect task models to system models [2] and [19] based on the Interactive Cooperative Objects formalism [20] and [21]. The

[3] It would have been possible to decompose the main model into sub-models within CTTe tool [25], but simulation would have been impossible.

notation presented hereafter provides additional extensions that are required for structuring models. Table 1 illustrates HAMSERTS' constructs, including:

- Abstract task is a task that involves sub-tasks of different types.
- System task is a task performed only by the system.
- User task is a generic task describing a user activity. It can be specialized as a Motor task (e.g. a physical activity), a Cognitive task (e.g. decision making, analysis), or Perceptive task (e.g. perception of alert).
- Interactive task represents an interaction between the User and the System; it can be refined into Input task when the users provide input to the system, Output task when the system provides an output to the user and InputOutput task which is a mix of both but performed in an atomic way.

Table 1. Tasks types in HAMSTERS

Task type	Icons in HAMSTERS task model
Abstract Task	Abstract task
System Task	System task
User Tasks	User task, Cognitive task, Perceptive task, Motor task
Interactive Tasks	Interactive task, Input task, Output task, InputOutput task

As in CTT, each particular task of the model can be either iterative, optional or both (see Fig. 3). *Iterative* property of tasks means that a task that can be executed 1 or several times; it can be interrupted or suspended by another task. It should not be a subtask of an ENABLE operator but of an INTERRUPT or SUSPEND_RESUME operator. *Optional* tasks do not require to be executed for the goal to be reached.

Fig. 3. Optional task Iterative task and of a task both iterative and optional

Additionally, (as in CTT again) temporal relationship between tasks is represented by means of operators as described in [26]. In HAMSTERS, the notion of objects can represent both information and knowledge needed for performing a task. This information might be modified when the task is performed. HAMSTERS offers two types of relationships between tasks: the first one describes how tasks are related to other tasks (using the temporal operators as presented above) and the second one represents the information flow between tasks (as illustrated in Fig. 4 where the PIN entered in the first task is conveyed to the next task by means of input and output ports). According to the discussion in the introduction about tasks notations it could be argued that adding information to behavioral models such as task models will degrade legibility for instance. However, as such information has an impact on the availability of tasks and the action they perform they must be exhibited in order to avoid under-specified tasks models.

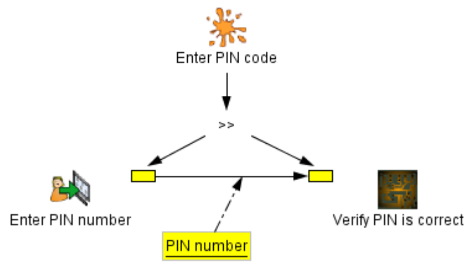


Fig. 4. Input (right-hand side of a task) and output (left-hand side of a task) flow in HAMSTERS

4.2 Structuring Mechanisms in HAMSTERS Notation

In order to address the scalability problems introduced in section 3, (as stated in section 2) we have identified 2 mechanisms: composition and abstraction/refinement. This section details how these mechanisms are integrated in HAMSTERS: composition is by means of sub-routines and the abstraction/refinement by means of sub-models. More precisely, we have added four types of sub-routines (see Table 2), parameters handling for sub-routines (see Table 3) and copy of tasks (or tasks sub-trees) (Table 4).

Table 2. Graphical representation of sub-routines in HAMSTERS

Structuring mechanisms	Related icon in HAMSTERS
Sub-routine with no input value and no output value	
Sub-routine with at least one input value and no output value	
Sub-routine with no input value and at least one output value	
Sub routine with at least one input value and at least one output value	

4.2.1 Composition Mechanism: sub-routines

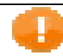
A sub-routine is a group of activities that a user performs several times possibly in different contexts which might exhibit different types of information flows. A sub-routine contains:

- The name of the sub-routine;
- The icon of an "Abstract" task type (as the sub-routine consists of a group of tasks that can belong to different types);

- Specialized input and output ports attached both to the left and to the right of the icon. The graphical symbol of these specialized ports can be filled (if they handle parameters) or not (if they don't). These ports are mechanisms for representing required parameters to and/or from sub-routines, thus providing explicit representation of data flow during task execution.

The sub-routine is then modeled in a dedicated model where the root task is the icon of the sub-routine. As stated above, the behavior of the sub-routine can be different according to the value of the parameters (see Table 3). In such a case the precondition in the sub-routine is represented by the symbol . In the main task model, the same symbol can be used by a task for changing the behavior of the model according to the value of the output parameter of the sub-routine. The value of the output parameter (if any) has to be assigned inside the sub-routine. This assignment is represented by the symbol . Such representation makes it very easy to identify in the sub-routine both the number of output parameters and where in the model those parameters are modified.

Table 3. Input and output parameters management

New artefact	Icon in HAMSTERS
Condition on a parameter	
Assignment of a value to an output parameter (of a sub-routine)	

4.2.2 Abstraction/Refinement Mechanism: Sub-models

Table 4 presents the task icon used to indicate that the task depicted by this icon is a copy of another task in the model. There can be several copies for a given task and all these copies correspond to the same sub-model. A modification in the sub-model (including its name) is thus replicated in all the sub-models. Table 4 only represents the copy icon for the "Output" task type but one "Copy" kind of icon is available for all the tasks types (see Table 1).

Table 4. Illustration of sub-models in HAMSTERS (only for an output task)

Abstraction/refinement	Icon in HAMSTERS
Sub-model (copy) of an existing Output task	

5 Case Study Modeling Using the Proposed Structuring Mechanisms

Fig. 5 shows the main task model for the main operator goal which is the monitoring and control of the Picard satellite. This main goal can be subdivided into 2 sub-routines: Satellite monitoring (highlighted by the black disc n°1) and Failure detection

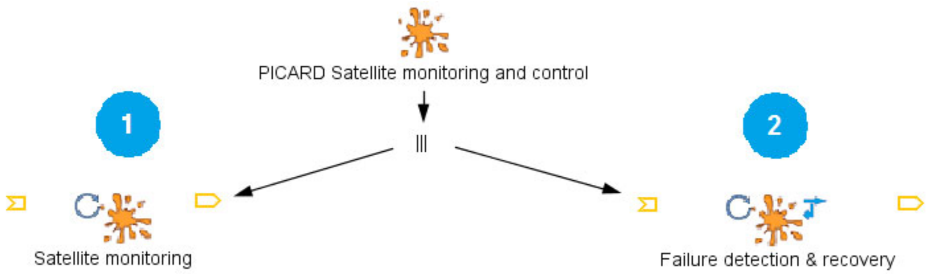


Fig. 5. PICARD satellite monitoring and control main task model with HAMSTERS

and recovery (highlighted by the black disc n°2). These 2 sub-routines are 2 sub-goals that can be achieved concurrently (||| temporal operator) and their detailed structures are presented respectively in Fig. 6 a) and Fig. 6 b).

5.1 Sub-routine with Empty Input and Output Parameters

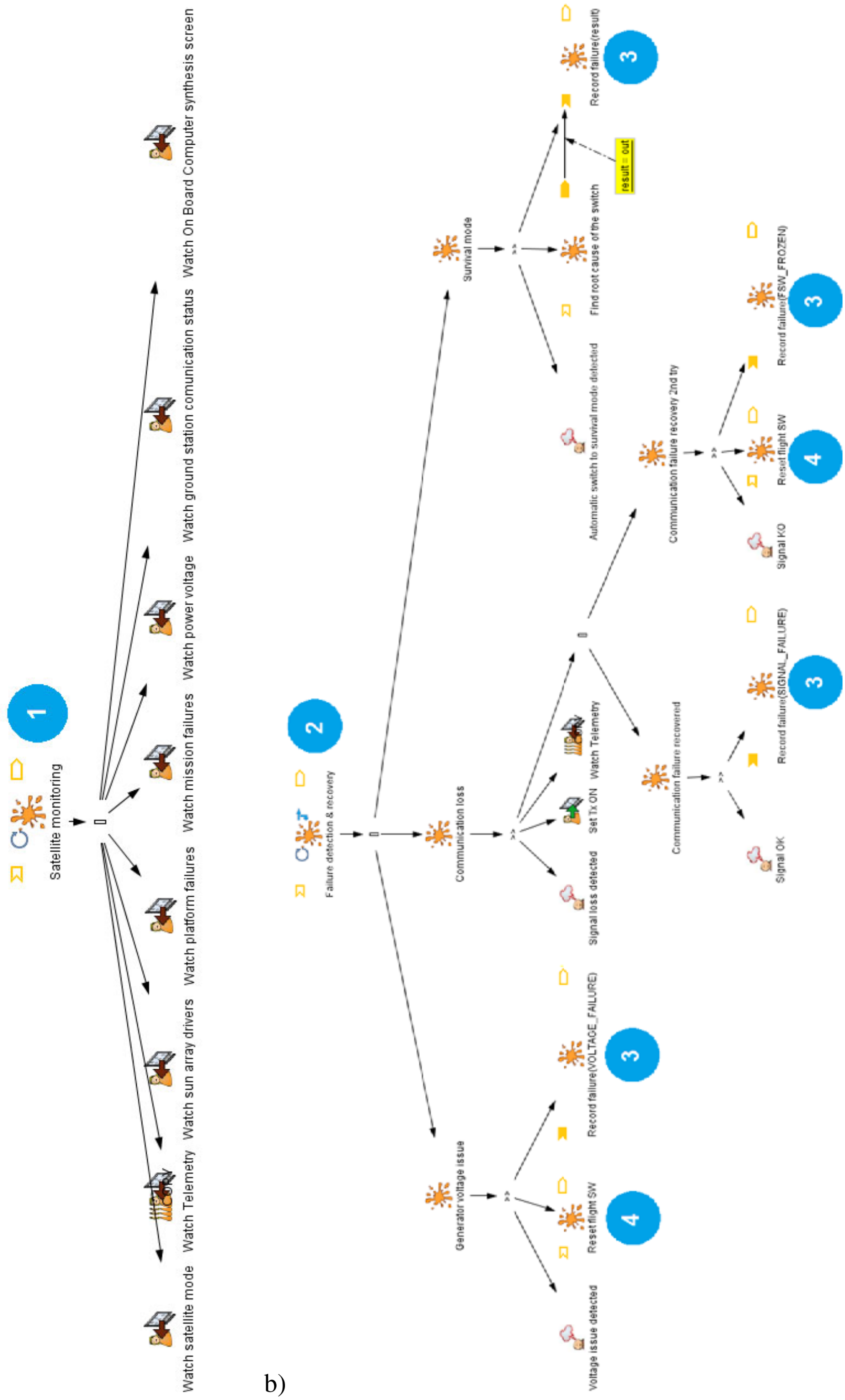
In Fig. 6 a), the sub-routine corresponding to the sub-goal "Satellite monitoring" (part 1 of Fig. 5) is decomposed into 8 output tasks that are corresponding to the monitoring activities detailed in section 3.1.2. Fig. 6 b) details the "Failure Detection and Recovery" sub-routine (part 2 of Fig. 5) and is decomposed into a choice ([] temporal operator) of 3 sub-goals, which are depending on the type of failure discovered by the controller.

The first failure type corresponds to the power voltage issue, and is detected by the controller if he/she understands (cognitive task) that the power voltage parameter is wrong. This is done by observing or just having observed (while monitoring, Fig. 6 a)) this voltage parameter. His/her activities will then consist in resetting the satellite flight software and then recording this failure in the dedicated application. Second failure type (Fig. 6 b)) is a communication link loss and is detected by the controller when he/she understands that Telemetry from the satellite is not available anymore. His/her activities will then consist in resetting the transmission link. Third and fourth failure types (last task at the second level of Fig. 6 b)) detail the activities that are performed by the controller when he/she detects that the satellite has automatically switched to survival mode.

The second and third failure types may require the controller to perform a "Reset flight SW" *procedure*, which has been modeled (disc 4 on Fig. 6 b) as a sub-routine with empty input and output parameters. This sub-routine is very useful as it avoids duplicating an entire sub-task model. The sub-routine "Reset flight SW procedure" is detailed in Fig. 7 b).

5.2 Sub-routine with at Least One Input Parameter and Empty Output Parameter

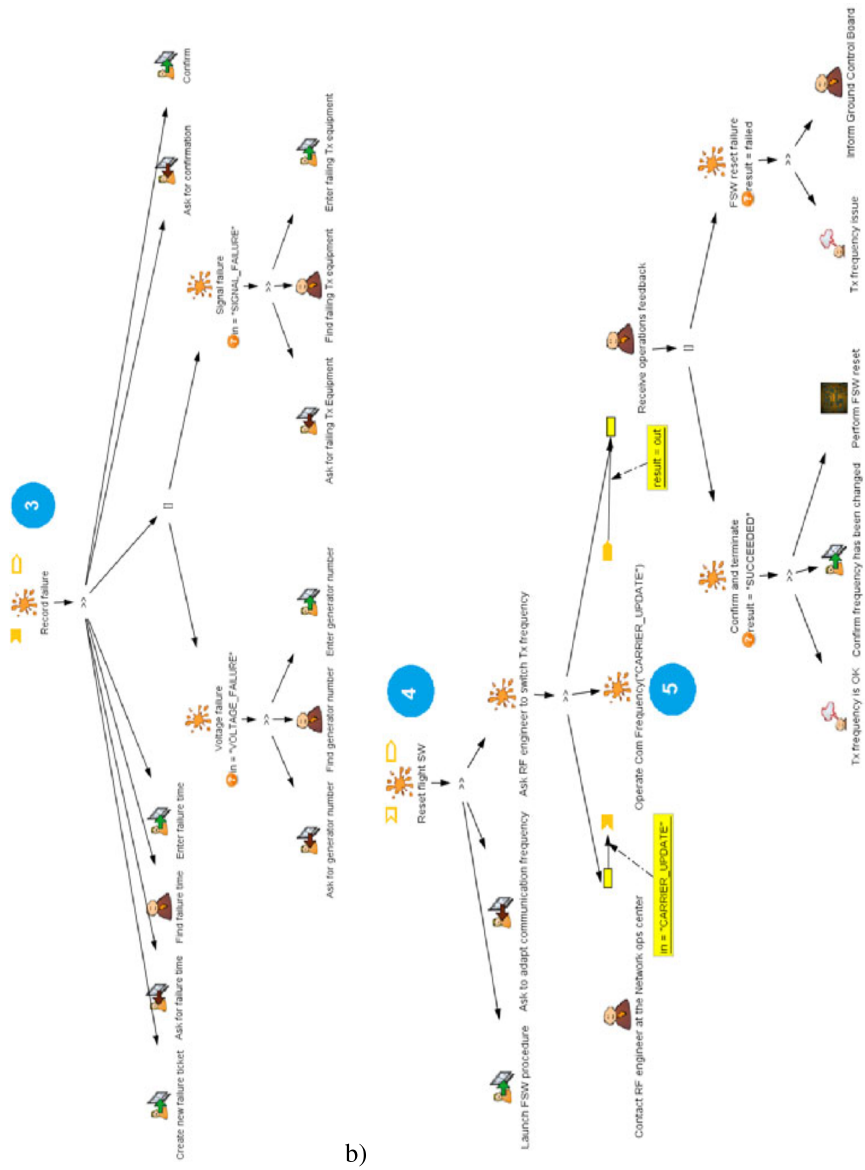
The "Record failure" sub-routine (disc 3 on Fig. 6 b detailed in Fig. 7 a) is performed by the controller each time a failure has been detected and aims at recording information in a dedicated desktop application about problems encountered in



a) b)

Fig. 6. Tasks of satellite Monitoring (a) and, failure detection and recovery (b)

operations. An input value is required, because one or more of the group of tasks that are composing the sub-routine need information to be executed. As shown on Fig. 7 a), the record of a failure can slightly differ from one failure to another, and the input parameter of the sub-routine will also help in modeling the differences into sub-task trees (using the condition parameter introduced in Table 3).



a) b)

Fig. 7. a) Record failure sub-routine b) and Reset flight SW (software) sub-routine

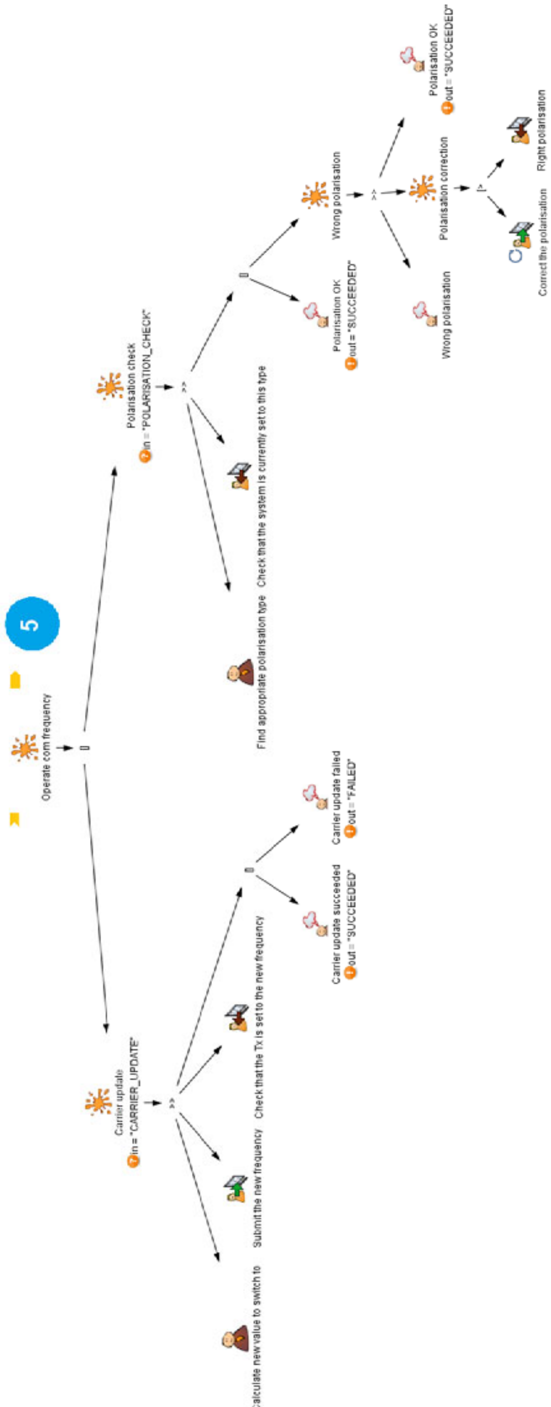


Fig. 8. Operate com (communication) frequency sub-routine

5.3 Sub-routine with Empty Input Parameter and at Least One Output Parameter

In case of an automatic switch of the satellite to the survival mode (third failure type on Fig. 6 b), the controller will perform a *procedure* to find out the root cause of this issue. It is modeled by the "Find root cause of the switch" sub-routine, and this sub-routine provides an output object that will be used by the next sub-routine "Record failure (out)" to record the failure type that has been found.

5.4 Sub-routine with Both Input and Output Parameters

In case of a "Reset flight SW" sub-routine (disc 4 on Fig. 6 b detailed in Fig. 7 a), the controller has to prepare the RF communication link. The "Operate Com frequency" sub-routine requires an input and provides an output. It indicates that the information produced while preparing the RF is then used by the controller to reach his/her goal.

5.5 Condition on a Parameter

Fig. 7 a) shows how the conditions on the input object can be used to customize the Record failure sub-routine. Indeed, in that model, there is a common part of the model dedicated to the recording of the failures but also specific parts which vary according to the failure types. In this model, if the failure type is a voltage issue ("VOLTAGE_FAILURE" is the input value of the sub-routine), the controller is asked to enter the power generator number in the application; if the failure type is a signal failure ("SIGNAL_FAILURE" is the input value of the sub-routine), the controller is asked to enter the failing transmission equipment. In both cases, failure time information and confirmation are required. In Fig. 7 a) this alternative is represented by the choice operator and the fact that each choice uses a precondition on the failure type in=SIGNAL_FAILURE for instance.

5.6 Assignment of a Value to an Output Parameter

Fig. 8 details the "Operate Com Frequency" sub-routine. It gives an example on how the assignment of a value to an output parameter (exclamation mark icon) is used to support storing of history information about what happened when the sub-routine has been performed. In Fig. 8 the assignment out=SUCCEEDED records in the out parameter the outcome of the sub-routine which can, in turn, be used as condition in the main task model. Besides, in each sub-task group, the output parameter is assigned which guarantees that the sub-routine always delivers an updated output parameter. The "out" parameter is assigned the value "SUCCEEDED" when the operation on the communication frequency succeeded, and it is assigned the value "FAILED" when the operation on the communication frequency failed.

5.7 Sub-model of an Existing Task (Copy)

On Fig. 6 b), in the second failure case "Communication loss", the controller has to send a command to set the transmission to ON (input task which corresponds to resetting the transmission) and look at the Telemetry status. This activity of observing the Telemetry status is already described in the Fig. 6 a) sub-routine as it is part of the "Monitoring" job of the controller. The artifact "Copy" of an existing task is here used to maintain consistency between the models and avoiding duplication (copy of the "Watch at Telemetry" task of the Monitoring task model described in Fig. 6 a)). This allows the reader of the model to understand that this user's activity is already referenced in another part of the task model. It is important to note that this sub-model structuring can be applied to any node in the task model. If the node is not a leave task, then the entire sub-tree is managed as a copy i.e. modification in one is reflected in all the copies. This issue has been raised in section 3.2 when the CTT model has been presented.

6 Analysis of the Case Study

6.1 Quantitative Analysis

Table 4 presents statistics on the number of model elements needed to describe the controller's activities, using both CTT and HAMSTERS notations (together with the structuring mechanisms proposed). It is important to note that part of the difference between the two representations lies for some aspects in the structuring mechanisms and for other aspects in the notations themselves. However, adding those mechanisms to CTT or to any other notation dedicated to task models would exhibit similar advantages.

The first line of the table (in italics) highlights the fact that the number of models is much higher using the mechanisms than without using them. Then, for each model, the number of nodes, arcs and operators is presented. This is due to the sub-routine extension that promotes splitting a model into several parts. The sub-model extension has no impact on the number of models.

Table 5. Quantitative comparison of the two modeling techniques

	Notations	
Number of elements used for the case study according to the notations:	**CTT**	**HAMSTERS**
Number of models	*4*	*6*
Tasks for representing the main model	6	3
Operators for representing the main model	3	1
Arcs for representing the main model	11	3
Tasks for representing the "Monitoring" task model	9	9
Operators for representing the "Monitoring" task model	7	1
Arcs for representing the "Monitoring" task model	22	9
Tasks for representing "Failure detection and recovery"	74	20
Operators for representing "Failure detection and recovery"	52	7
Arcs for representing "Failure detection and recovery"	175	27
Tasks for "Operate communication frequency"	19	17
Operators for "Operate communication frequency"	11	7
Arcs for "Operate communication frequency"	40	23
Tasks for representing the "Record failure" task model	-	15
Operators for representing the "Record failure" task model	-	4
Arcs for representing the "Record failure" task model	-	18
Tasks for representing the "Reset flight SW" task model	-	14
Operators for representing the "Reset flight SW" task model	-	5
Arcs for representing the "Reset flight SW" task model	-	20
Total number of tasks	**102**	**78**
Total number of operators	**73**	**25**
Total number of arcs	**248**	**100**

The total number of tasks required to model the controller's activities is decreased by about 20% using HAMSTERS notation together with the structuring mechanisms. There is a bigger difference (66% reduction) as far as both the operators and the arcs are concerned. This is due to the fact that operators are considered as part of the node in HAMSTERS and don't need to be duplicated if the temporal operator between several tasks in a row is the same (as this is the case in CTT notation). The light grey lines in the lower part of the table correspond only to models appearing in the HAMSTERS modeling. They have no counterpart in CTT.

6.2 Qualitative Analysis

Qualitative assessment is a more complex task which would require very detailed usability analysis of both the use of the structuring mechanisms and the idiosyncrasies of each notation. This is premature to the work presented here as it would require the distribution of HAMSTERS at a large scale (as this has been done for a long time with CTT) and then the setting up of an ethnographic study. However, even though such experiments were conducted we would face the same difficulties as those encountered in the late 80's when the software engineering community was struggling to assess whether Object-Oriented programming was "better or not" with respect to "structured programming" [37]. Indeed, there are so many parameters involved such as training, maturity of tools, application domain, background of the developers … that a definite result is out of reach. Another related element is the user interface and interaction technique embedded in the tool supporting the notation. Indeed, even CTTE proposes the CTTVis extensions providing zooming and enhanced interaction techniques for task models edition. Such aspects have a he impact on the adoption and performance of user while building task models.

However, the current version of HAMSTERS tool is publicly available at http://www.irit.fr/ICS/hamsters/ and we plan to gather feedback from the future users.

7 Discussion and Conclusions

This paper has presented two new structuring mechanisms for notations dedicated to task modeling. We have shown how these mechanisms are different from the mechanisms currently available in the various existing notations. These mechanisms have been applied on a large case study from the space domain. We have used CTT notation as a reference point to support the reader who might be familiar with this notation and to compare quantitatively the size of the resulting models.

While this paper has focused on the mechanisms and the notation themselves this work belong to a long research program aiming at exploiting in a systematic way models in the design and validation of large (potentially safety critical) interactive systems. Indeed, these mechanisms would be extremely useful for addressing the issue of user errors in tasks models (in order, for instance, to identify scenarios exhibiting user errors) as in [24] or to support training of operators as in [18]. In the context of safety critical systems such elements are of primary importance as the resilience of the entire socio-technical system (including operators, procedures, training and computing system) has to be assessed prior to deployment.

References

1. Amboss,
   http://wwwcs.uni-paderborn.de/cs/ag-szwillus/lehre/ws05_06/PG/PGAMBOSS
2. Barboni, E., Ladry, J.-F., Navarre, D., Palanque, P., Winckler, M.: Beyond modeling: an integrated environment supporting co-execution of tasks and systems models. In: Proc. of the 2nd ACM SIGCHI Symp. on Engineering Interactive Computing Systems (EICS 2010), pp. 165–174 (2010)
3. Breedvelt, I., Paterno, F., Sereriins, C.: Reusable structures in task models. In: Harrison, M.D., Torres, J.C. (eds.) DSVIS 1997, pp. 251–265. Springer, Heidelberg (1997)
4. Card, S., Moran, T., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum, Hillsdale (1983)
5. Chen, P.: The Entity-Relationship Model - Toward a Unified View of Data. ACM Transactions on Database Systems 1(1), 9–36 (1976)
6. Chattratichart, J., Kuljis, J.: Exploring the effect of control-flow and traversal direction on VPL usability for novices. Journal of Visual Languages and Computing 13, 471–500 (2002)
7. Dennis, A., Valacich, J.: Computer Brainstorms: More Heads are Better than One. Journal of Applied Psychology 78(4), 531–537 (1993)
8. Diaper, D., Stanton, N.A.: The Handbook of Task Analysis for Human-Computer Interaction, 650 pages. Lawrence Erlbaum Associates, Mahwah (2004)
9. Diehl, M., Stroebe, W.: Productivity Loss in Brainstorming Groups: Towards a Solution of a Riddle. Journal of Personality and Social Psychology 53(3), 497–509 (1987)
10. Dittmar, A.: More precise descriptions of temporal relations within task models. In: Paternó, F. (ed.) DSV-IS 2000. LNCS, vol. 1946, pp. 151–168. Springer, Heidelberg (2001)
11. European Cooperation for Space Standardization, Space Engineering, Ground Systems and Operations, ECSS-E-70C (July 31, 2008)
12. Gaffar, A., Sinnig, D., Seffah, A., Forbrig, P.: Modeling patterns for task models. In: Proceedings of the 3rd Annual Conference on Task Models and Diagrams (TAMODIA 2004), pp. 99–104. ACM, New York (2004)
13. Hartson, R.H., Gray, P.D.: Temporal aspects of tasks in the User Action Notation in Human Computer Interaction, vol. 7, pp. 1–45. Lawrence Erlbaum Associates, Mahwah (1992)
14. Hopcroft, J., Ullman, J.: Introduction To Automata Theory, Languages, and Computation. Addison-Wesley Longman Publishing Co., Inc., Boston (1990)
15. Johnson, H., Johnson, P.: Task Knowledge Structures: Psychological basis and integration into system design. Acta Psychologica 78, 3–26 (1991)
16. Limbourg, Q., Pribeanu, C., Vanderdonckt, J.: Towards Uniformed Task Models in a Model-Based Approach. In: Johnson, C. (ed.) DSV-IS 2001. LNCS, vol. 2220, pp. 164–182. Springer, Heidelberg (2001)
17. Mavin, A., Maiden, N.A.M.: Determining Socio-Technical Systems Requirements: Experiences with Generating and Walking Through Scenarios. In: Proc. 11th Int. Conf. on Requirements Engineering, pp. 213–222. IEEE Computer Society Press, Los Alamitos (2003)
18. Martinie, C., Palanque, P., Navarre, D., Winckler, M., Poupart, E.: Model-Based Training: An Approach Supporting Operability of Critical Interactive Systems: Application to Satellite Ground Segments. In: Proceedings of ACM SIGCHI EICS 2011. ACM Press, Pisa (2011)

19. Navarre, D., Palanque, P., Bastide, R., Paternó, F., Santoro, C.: A tool suite for integrating task and system models through scenarios. In: Johnson, C. (ed.) DSV-IS 2001. LNCS, vol. 2220, pp. 88–113. Springer, Heidelberg (2001)
20. Navarre, D., Palanque, P., Ladry, J., Barboni, E.: ICOs: A model-based user interface description technique dedicated to interactive systems addressing usability, reliability and scalability. ACM TOCHI 16(4), 1–56 (2009)
21. Navarre, D., Palanque, P., Bastide, R.: A Tool-Supported Design Framework for Safety Critical Interactive Systems. In: Interacting With Computers, vol. 15(3), pp. 309–328. Elsevier, Amsterdam (2003)
22. Norman, D., Drapper, S. (eds.): User Centred System Design. L. Erlbaum, U.S (1986) ISBN-10: 0898597811
23. Petri, C.A.: Kommunikation Mit Automaten. Technical University Darmstadt (1962)
24. Palanque, P., Basnyat, S.: Task Patterns For Taking Into Account In An Efficient and Systematic Way Both Standard And Erroneous User Behaviours. In: IFIP 13.5 Working Conf. on Human Error, Safety and Systems Development (HESSD), pp. 109–130. Kluwer Academic Publisher, Dordrecht (2004)
25. Paternò, F.: CTTE: An Environment for Analysis and Development of Task Models of Cooperative Applications. In: ACM CHI 2001 (Seattle, 2001) Extended Abstracts. ACM Press, New York (2001)
26. Paternò, F.: ConcurTaskTrees: An Engineered Notation for Task Model. In: The Handbook of Task Analysis for Human-Computer Interaction, pp. 483–503. Lawrence Erlbaum Associates, Mahwah (2003)
27. Paternò, F., Mancini, C., Meniconi, S.: ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models. In: Proc. of Interact 1997, pp. 362–369. Chapman & Hall, Boca Raton (1997)
28. Paternò, F., Zini, E.: Applying information visualization techniques to visual representations of task models. In: Proc. of TAMODIA 2004, pp. 105–111. ACM, New York (2004)
29. Rational Software Corporation. UML Notation Guide. 1.1 ed (1997)
30. Alur, R.: Timed Automata. In: Halbwachs, N., Peled, D.A. (eds.) CAV 1999. LNCS, vol. 1633, pp. 8–22. Springer, Heidelberg (1999)
31. Rettig, M.: Prototyping for Tiny Fingers. Communication of the ACM 37(4), 21–27 (1994)
32. Sebillotte, S.: Hierarchical planning as method for task analysis: the example of office task analysis. Behaviour & Information Technology 7(3), 275–293 (1988) 1362-3001
33. Scapin, D.L.: K-MADe. In: COST294-MAUSE 3rd International Workshop, Review, Report and Refine Usability Evaluation Methods (R3 UEMs), Athens (March 5, 2007)
34. Sinnig, D., Wurdel, M., Forbrig, P., Chalin, P., Khendek, F.: Practical Extensions for Task Models. In: Winckler, M., Johnson, H. (eds.) TAMODIA 2007. LNCS, vol. 4849, pp. 42–55. Springer, Heidelberg (2007)
35. van Welie, M., van der Veer, G.C., Eliëns, A.: Euterpe - Tool support for analyzing cooperative environments. In: Ninth European Conf. on Cognitive Ergonomics, pp. 25–30 (August 24-26, 1998)
36. van Welie, M., van der Veer, G.C., Eliëns, A.: An Ontology for Task World Models. In: 5th International Eurographics Workshop on Design Specification and Verification of Interactive Systems DSV-IS 1998, Abingdon, UK, pp. 57–70 (June 3-5, 1998)
37. Wieringa, R.: A survey of structured and object-oriented software specification methods and techniques. ACM Comput. Surv. 30(4), 459–527 (1998)
38. Wood, W.A.: Transition network grammars for natural language analysis. Communications of the ACM 13(10), 591–606 (1970)

User Driven Evolution of User Interface Models – The FLEPR Approach

Stefan Hennig[1], Jan Van den Bergh[2], Kris Luyten[2], and Annerose Braune[1]

[1] Institute of Automation
Technische Universität Dresden
01062 Dresden, Germany
{stefan.hennig,annerose.braune}@tu-dresden.de
[2] Hasselt University – tUL – IBBT
Expertise Centre for Digital Media
Wetenschapspark 2, B-3590 Diepenbeek, Belgium
{jan.vandenbergh,kris.luyten}@uhasselt.be

Abstract. In model-based user interface development, models at different levels of abstraction are used. While ideas may initially only be expressed in more abstract models, modifications and improvements according to user's feedback will likely be made at the concrete level, which may lead to model inconsistencies that need to be fixed in every iteration. Transformations form the bridge between these models. Because one-to-one mappings between models cannot always be defined, these transformations are completely manual or they require manual post-treatment.

We propose interactive but automatic transformations to address the *mapping problem* while still allowing designer's creativity. To manage consistency and semantic correctness within and between models and therefore to foster iterative development processes, we are combining these with techniques to track decisions and modifications and techniques of *intra-* and *inter-model* validation. Our approach has been implemented for abstract and concrete user interface models using Eclipse-based frameworks for model-driven engineering. Our approach and tool support is illustrated by a case study.

Keywords: User interface models, model transformations, interactive model transformations, model consistency, model synchronization.

1 Introduction

Industrial and particularly safety critical systems have to meet extensive requirements since the cost of failure is high and might result in loss of life. Thus, also the *User Interfaces* (UI) to operate such a industrial system need to be clear and without ambiguities with respect to defined behavior of the system. Specifically in the industrial automation domain, Visualization Systems enable factory-trained operators to monitor the operative states of automation systems as well as to operationally intervene in the process.

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 610–627, 2011.

Since these operators should react appropriately to exceptions in the technical process even in extreme situations, they are required to give an early feedback to the system design. Therefore, a User-centered Design (UCD) [17] process is recommended. UCD is an iterative process: Ideas and concepts of the UI will be concretized into prototypes that are tangible for end-users in an early point of time. According to the end-users' feedback after evaluation, the prototype will be improved or modified and thereafter be evaluated again. Fig. 1 gives an overview of this iterative process.

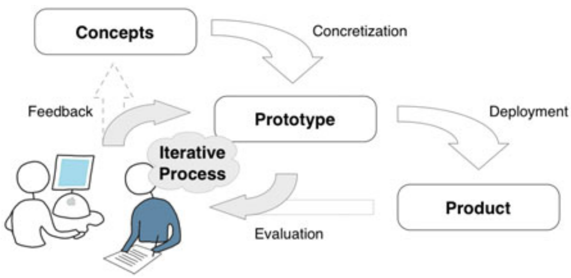


Fig. 1. Overview of the User-centered Design procedure

Moreover, *Model-based User Interface Development* (MBUID) procedures have established to ensure the quality, usability, and sustainability of interactive applications. Since visualization systems are interactive applications [24], it is recommended to apply the methods proposed by MBUID: Basically, it uses models at different levels of abstraction, whereas moving from more abstract models to more concrete models means adding new information related to the UI design of the particular level of abstraction. This concretization is achieved by transformations. These are completely manual or require manual post-treatment because one-to-one mappings between models cannot always be defined.

User-centered design processes are about continuous improvement of user interface prototypes, already in early development stages. In MBUID, these modifications can take place at a concrete level where the UI is getting more concrete. However, altering UI models at the concrete level might not influence the abstract models, which leads to model inconsistencies. Therefore, we can identify two main challenges which arise if we want to combine MBUID with UCD: (1) *Ambiguities during transformation* and (2) *Model Consistency*.

In this paper, we propose a *Flexible Workflow for early User Interface Prototyping* (FLEPR) – an approach to resolve ambiguities during transformation interactively, thus keeping the user in control over design decisions. FLEPR combines interactive transformations with intra- and inter-model validation and with techniques to track decisions and modifications. Therefore, FLEPR combines UCD with MBUID by keeping the models consistent and preserving their semantic correctness at any time. The FLEPR is supported by a proof-of-concept tool. This tool was used in a small project to show the feasibility of this approach.

2 Problem Classification

UCD as an iterative process advocates the use of tangible prototypes in order to involve end-users in the design process as early as possible. These prototypes will be altered according to end-users' feedback after an evaluation — and then be evaluated and be improved again. This is what the term *evolution* implies — modifications to artifacts at all stages of software development [13]. Once the end-users' needs are satisfied, the prototype can be provided with complete functionality and deployed as a product. Fig. 1 illustrates the UCD process. The input of end-users is of great importance to create an interactive system that is well received by and tailored for the target end-user group. However, for industrial automation (and by extension safety critical systems), we need to be wary of inconsistencies between user preferences and required system behavior. After all, a correct operation of the system is more important in this case.

The *Cameleon Reference Framework* [2] defines development steps for *Model-based User Interface Development* (MBUID) by introducing UI models at different levels of abstraction. The *Final UI* level captures UIs ready for execution on a particular platform (see *Product* in Fig. 1). A *Concrete UI* level abstracts the *Final UI* from a particular platform. *Movisa* (see Section 3.2) is our representative for this level. It defines a high-fidelity concrete syntax which is tailored to the needs of automation engineers. Therefore, it corresponds to the prototype level of UCD depicted in Fig. 1 – an evaluation by end-users can be provided. *Abstract UI* definitions form the next level of abstraction proposed by the Cameleon Reference Framework. *CAP3* (see Section 3.1) forms an appropriate realization. It provides a concrete syntax which supports interaction designers in their work (see *Concepts* in Fig. 1).

Transformations are the key to progress from one model to another in MBUID. Three types of transformations have been defined in [30]: (1) *Abstraction* creates a more abstract model from a given model, (2) *reification* concretizes a model, and (3) *translation* produces models at the same level of abstraction. Another taxonomy, introduced by Mens et al. [16], distinguishes in that context *vertical* (abstraction, reification) and *horizontal* (translation) transformations. Furthermore, they define an additional dimension: A transformation is *endogenous* if source and target models are compliant to the same metamodel; it is *exogenous* if they are compliant to different metamodels. In [3], the authors refer to transformations as mappings and explicitly include human intervention during the mapping as a possibility. This has not been fully explored in MBUID literature to the authors' knowledge. For resolving ambiguities while ensuring consistent system behavior, human intervention during transformations is inevitable.

When combining MBUID and UCD for the problem at hand, the starting point is a CAP3 model (see Fig. 2) situated on the abstract UI layer. Using a vertical, exogenous transformation, a CAP3 model will be *refined* into a Movisa model (① in Fig. 2) whereas unambiguous one-to-one mappings cannot be ensured. According to the end-user feedback, the Movisa model will be modified or improved using horizontal, endogenous transformations (② in Fig. 2). In other words, the Movisa model will be *refactored*. Refactoring the Movisa models might lead to inconsistencies with the

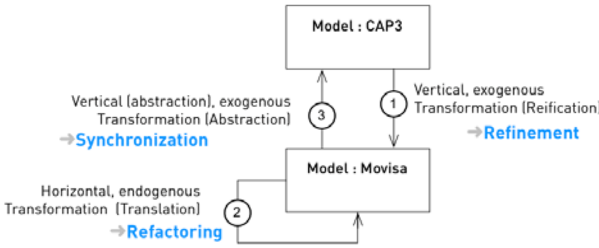


Fig. 2. Overview of the different types of transformation

CAP3 model. Resolving these issues requires *Model Synchronization* (③ in Fig. 2, see also dashed arrow in Fig. 1). According to Ivkovic and Kontogiannis [13], model synchronization is "[...] the process of establishing an equivalence between two models when one of them is altered."

The aim of this paper is to combine the *User Centered Design* process and *Model-based User Interface Development* techniques complying with the *Cameleon Reference Framework*. FLEPR enables a flexible workflow that solves ambiguous mappings during model refinement by means of interactive transformations (① in Fig. 2). It enables incorporating end-user feedback into concrete models and keeps track of these manual model refactorings (② in Fig. 2). As this can cause inconsistencies with abstract models, FLEPR provides model synchronization means (③ in Fig. 2. After explaining the basics of both CAP3 and Movisa, the following sections discuss our FLEPR approach on a conceptual and on a technical basis. A case study proves its feasibility.

3 Background

This section gives a brief introduction of the used modeling concepts *CAP3* (Section 3.1) and *Movisa* (Section 3.2).

3.1 CAP3

CAP3 is a modeling language for abstract user interface models that provides both a concrete syntax (a graphical notation) and an abstract syntax (meta-model). Its concrete syntax is based upon the *Canonical Abstract Prototypes* notation [4] (CAP). CAP provides a number of abstract UI components to describe the structure of a (graphical) user interface in a way that is independent of any concrete toolkit or even modality. There are three main UI components: *tool*, *container* and *active material*. A tool allows a user to trigger a change in the UI or in the functional core (e.g. a button), a *container* can hold data (e.g. a label, image) or any other UI component (e.g. a window), and an active material is a combination of both previous components and thus can both hold data and trigger a change in the UI, data or functional core (e.g. a textfield, the Microsoft Ribbon). There are many other UI components besides these three that have much richer semantics and visuals, but are entirely optional as they are special cases of them.

The difference is illustrated in Fig. 3. Fig. 3a only uses the three basic UI components to specify a login dialog: A user can enter a login and a password as well as confirm or clear this information. Fig. 3b shows the same login dialog, but this time using richer UI components where appropriate. It shows that the login can be selected from a set of available options (using a *selectable collection*) and that the password has to be entered from scratch each time (using an *input*). The confirmation ends the activities in this dialog (using an *end component*) and visually confirms that *Clear* removes the data in the interface. The dashed rectangle (conceptual *group*) around the login and password show that both UI components belong together.

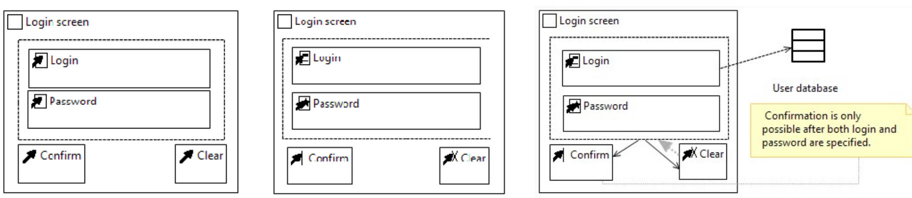


Fig. 3. CAP3 login window using (a) the basic CAP UI components; (b) richer CAP UI components suggesting possible options, and (c) UI components specifying behavior and relations to the domain model

Both versions of the login dialog, however, do not specify the relations between components. For example, they do not specify that confirmation can only be done when both login and password are specified or that the *clear* tool clears both the login and password. This behavior can be specified by using the additions made in CAP3; it adds the capability to specify behavior and relations to the domain model or functional core, as can be seen in Fig. 3c. It specifies that *Confirm* and *Clear* are only available after the login and password are specified using the *enable* relation. The fact that *Clear* clears both the login and password is shown using the *update* relation (thick gray dotted arrow). Fig. 3c also shows that the logins are fetched from the *user database*. Note that the *conceptual group* is used to reduce the number of arrows; all relations that are connected to a *conceptual group* can be replaced by relations to all its contained components.

3.2 Movisa

Movisa is a *Domain Specific Language* (DSL) to create models for technology independent development of interactive systems in industrial automation that require complex graphical representations (e.g. visualization of a plant process). This domain has certain requirements for the UI components, the behavior of those components, and the protocols to gather the data that influence this behavior. A Movisa compliant model contains three sub-models (see Fig. 4): The *Presentation Model*, the *Algorithm Model*, and the *Client Data Model*.

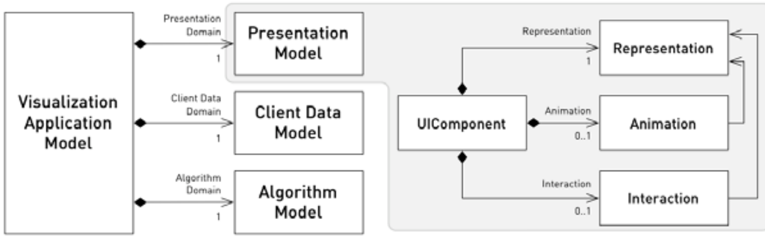


Fig. 4. Excerpt of Movisa's metamodel, the relevant part for this paper is the *Presentation Model*, which defines UI components

As shown in Fig. 4, the *Presentation Model* defines *UI Components* whose properties where classified in the following three categories: (1) The *Representation* category defines the initial appearance of a UI component on the screen. It encapsulates representations for properties such as *Position*, *Size*, or *Border Color*. (2) An *Animation* category captures properties defining the dynamic behavior of UI components in order to reflect the current state of the process. Thus, properties of the *Animation* influence properties of the *Representation* during runtime. (3) The *Interaction* category allows to configure the user's interaction means as well as the resulting actions. Based on a survey of conventional visualization systems for industrial automation, we deducted that these properties – and the relations between them (see Fig. 4) – meet the requirements of *industrial automation*.

Conventional visualization systems provide scripting environments in order to enable developers to integrate application specific logic. This is what the *Algorithm Model* provides in a technology independent manner using a customized *Executable UML* realization. *Executable UML* is a computationally complete abstract software specification or modeling language.

Movisa's *Client Data Model* contains modeling elements dedicated to configure communication parameters. Basically, these parameters depend on concrete data server specifications. The *Client Data Model* therefore provides an adapter-like architecture: An adapter for each data server specification allows fine-grained parameterization so that reliable communication relationships with any automation specific data server can be ensured. A common information model provides these data to elements of the *Algorithm Model* and to elements of the *Presentation Model*.

To illustrate Movisa's principles, we use the CAP3 model introduced in Section 3.1 and realize the *Selectable Collection* "Login" exemplarily on the concrete UI level. (This is the level of the *Cameleon Reference Framework* where Movisa resides; it has been discussed in Section 2.) Therefore, the *Presentation Model* defines a UI component *Drop Down*. This component has an appropriate *Animation Property* that is responsible to fill up the *Drop Down* component with available options from the "User database". An additional requirement may be not to show up the user name "Administrator". This can be realized with an appropriate set of actions — defined within the *Algorithm Model* — dedicated to remove this particular user name from the data set. Finally, the *Client Data Model* defines the required data items and the particular data server from which the data will be fetched.

4 FLEPR: UCD Meets MBUID

In order to define an effective approach for the problem described in Section 2, we consider the following three perspectives: (1) A *User Perspective* examines the different users and roles involved in the UI development process; (2) a *Conceptual Perspective* derives an appropriate concept from the tasks of the user perspective; and (3) a *Technology Perspective* introduces the concrete technical realization of the derived concepts.

4.1 User Perspective

Users involved in the UI development process are characterized by different responsibilities and tasks, different knowledge and different skills. They are therefore assigned one-time to one of the following user roles:

Interaction Designer/Information Architect is familiar with the standards concerning usability (e.g. [8]), dialog design (e.g. [6]), and/or information representation (e.g. [7]) – with respect to safety critical systems. All of these can be described as common knowledge of the interface design; it is captured by conceptual models such as CAP3. This user role is therefore responsible for creating and maintaining the abstract UI model CAP3.

Domain Expert is an Application Engineer. She or he has advanced knowledge about the technical process to be monitored and operated by the visualization system. Furthermore, he or she is acquainted with the devices realizing the technical process and the relationships between them. Domain experts are additionally characterized by knowing the physical backgrounds, possible events and hazardous situations as well as the details on the product to be manufactured (e.g. product compounds). Movisa has been designed to capture specific knowledge of the automation domain. As this user role specifies these implementation details, it is mainly responsible for creating and maintaining the concrete UI model Movisa.

End User knows about the (manufacturing) process and the product to be manufactured. She or he also has knowledge of how to react to particular events. They are intended to work with the *product* of the user-centered design process. Therefore, users of this role are responsible for evaluating the UI prototypes.

Fig. 5 shows these user roles assigned to the particular development step (see Fig. 2) which the users are mainly responsible for.

There is, however, no clear border between the responsibilities of the participating user roles. Interaction designers might collaborate with domain experts at the CAP3 level. Domain experts might collaborate with interaction designers at the Movisa level. In any case, the users have deep knowledge about there belonging domain, but they are neither familiar with model transformations nor with details of the models itself. Therefore, they have to be supported appropriately by dedicated tools.

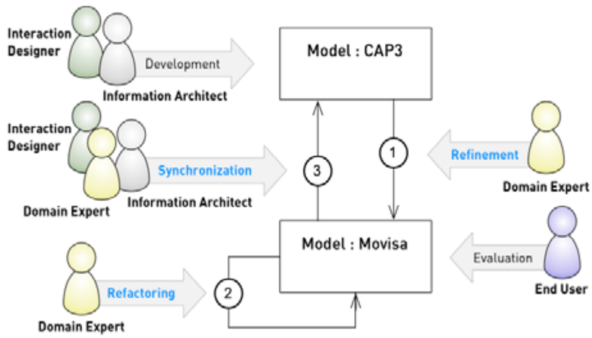


Fig. 5. Different types of users are involved in the UI development process at different development steps

4.2 Conceptual Perspective

The basis we use for this Section is a CAP3 model initially developed by the responsible users (see Section 4.1). Starting from that, Fig. 2 points out the different kinds of transformation while Fig. 5 introduces particular user roles mainly responsible for each part together with their particular task. Based on this, Fig. 6 introduces the overall concept which is composed by the following steps: ① *model refinement*, ② *model refactoring*, and ③ *model synchronization*.

Model Refinement(①). This is the development step which transforms a source model (CAP3) into a target model (Movisa) using a vertical, exogenous transformation. In this development step, an unambiguous one-to-one mapping between elements of the CAP3 model and elements of the Movisa model cannot be ensured. The following list presents several fundamental ambiguities we identified:

- ❶ CAP3's *Selectable Collection*[1] can be mapped to one of the following Movisa elements: (1) *Radio Button Group*, (2) *Check Box Array*, or (3) *Drop Down*.
- ❷ A CAP3 *Tool* can be mapped to a Movisa *Button* or to any other Movisa UI component with an appropriate *interaction property*.
- ❸ CAP3 defines an element *Notification* which can be transformed to Movisa's *AlarmWidget* or any other Movisa UI component with an appropriate *animation property*.
- ❹ CAP3 introduces an Update relationship, however it makes no assertions which property of the particular Movisa UI component should be updated under which condition.

Clearly, these ambiguities cause problems because one element can be mapped to multiple alternatives. Decisions regarding these mappings are always subject to project specific requirements or assumptions, company guidelines, or even personal

[1] A *Selectable Collection* can also be represented by a *Collection* and a separate (nested) *Selection Tool*.

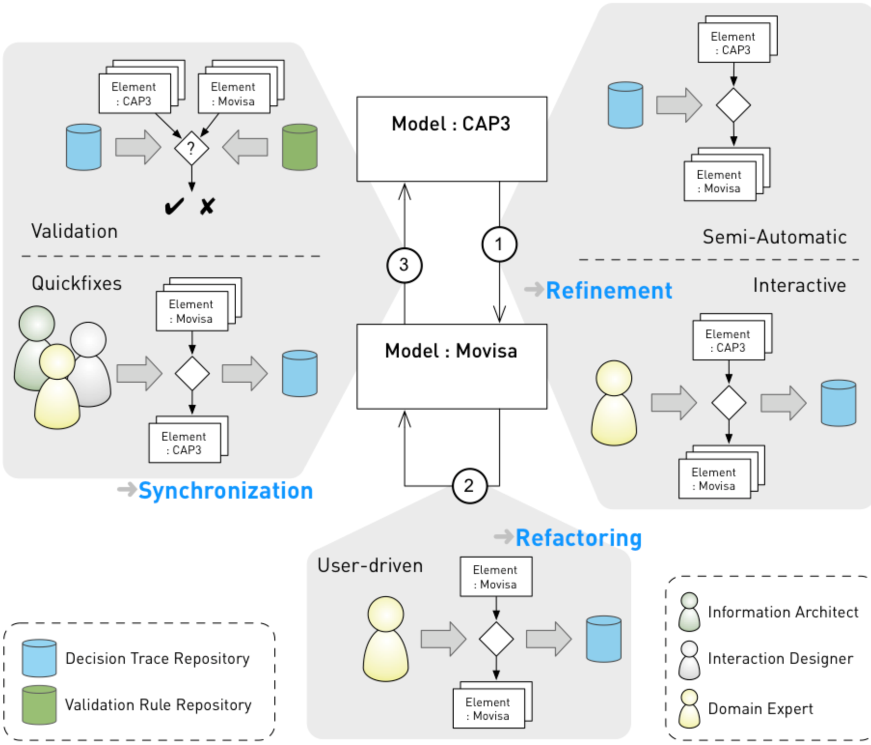


Fig. 6. The overall concept showing details about each development step – ① Model Refinement, ② Model Refactoring, and ③ Model Synchronization – using the trans-formations defined in Fig. 2. It also shows in what extend the users introduced in Fig. 5 are involved in the procedure.

taste. In the best case, there is only a single one-to-one mapping possible avoiding all ambiguities. Nevertheless, these decisions influence the transformation process — it needs to be customized based on them. Since Domain Experts, as the users that steer this development step, are not familiar with transformations (see Section 4.1), the following four different approaches can increase accessibility of these transformations:

1. A default mapping is used within an automatic transformation process. The domain expert then needs to customize the resulting Movisa model subsequently according to project specific requirements.
2. A separate *Mapping Model* can be used which captures the project specific mappings between elements. This concept has been introduced by the *Human Computer Interaction* (HCI) community, e.g. in [3]. Sottet et al. [26] have defined a metamodel to specify mappings between different kinds of models. Elements of a source model can be connected to elements of a target model on both the metamodel and the model level. UsiXML [29] defines a mapping model, too, which only provides mappings on the model level. According to Siikarla and Systa, transformations relying on this additional information are called *semi-automatic* since they require “the transformation engineer [to make] decisions that guide an [...] ambiguous automatic transformation” [25].

3. The source model (CAP3) can be annotated in order to remove ambiguities in a semi-automatic transformation. For example, the *Eclipse Modeling Framework* (EMF) [27] provides an*EAnnotation* element which enriches arbitrary modeling elements with additional information. XML introduces a similar mechanism with the <annotation> element. In both cases, such mechanism can be used to specify mappings between elements of two models.
4. In contrast to the previous two approaches which require user interventions before the actual transformation, *interactive transformations* require the user to intervene only during the transformation process.

We argue that the last (4) approach is most suitable, despite the others being used more often in existing solutions. Defining mappings on metamodel as well as on model level (2) or annotating a model (3) requires deep knowledge of the model's abstract syntax. It cannot be assumed that the responsible user roles have this knowledge (see Section 4.1). Interactive transformations (4), however, do not expect users to have deep knowledge neither of abstract syntax nor of the transformation itself. While the user will guide the transformations, the transformations also have to guide the user; the more precise the descriptions of ambiguities are the more powerful is the procedure.

Model Refactoring. This development step uses a horizontal, endogenous transformation to modify or improve a model: Once the Movisa model has been derived by transformation, a manual post-treatment is necessary because information e.g. regarding an element's position or size are not part of the CAP3 model. Moreover, the role *end-user* comes into play: He or she evaluates the actual design and according to this she or he gives her or his feedback to the domain expert. It entails the following tasks:

1. Improvements such as element's size or border color can simply be made using Movisa's model editor.
2. Modifications such as changing a *Radio Button Group* to a *Drop Down* element while preserving the functional behavior is a more sophisticated task. It can be done using appropriate transformations. Kolovos et al. define them as *update transformations in the small*: They "are applied in a user-driven manner on model elements that have been explicitly selected by the user" [15].

Model Synchronization. This is the process to manage consistency between two or more models if one of these models has been altered: After the Movisa model has been modified due to end-user feedback, the CAP3 model has to reflect these modifications. As stated in Section 2, there are two classes of modifications, those which do not entail updates in the CAP3 model and those which do. For example, changing only the border color of a particular UI element in the Movisa model has no effect to the CAP3 model. Adding a further item to a *Drop Down* box requires an update of the CAP3 model. Otherwise, the consistency is violated.

According to Nuseibeh et al. [18], inconsistencies can be tolerated as long as they will be fixed in a future point in time. Therefore, we do not artificially constrain the domain expert in improving a model and treat consistency management in a subsequent validation procedure which is explicitly driven by the user (see ③, Fig. 5). Nuseibeh et al. [18] describe model consistency management basically with the tasks (1) detecting inconsistencies, (2) characterizing inconsistencies, and (3) handling inconsistencies. Following that scheme, our approach for model synchronization is a two-staged process:

Intra- and inter-model validation: The models are consistent if both syntactical and semantical correctness is guaranteed. While syntactical correctness appears if the model is compliant to its metamodel, the process of ensuring semantical correctness is subject to a check against a set of appropriate rules. Basically, these rules define to what extent the Movisa model is allowed to alter to be consistent to the CAP3 model. Additionally, semantical correctness has to be proven not only within one model but also between them.

Solving inconsistencies: If inconsistencies were identified by the preceding validation, members of the particular user role (see ③, Fig. 5) decide whether to roll back the corresponding modifications or to update the counterpart model. The latter option will be supported by user-driven update transformations. The system therefore suggests valid alternatives.

Traceability. Inter-model validation, as previously introduced, needs knowledge about which elements of both models are interrelated. It is desirable to be able to reproduce all previous decisions since models may contain hundreds of elements and therefore require hundreds of decisions. Thus, a transformation (①, Fig. 5) performed for a second time – after all other transformations (①, ②, and ③, Fig. 5) are executed – should automatically produce the model which was the result of a previously performed transformation (②, Fig. 5).

Czarnecki and Helsen [5] point out that "[t]ransformations may record links between their source and target elements" [5] which fosters synchronization between models. It also enables reproducibility of transformation results. They prefer to store these links separately which has been consolidated by Van Gorp who stated in [11]: "[T]raceability links should be treated as first class software artifacts."

4.3 Technology Perspective

Fig. 6 shows that the *Decision Trace Repository* is the core of the entire approach. It feeds the refinement transformation (①) in order to make all steps reproducible. It also supports the process of model synchronization (③) by providing information about how the elements of the CAP3 model are linked to the elements of the Movisa model. For that purpose, a model dedicated to capture these relationships should be provided separately, as [5] and [11] suggest. Czarnecki and Helsen [5] propose to add an additional GUID[2] to each model element (in our case elements of both the CAP3 and the Movisa model). The third model responsible for capturing the relationships between these models only establishes links between these GUIDs.

Fig. 7 shows the metamodel of our *User Decision Repository*, further on named *FleprMap*. As Section 4.2 claims, the *FleprMap* only establishes a relationship between

---

[2] *Globally Unique Identifier* (GUID).

source and target elements. Therefore, a particular *Mapping* points to the respective elements with the *Source* and *Target* references. These references can be of any type as long as they are derived from an *Eclass*[3].

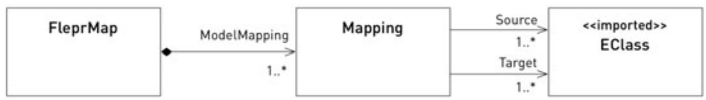


Fig. 7. Simplified metamodel of the FleprMap, our *Decision Trace Repository*

Transformations guide the different users through the entire user interface development process. Mens et al. [16] distinguish mainly *declarative* and *operational* transformation mechanisms. While declarative approaches have their strengths in compactness and maintainability because they hide procedural details, operational approaches may unfold their advantages if incremental model updates are required [16]. Nevertheless, a suitable transformation approach has to deal with the following model management tasks (see Section 4.2): (1) transformation, (2) comparison, (3) validation, and (4) merging. There are different types of languages and transformation approaches each addressing only some of these tasks. For example, OCL[4] [21] is only for model validation; *openArchitectureWare* [20] captures besides validation only transformation. With its set of dedicated languages, the model management tool *Epsilon* [10,14] addresses all of the tasks. Furthermore, the Epsilon languages combine declarative and operational approaches. A further advantage is that it is tightly integrated into the *Eclipse* [9] platform, since both the CAP3 and the Movisa editors are distributed as *Eclipse* plugins. Epsilon's transformations also allow for user interaction. We used (1) the *Epsilon Transformation Language* (ETL) to realize the interactive transformations (derived in Section 4.2) to be applied during model refinement (see ①, Fig. 2), whereas the concrete mapping options are implicitly contained in the transformations; (2) the *Epsilon Wizard Language* (EWL) to foster user-driven model refactoring (see ②, Fig. 2) which has been introduced in Section 4.2; and (3) the *Epsilon Validation Language* (EVL) to apply a set of predefined validation rules in order to achieve model synchronization (see ③ Fig. 2) mentioned in Section 4.2.

Finally, the result of our investigation is a prototypical implementation of the aforementioned concepts (see 4.2) in the form of suitable Eclipse plugins realized with the technology introduced in this Section.

5 Case Study

Our laboratory plant is equipped with three tanks and pipes, pumps, or valves, respectively, between them. Fuel level sensors ensure that the water in the tank does not spill over; a programmable controller enables monitoring and operating. The requirement is to provide a user interface for observing the automatic process as well as enabling manual operator intervention.

[3] *EClass*: Metamodel entity of the *Eclipse Modeling Framework*.
[4] Object Constraint Language (OCL).

Fig. 8 depicts the *Abstract UI* modeled in CAP3 using the CAP3 model editor. The UI contains four *Container* elements: (1) a top level container "Overview" contains all other UI elements, (2) an "Interaction" container captures interaction elements, (3) the container "Fuel level map" defines elements for monitoring the process, and (4) a *Repeated Conceptual Group* "Fuel Level Trends" providing appropriate trend charts. CAP3's *Domain Objects* represent process variables.

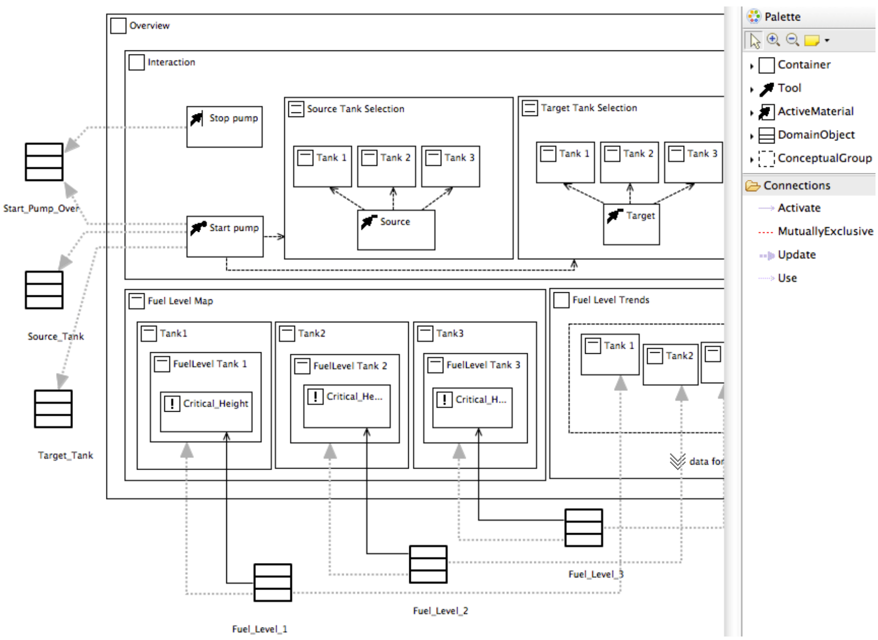


Fig. 8. Excerpt of the CAP3 model created in its dedicated model editor

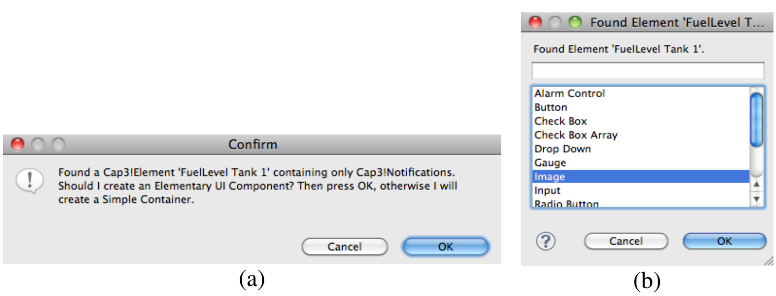


Fig. 9. Model Refinement: (a) notifies the user about ambiguities to be resolved; (b) suggests possible options

During a refinement transformation (see ①, Fig. 6) a first ambiguity to be resolved appears when detecting the *Element* "FuelLevel Tank 1" which is connected to a *Domain Object* via an *Update* relationship. The domain expert therefore has to decide which Movisa element is suitable to monitor a process variable and in which way it can be animated according to its current value. Fig. 9 illustrates this interactive task where the user chooses to create an *Image* element. The transformations store these decisions in the FleprMap, the mapping model. (If the user performs this transformation again, the transformations recognize that both models exist and that the links stored within the FleprMap are valid – in that case no user interaction is required.)

Once all ambiguities are resolved, a complete and valid Movisa model is available. Fig. 10a shows the entire model using Epsilon's model tree view. The "Interaction" container, e.g., has been transformed to a *Simple Container* containing two *Buttons* in order to "Start" and "Stop" an inter-tank transfer process and two *Drop Down* elements dedicated to choose the source and the target tank. Fig. 10b shows the *Drop Down* elements in the Movisa model editor.

An evaluation points out that *Drop Down* elements are not suitable. Therefore, the domain expert invokes an appropriate EWL script directly within the model editor (see Fig.10b) — this invokes a refactoring transformation. First, the user chooses the element to be created using the same procedure as provided for the model refinement process (see Fig. 9b). A transformation (1) creates this new model element, (2) adds it to the right position within the Movisa model, (3) ports all properties of the old element to the new one, (4) deletes the old element, and (5) updates the FleprMap.

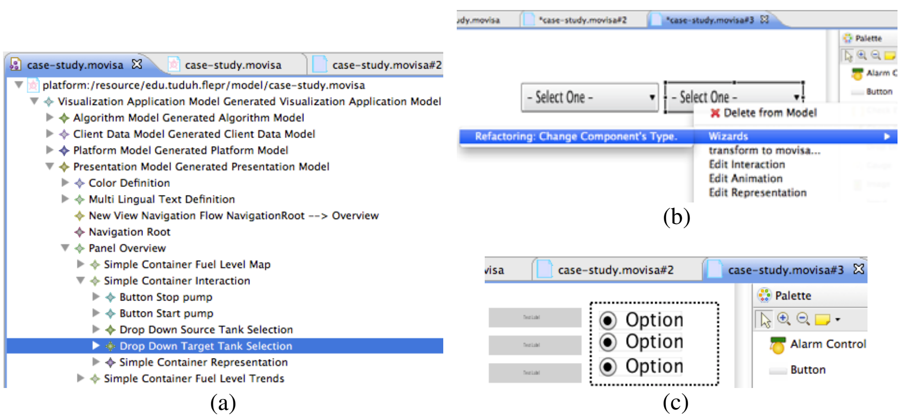
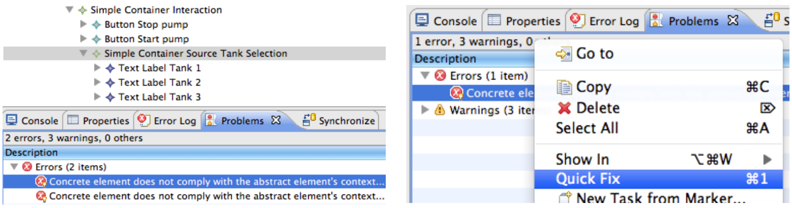


Fig. 10. *Model Refactoring*: (a) shows a compact tree view of the Movisa model after the refinement process; (b) shows an excerpt of Movisa's model editor where a user starts to refactor it; (c) depicts the result of the refactoring process

Fig. 10c shows the resulting interaction elements. The first *Drop Down* element was refactored to a set of *Text Labels*, the second one to a *Radio Button Group*. A subsequent model validation, as part of a *Model Synchronization* process (see ③, Fig. 6), points out that the *Text Label* alternative has led to semantical incorrectness (see Fig. 11a). The reason is that the CAP3 model defines the *Tool* "Start pump" to initialize an inter-tank transfer process as an atomic task – the end-user has to set source and target tank before the particular process variables will be updated[5] using the *Tool*. Thanks to the FleprMap and to the validation rules, expressed in Epsilon EVL, the individual inconsistency can be resolved (see Fig. 11b) using the same procedure as provided for the model refinement process – clarify ambiguities interactively.



(a) (b)

Fig. 11. Problems View of the Eclipse tool: The validator has recognized semantical incorrectness between the CAP3 model and the altered Movisa model (a) and provides means to fix it (b)

The first step after selecting the *Quick Fix* (see Fig. 11b) is deciding whether to roll back the recent modifications or to update the CAP3 model. In this case study, the user chooses the latter option. Fig. 12 highlights the relevant parts of the updated CAP3 model introduced in Fig. 8.

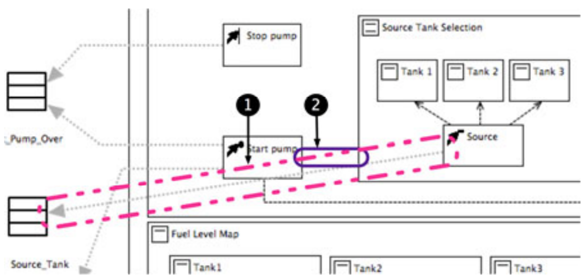


Fig. 12. Excerpt of the CAP3 model after Model Synchronizing

Since the "Source Tank Selection" is now realized by a set of *Text Labels*, the user decides to let the particular *Text Label* update the process variable immediately after clicking it. (It does not need to be an atomic task.) This can be achieved using an appropriate interaction property defined in Movisa's metamodel (see Section 3.2). A set of transformations update both the Movisa model by adding interaction properties to the *Text Label* components and the CAP3 model by moving the origin of the *Update* relationship from the Tool "Start pump" to the Tool "Source" (shape ❷ in Fig. 12). Since the Tool "Start pump" does not depend on the "Source Tank

[5] In industrial automation intervention in the process often requires to modify several process variables as an atomic task.

Selection" anymore, the transformations remove the *Use* relationship between the "Start Pump" *Tool* and the *Selectable Collection* (see shape ❶ in Fig. 12 compared to the same area of Fig. 8). Finally, these transformations update the FleprMap according to the latest modifications in the CAP3 model. All three models – the CAP3 model, the Movisa model, and the FleprMap – are consistent and semantically correct; together they describe a *User Interface* which is appreciated by end-users.

6 Related Work

Model-based UI development is most completely addressed by UsiXML [29], a sophisticated *XML-based User Interface Description Language*. It fosters the MBUID process at each level of abstraction by defining appropriate models and providing tools to reify abstract models. Moreover, the tool *ReversiXML*, proposed by Vanderdonckt et al. [31], can automatically reverse engineer web pages into UsiXML's CUI and AUI models. Another tool for *UI Reverse Engineering* is *VAQUITA* proposed by Bouillon et al. [1]. With it, one can create an XIML [22] *Presentation Model* from any web site. For that purpose, VAQUITA lets developers participate in the engineering process. The purpose of both tools is the migration of an existing UI to another context of use. Beyond that, Stroulia et al. [28] as well as Ramón et al. [23] use reverse engineering methods to extract the knowledge included in legacy UIs into models for further treatment. The presented approaches have in common, that they always create the more abstract UI model from scratch. However, aspects of *Model Synchronization* are desirable, too: Information that is encapsulated in abstract models but not part of concrete models needs to be preserved.

Model synchronization, also referred to as *Model Inconsistency Management*, is addressed by *xlinkit* [19] which uses *XLink* [32] to establish relations between elements of XML based models. *CAViT* [12] is a consistency maintenance framework that binds particular transformation rules to OCL invariants. OCL checks if consistency constraints are satisfied; the particular transformation rule can resolve a possibly identified inconsistency. While the former approach requires XML based models, the latter one relies on OCL. Both prerequisites are not addressed by Eclipse based approaches.

7 Conclusion and Future Work

We investigated to what extend the User-centered Design process can be combined with the Model-based User Interface Development process. A prototype Eclipse implementation using different types of (interactive) transformation and intra- and inter-model validation in combination with a model dedicated to trace decisions, proves our approach to be feasible. It provides an iterative development procedure where modifications can take place at each level of abstraction defined by MBUID. A subsequent model synchronization ensures consistency during this process. Since the methods introduced in this paper enable a forward engineering as well as a reverse engineering, our approach provides a *Round Trip Engineering* in MBUID. While Clerckx et al. [3] state that "[t]he aggregation of all of the abstract and concrete models is called the interface model", we have also to assign our *FleprMap* model to be that part of the *interface model* which ensures consistency.

Future work should be dedicated in a first instance to the usability during the development process. For example, the transformation environment should highlight the particular element if ambiguities have been identified, so that users can reason about the element's context in order to make the right decision. Future work should also be dedicated to maintainability of the transformation environment. The validation rules, for example, could be made explicit with the following consequences: (1) The transformation environment will be more generic and (2) the validation rules can simply be enhanced without knowledge of the transformation itself.

Acknowledgements. The authors would like to thank *Henning Hager* and *Pieter Van Gorp* for their contribution to this paper in the form of fruitful discussions. Part of this work is supported by the FWO project G.0296.08 and by the ESF project 080904707 (together with the Free State of Saxony).

References

1. Bouillon, L., Vanderdonckt, J., Chow, K.C.: Flexible re-engineering of web sites. In: Proceedings of the 9th International Conference on Intelligent User Interfaces (2004)
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 (2003)
3. Clerckx, T., Luyten, K., Coninx, K.: The mapping problem back and forth: customizing dynamic models while preserving consistency. In: Proceedings of the 3rd Annual Conference on Task Models and Diagrams (2004)
4. Constantine, L.L.: Canonical abstract prototypes for abstract visual and interaction design. In: Jorge, J.A., Jardim Nunes, N., Falcão e Cunha, J. (eds.) DSV-IS 2003. LNCS, vol. 2844, pp. 1–15. Springer, Heidelberg (2003)
5. Czarnecki, K., Helsen, S.: Classification of model transformation approaches. In: OOPSLA 2003 Workshop on Generative Techniques in the Context of Model-Driven Architecture (2003)
6. EN ISO 9241-10: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 10: Dialog principles (1996)
7. EN ISO 9241-8: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 8: Requirements for displayed colours (1998)
8. EN ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability (1999)
9. Eclipse Website, `http://www.eclipse.org`
10. Epsilon Project Website, `http://www.eclipse.org/gmt/epsilon`
11. Van Gorp, P., Altheide, F., Janssens, D.: Traceability and Fine-Grained constraints in interactive inconsistency management. In: Neple, T., Oldevik, J., Aagedal, J. (eds.) Second ECMDA Traceability Workshop, ECMDA-TR 2006 (2006)
12. Van Gorp, P., Janssens, D.: CAViT: a consistency maintenance framework based on visual model transformation and transformation contracts. In: Transformation Techniques in Software Engineering (2005)

13. Ivkovic, I., Kontogiannis, K.: Tracing evolution changes of software artifacts through model synchronization. In: Proceedings of the 20th IEEE International Conference on Software Maintenance (2004)
14. Kolovos, D.S., Paige, R.F., Polack, F.A.C.: Epsilon development tools. In: Eclipse Summit 2006 (2006)
15. Kolovos, D.S., Paige, R.F., Polack, F.A.C., Rose, L.M.: Update transformations in the small with the epsilon wizard language. Journal of Object Technology (2007)
16. Mens, T., Czarnecki, K., Van Gorp, P.: A taxonomy of model transformation. Electronic Notes in Theoretical Computer Science (2006)
17. Norman, D.A., Draper, S.W.: User Centered System Design: New Perspectives on Human-computer Interaction (1986)
18. Nuseibeh, B., Easterbrook, S., Russo, A.: Making inconsistency respectable in software development. Journal of Systems and Software (2001)
19. Nuseibeh, B., Kramer, J., Finkelstein, A.: ViewPoints: meaningful relationships are difficult. In: Proceedings of the 25th International Conference on Software Engineering (2003)
20. openArchitectureWare Project Website, `http://www.eclipse.org/workinggroups/oaw`
21. Object Constraint Language, `http://www.omg.org/spec/OCL/`
22. Puerta, A., Eisenstein, J.: XIML: a common representation for interaction data. In: Proceedings of the 7th International Conference on Intelligent User Interfaces (2002)
23. Ramón, Ó.S., Cuadrado, J.S., Molina, J.G.: Model-driven reverse engineering of legacy graphical user interfaces. In: Proceedings of the IEEE/ACM International Conference on Automated Software Engineering (2010)
24. Sheridan, T.B.: Supervisory control. In: Handbook of Human Factors (1987)
25. Siikarla, M.P., Systa, T.J.: Decision reuse in an interactive model transformation. In: Proceedings of the 2008 12th European Conference on Software Maintenance and Reengineering (2008)
26. Sottet, J.-S., Calvary, G., Favre, J.-M.: Mapping Model: A First Step to Ensure Usability for Sustaining User Interface Plasticity. In: Model Driven Development of Advanced User Interfaces, MDDAUI 2006 (2006)
27. Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF Eclipse Modeling Framework, 2nd edn. Addison-Wesley, Reading (2008)
28. Stroulia, E., El-Ramly, M., Iglinski, P., Sorenson, P.: User interface reverse engineering in support of interface migration to the web. Automated Software Engg. (2003)
29. UsiXML V1.8 Reference Manual (February 2007)
30. Vanderdonckt, J.: A MDA-Compliant environment for developing user interfaces of information systems. In: Advanced Information Systems Engineering (2005)
31. Vanderdonckt, J., Limbourg, Q., Michotte, B., Bouillon, L., Trevisan, D., Florins, M.: UsiXML: a user interface description language for specifying multimodal user interfaces. In: Proceedings of W3C Workshop on Multimodal Interaction WMI 2004 (2004)
32. XML Linking Language (XLink) Version 1.1, `http://www.w3.org/TR/xlink11/`

Adapting Desktop Web Pages for Vocal Browsing

Fabio Paternò and Christian Sisti

CNR-ISTI, HIIS Laboratory,
Via Moruzzi 1, 56124 Pisa, Italy
{fabio.paterno,christian.sisti}@isti.cnr.it

Abstract. In this paper we describe a solution to make Web pages more suitable for vocal browsing by analyzing and modifying their logical structure. The solution exploits intermediate logical descriptions that are automatically created by reverse engineering techniques. The adaptation engine aims to identify the main logical structure of the Web page components and remove the aspects specific to the graphical modality. Then, a vocal implementation is generated to support browsing, which begins by the user's selecting from the main components.

Keywords: Adaptation, Web sites, Vocal Browsing, User Interface Models, Accessibility.

1 Introduction

Emerging ubiquitous environments call for supporting access through a variety of interactive devices. Vocal technologies and related applications are steadily improving and this makes possible services such as vocal searches and map navigation by Google or others.

Web applications are widely used and a number of approaches and techniques have been proposed to support their adaptation to mobile devices. Little attention, however, has been paid so far on how to adapt Web applications for vocal access. This seems useful and important, given the recent improvements in vocal technologies, which can allow users to access their applications when the visual channel is busy (e.g. when driving) or when they are visually-impaired. The assistive technology for blind people (mainly screen readers) have a number of usability limitations because they usually access the page implementation and require keyboard selection of the elements of interest, with little ability to filter content; only type filtering is currently available (e.g. listing links, headers, or other types). Thus, often users have to listen to a lot of useless content before reaching the point of interest or are provided with information that is difficult to interpret because they do not know the surrounding context in the Web page.

A work that has similar goals is DANTE [3], which provides semantic transcoding for visually-impaired users. For this purpose it uses ontology-based annotations that are manually created and included in the Web page. Such annotations are then used

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 628–635, 2011.

for automatically transforming the original pages into pages more easily accessible for the visually-impaired. In our case we aim to obtain a tool that is able to automatically analyze the content of Web pages and transform them into vocal applications that are sufficiently usable, with the possibility to customize some adaptation parameters. For this purpose we consider intermediate model-based representations [1], which aim to represent the logical structure of the user interface and enable relevant reasoning to derive a more suitable structure for the vocal channel. While previous work (e.g. [5]) has considered the use of model-based techniques in the authoring of vocal or multimodal applications, we exploit them in the graphical-to-vocal adaptation obtained by transforming graphical Web pages in order to make their content browsable through (*menu-based*) vocal interaction.

2 The Proposed Approach to Vocal Adaptation

We exploit a model-based language [4] for performing a more semantically-oriented transformation. The MARIA framework provides abstract (independent of the interaction modality) and concrete (dependent on the interaction modality but independent of the implementation) languages. Such languages share the same structure with different levels of refinements. A user interface is composed of one or multiple *presentations*. While in graphical interfaces the concept of presentation can be easily mapped onto that of a set of user interface elements perceivable at a given time (e.g. a page in the Web context), in the case of a vocal interface we consider a presentation as a set of communications between the vocal device and the user that can be considered as a logical unit. One example is the case in which the dialogue between the user and the vocal application aims to provide some pieces of information to fill in a form regarding the user. The user interface elements are called *interactors.* Examples of interactors are *navigators* (allow users to move from one presentation to another) and *description* (allow provide output to the user). Each presentation is structured using *interactor compositions* such as *groupings* and *relations.* The grouping operators identify the main logical parts of the interface and contain both interactors and interactor compositions (such as groupings itself). The relation operator defines a kind of relation between two groups of elements, typical example is a form with one group for editing input values and one for controlling them (clearing, sending to the server, …).

Our solution is based on an **adaptation server**, which provides a number of functionalities (see Figure 1). Firstly the *Reverser* automatically parses the content of the Web page and the associated style sheets, and builds a corresponding concrete logical description. Secondly the *Graphical-to-Vocal Adapter* transform the graphical concrete description into a vocal one, which is optimized for vocal browsing. Lastly the *VoiceXML Generator* produce the final vocal interface so that the final result can be loaded onto a vocal browser for execution.

The approach has addressed various issues: *content*, some graphical content is badly rendered on vocal interfaces; *structure***,** the structure of a graphical page is

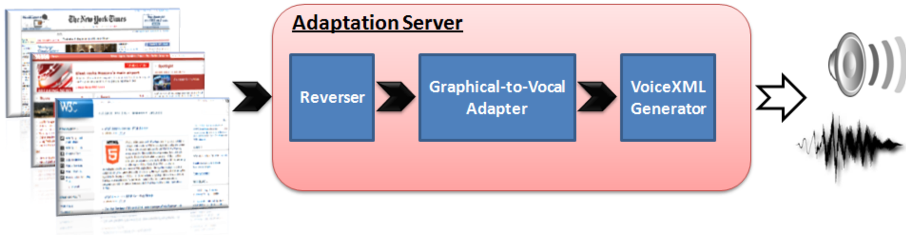


Fig. 1. The global adaptation architecture

typically deeply nested and not suitable for vocal menus; *menu items*, it is crucial to find names for the activation of the menu items that allow users to understand what the corresponding content will be if they are selected.

3 An Example

In order to illustrate how our approach works, we consider an example of a widely known Web site: the W3C. Figure 2 shows an example of the adaptation process, which generates two hierarchical levels. The screen dump (left) has been annotated with solid lines representing the higher sections while the dotted lines represent the nested sub-sections. In addition, we have added the text labels corresponding to the items of the corresponding vocal menus. Figure 2, right, shows the vocal menu structure of the example. Intermediate nodes are menus while leafs represent content.

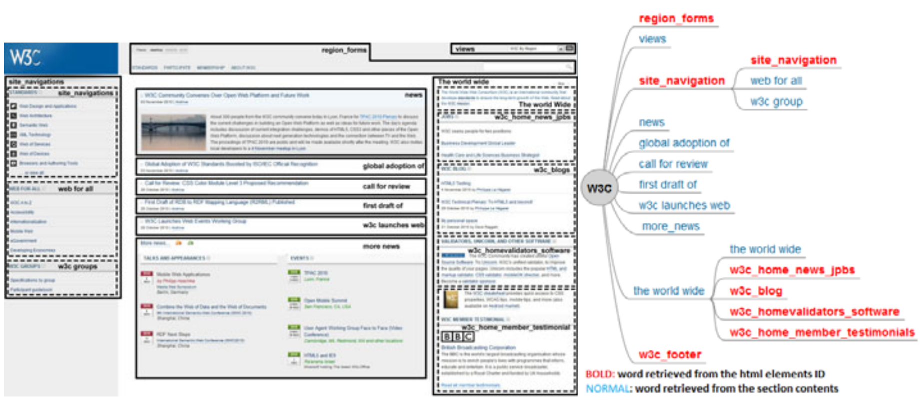


Fig. 2. W3C Interface Splitting and Vocal Menu Structure

Adjusting the adaptation process parameters (described in next sections) allows users to increase or reduce the level of nesting in the vocal menus navigation and, consequently, to control the fragmentation of the resulting vocal interfaces. The table below shows a possible resulting dialogue for the example in Figure 2.

Table 1. The Vocal Dialogue in the Example

	Input / Output	Action
System:	Welcome to Word Wide Web Consortium W3C Main Menu. Choose one of the following sections: region form, views, …	
User:	Call for review	*Menu Activation*
System:	You are in Call for review. Call for Review: CSS Color Module Level 3 Proposed …	
User:	Repeat	*Ask the system to repeat*
System:	Call for review. Call for Review: CSS Color Module Level 3 Proposed …	
User:	Next	*Go to the next part*
System:	Moving to part "The Cascading Style". The Cascading Style Sheet Working Group has published …	
User:	Previous	*Go back to the previous menu*
System:	Moving back to the main menu. Welcome to Word Wide Web Consortium W3C Main Menu. Choose one of the following sections: region form, views, …	
User:	Exit	*Close the interface.*
System:	Goodbye!	

4 The Graphical-to-Vocal Adaptation

This transformation is performed after the reverse engineering phase, in which an analysis of the implementation in terms of all the HTML tags and associated CSS files is performed on the Web page considered in order to build the corresponding logical description.

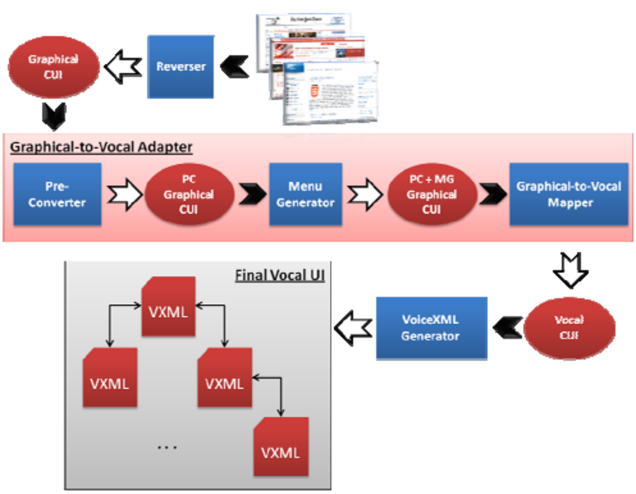


Fig. 3. The adaptation process in detail

The transformation is carried out through various phases (see Figure 3). Firstly the *Pre-Converter* processes the graphical concrete logical descriptions by removing elements that are considered useless for the overall conversion (e.g. images without alternative text description). Secondly, the *Menu Generator* produces a new hierarchical structure in order to allow users to navigate the interface through menus/submenus. Finally, the *Graphical-to-Vocal Mapper* transforms the graphical concrete elements onto concrete vocal ones having the same semantic effects but using different attributes.

4.1 Concrete Graphical Description Pre-conversion

A number of steps are necessary to transform the logical graphical description in input into a cleaner and more suitable one, which is then used for deriving the vocal logical description. In this work we have identified a number of parameters that can be modified in order to customize the adaptation process.

We perform four steps for the pre-conversion that should be accomplished in sequence. The first consists of removing the elements that are problematic for vocal interfaces. In particular we firstly remove the interactors that contain *special characters* (this option is user settable): the use of special symbols (e.g., chinese alphabet if not supported) by the voice browser may be problematic. Secondly, we remove all the image interactors (description, navigator) *without the ALT attribute*: without alternative description it is impossible to render an image vocally. Then we remove the description/image interactors that have *width or height less than a threshold* (this parameter is user settable): usually small images (but not image link) are used more for decoration rather than to provide content. We also remove *redundant navigator interactors*: if we find navigators with more than one label (image, text), then we maintain only one of them, preferably the textual one.

In the second step we simply remove any grouping that contains zero or one child. This step has to be performed recursively until the specification no longer requires further modifications.

In the third step we identify the words/sentences that are candidates as the title/subtitle for menu items. A *candidate word/sentence* is a description/text that has a *role* attribute of “heading1”, “heading2” or “heading3”, and that is placed just before a grouping. In this way we detect an association between title/subtitles and grouping that provides better results in the menu generation process. Candidate words/sentences are moved inside the grouping in order to facilitate their use in the corresponding menu item generation, which is performed afterwards.

The purpose of the fourth step is to avoid too heavily nested groupings, which would then be transformed into too many vocal menus. In this step after removing the groupings *with only one child*, which may be created by the previous step, the tool removes the groupings *whose children are only groupings,* only when the children are deeply nested (compared with the global structure). In this way we reduce the complexity of the structure, moving towards the root the information.

4.2 Menu Generation

The result of the previous phase makes it possible to generate menu sections that define a new logical structure of the interface itself. Moreover, through the choice of appropriate parameters, it is possible to reduce the number of menu items, in case they are too many, in order to avoid short term memory overload.

We designed and implemented a recursive process that analyzes the presentations in the graphical logical description and decides whether to split them into connected multiple presentations. At the end of the process, there will be one presentation containing a menu and the others the actual content. The purpose of the menu generation phase is to identify when we should create menus that will allow the users to choose among various options. A presentation is split (and a new menu is generated) if at least one of the following conditions is verified: *it contains more than one relation*, we consider relations to be atomic and consistent interactions corresponding to form filling, so it is desirable to have a menu item for each of them. *The number of navigators is much greater than the number of description interactors*, this is the case of navigator bars, which should correspond to one single vocal presentation with the associated menu. The threshold ratio of navigators to description interactors can be defined by the designers. *The length of the overall text content exceeds a fixed threshold (user settable)*: this avoids the generation of unbalanced sections in terms of content. Parameterization can be performed through a user interface, which has on the left the modifiable parameters, and on the right the graphical representation of the structure of the corresponding vocal interface, useful for users to have an idea of the structure of the resulting vocal interface.

One of the main problems in the menu generation is how to find keywords (or short sentences) that describe the content of the destination. One possibility is to use the grouping ID originally defined by the designer of the graphical interface. In the case that this value is missing the tool performs an analysis of the grouping content and uses the string of the first description/text with the highest text *role* present in the content considered. The text roles indicated in the graphical concrete description are sorted from the most to the least important in this way: *Heading1, Heading2, Heading3, Heading4, Heading5, Heading6, Strong, Emphasis, List_element, Paragraph, Normal*. If the string is a sentence and not a simple word we use the first three words in the sentence. If no text role is specified, then we select the first string that we find (max 3 words) in the actual content.

4.3 Desktop to Vocal UI Mappings

This is a transformation that takes the elements of the graphical concrete presentations, as modified in the previous stages, and translates them into corresponding elements of the vocal concrete language that have similar effects.

The presentation *title* is used to build a suitable introduction sentence to use as the *welcome message* for the first time the user accesses the application.

Regarding the output-only interactors, we provide the following processing. The text elements are directly mapped onto speech elements. The audio element is mapped onto a *prerecorded message* vocal interactor for rendering. In the case of images, if the ALT attribute is specified in the source document, then we transform the element *image* into a *speech* element that renders the attribute. As in the case of an image, for videos we use the ALT attribute (when present) to vocally render its description, otherwise it is discarded. The *audio* elements are simply mapped onto the *prerecorded message* interactors. Data tables are mapped into *vocal tables*.

Regarding the control interactors we provide for the following processing. The text links are mapped onto the *vocal link* element. The name of the link is used as *activation vocal sentence*. The sentences that users must say to activate a link will be synthesized with higher pitch in order to let the user recognize it. For the image link currently we use the *navigator ID* as the corresponding vocal command. Graphical buttons are mapped onto *vocal links,* triggered by pronouncing their labels.

In the case of graphical interactors supporting selection, they are managed differently depending on whether they support *single* (radio button, list box, drop down list) or *multiple* selections (check box, list box), and consequently associated with the single or the multiple vocal selection interactors. In both single and multiple vocal selection, the labels of the graphical choice elements are used to build either the request or the list of accepted inputs.

The edit interactors allow users to enter pieces of information of various types. They raise the issue of the generation and use of grammars for vocal interfaces. For *numerical input* (spin boxes, track bars) we set a grammar able to recognize numbers. In case of *textual input,* we instead introduce two pre-generated grammars: a vocal grammar able to recognize short sentences composed of a well-defined list of words; and a DTMF grammar able to transform the keyboard inputs into words in a way similar to that used to write text messages in mobile phones. Lastly, once the vocal concrete description has been obtained it can be transformed into a VoiceXML implementation. The type of approach used for this purpose is described in [5].

4.4 Evaluation

We conducted a first user test of the results of our adaptation process on two Web sites: W3C[1] and BBC[2]. We gathered a number of positive user impressions and criticisms that gave us further motivation.

Accessibility was one of the strong points underlined by some users. This approach in fact expands the accessibility horizons not only for visually-impaired users, but also for people without Internet connection or, moreover, when the visual modality is busy (e.g. checking email when driving a car, follow a recipe while cooking, etc.). Another strong point is the simple navigation, also useful for people who are not used to interacting through vocal interfaces.

Some users criticized the TTS and the Speech Recognizer, but these features are not part of our architecture. Three users considered some of the sentences for menu activation too long, while another one asked for more attention in terms of re-prompting and help hints when the user seems to stop during navigation. Some users seemed confused by the terms selected to identify the structure of the generated vocal interface. The term “section part”, for example, misleads users to draw a parallel with book structures. A related problem was the name of some commands that seemed hard to memorize. For example using “next/previous part” instead of “next/back” for the navigation. Some users suggested expanding the command list.

---

[1] `http://www.w3.org/`

[2] `http://www.bbc.co.uk/`

5 Conclusions and Future Work

Our automatic adapting process has been applied to twenty Web pages including W3C and BBC Web sites, which have been used in the user test. The generated VoiceXML vocal interfaces have been tested with the VOXEO Voice Browser[3] and have passed the validation test integrated in it. The applications have been used through VoIP access to the vocal server. A configuration tool was developed in order to change the parameters of the adaptation process and have a preview of the structure of the generated vocal application. This allows the possibility for developers of vocal applications to control the adaptation process and better tailor it to their needs.

In general, the results depend on the original content. The solution is not able to manage content such as Flash or Java applets and the results are not as good with Web content that does not follow the standard W3C guidelines for Web applications.

Future work will be dedicated to improving the transformation rules and further performing empirical validation. We also plan to extend the adaptation support to HTML5 applications.

Acknowledgments. We gratefully thank the support from the EU ICT SERENOA Project (http://www.serenoa-fp7.eu/).

References

1. Fonseca, J.M.C. (ed.): W3C Model-Based UI XG Final Report (May 2010), `http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui-20100504/`
2. Franz, A., Milch, B.: Searching the web by Voice. In: Proceeding of the 19th International Conference on Computational Linguistic, Stroudsburg, PA, USA, vol. 2, pp. 1–5 (2002)
3. Yesilada, Y., Stevens, R., Harper, S., Goble, C.: Evaluating DANTE: Semantic transcoding for visually disabled users. ACM Transactions on Human-Computer Interaction 14(3), Article 14 (2007)
4. Paternò, F., Santoro, C., Spano, L.D.: MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM Trans. Comput.-Hum. Interact. 16(4) (2009)
5. Paternò, F., Sisti, C.: Deriving Vocal Interfaces from Logical Descriptions in Multi-device Authoring Environments. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) ICWE 2010. LNCS, vol. 6189, pp. 204–217. Springer, Heidelberg (2010)

[3] `http://www.voxeo.com/`

Using the Journalistic Metaphor to Design User Interfaces That Explain Sensor Data

Martin Molina[1], Enrique Parodi[1], and Amanda Stent[2]

[1] Department of Artificial Intelligence
Technical University of Madrid, Spain
martin.molina@upm.es, enrique.parodi@fi.upm.es
[2] AT&T Labs - Research
Florham Park, NJ, USA
stent@research.att.com

Abstract. Facilitating general access to data from sensor networks (including traffic, hydrology and other domains) increases their utility. In this paper we argue that the journalistic metaphor can be effectively used to automatically generate multimedia presentations that help non-expert users analyze and understand sensor data. The journalistic layout and style are familiar to most users. Furthermore, the journalistic approach of ordering information from most general to most specific helps users obtain a high-level understanding while providing them the freedom to choose the depth of analysis to which they want to go. We describe the general characteristics and architectural requirements for an interactive intelligent user interface for exploring sensor data that uses the journalistic metaphor. We also describe our experience in developing this interface in real-world domains (e.g., hydrology).

Keywords: Intelligent user interface, multimedia presentation, interactive data exploration, user interface for web applications.

1 Introduction

Today, organizations are collecting large quantitative datasets from distributed sensors including GPS devices, RFID tags, and point of purchase devices. There exist advanced visualization systems to help domain experts analyze, interpret and understand data from sensor networks [5][11]. Non-expert users can also benefit from analyses of sensor network data. Web-based systems that automatically construct textual, spoken or multimedia explanations of the meaning of sensor data using non-technical language can facilitate access by non-expert users and consequently increase the utility of sensor network infrastructures.

An example of this situation in the hydrological domain is sensor network with real time data about water levels, water flows and other meteorological information. In contrast to a specialized system that only presents graphics and quantitative results of analyses to expert hydrologists, an interactive web application with additional text and multimedia *explanation* capabilities is potentially useful to a wide range of non-expert users (e.g., municipalities, civil protection, engineering consultants, and educators).

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 636–643, 2011.

Other examples of large quantitative datasets potentially useful for different communities of users are environmental data from ship traffic [23], crime statistics [22], and wildlife tracking data [18].

The addition of text explanations to visualizations of sensor data can facilitate better understanding and more effective use of the data. The automatic generation of presentations combining text and graphics is a well-known but complex problem [6][7][12]. In this paper we analyze two particular questions: (1) what presentation style (combining text and graphics) can be used to help non-expert users analyze sensor data, and (2) what techniques can be used to support the automatic generation of this type of presentation. Our answer to the first question is to use the journalistic metaphor, because the journalistic layout and presentation style are widely understood. Our answer to the second question is a general design that combines and adapts techniques from multimedia presentation systems and intelligent data analysis.

In the following sections, we first describe our concept of an user interface following the journalistic metaphor. Then, we describe the main components of a general architecture to support this type of interface. Finally, we describe applications that we developed to evaluate our approach, and a comparison with related work.

2 A User Interface Based on the Journalistic Metaphor

Journalists, who face the challenging task of presenting complex situations to non-experts in a way that quickly and effectively promotes understanding, have converged on practical guidelines for effective communication to wide audiences. We use these journalistic guidelines to design a user interface that explains sensor data to wide audiences. In general, the journalistic style uses non-technical language understandable by general users and follows guidelines such as the following:

- *Writing using the inverted pyramid*. A news story is usually structured to put essential information first, with supporting information following in order of diminishing importance. The *headline* captures the user's attention and quickly communicates the main idea of the story. Then, the *text body* develops the headline. The text body is structured to put essential information first, with supporting information following in order of diminishing importance. It is structured so that the user can stop at any point with a high-level understanding, or continue to read to get more detailed analysis.
- *Effective use of graphics to elaborate on the text*. A feature story is often accompanied by illustrations including 2D maps and timelines, bar charts, line charts, pie charts, and images. Figure captions summarize the content of the figure. Figure legends explain the meaning of certain symbols and act as quick references to subsets of the data identified during data analysis. The text body also explains the meaning of the data analyses displayed in the graphics.
- *Page layout to organize related information*. A newspaper organizes news stories into different sections, with the most important stories on the first page. Graphics and text may be presented in a one-column layout (with graphics located near the relevant text), or in multi-column layouts (with graphics running parallel to the text).

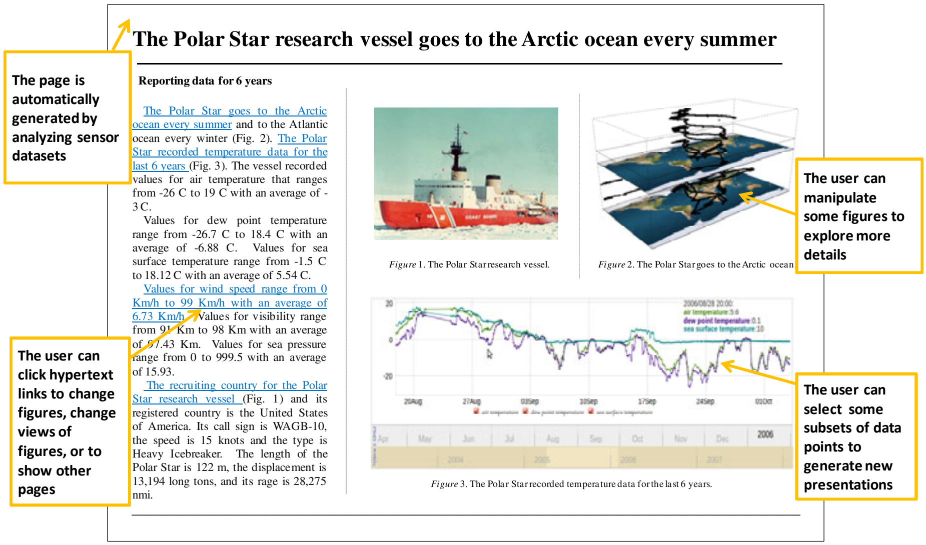


Fig. 1. Example presentation for interactive exploration of sensor datasets

Figure 1 shows an example presentation in a user interface that we developed following the journalistic metaphor. This presentation is based on a large set of environmental information collected by ships over the last ten years as part of the international VOSclim project [23]. This data set includes tens of measures for each of hundreds of thousands of events for hundreds of ships. Our web-based, interactive interface works as a virtual newspaper that automatically generates multimedia news stories that highlight and explain analyses of sensor data.

3 Automatic Generation of Presentations

To implement a user interface that explains sensor data using the journalistic metaphor, we need to simulate tasks that human journalists perform to construct presentations. These tasks are: analyze the data and find the important elements, create a plan to present the information to the reader, generate the text in natural language using general terminology, generate graphics that complement the text, and generate a logically consistent layout for the presentation. In order to automatically perform these tasks, we have designed a general architecture that combines and adapts solutions from the fields of multimedia presentation systems and intelligent data analysis. Figure 2 shows the main components of our architecture.

This architecture shares some general components of multimedia presentation systems [1] (e.g, content planning and presentation generation). Some original contributions of our design are: (1) *data analysis of large spatio-temporal datasets*, i.e. we automatically analyze large datasets with spatio-temporal information to find patterns and perform aggregations (e.g., spatial and temporal aggregations), (2) *explanatory multimedia descriptions*, i.e. we automatically construct complex descriptions (with multi-paragraph natural language summaries and graphics) with informative and persuasive discourses that summarize and explain the sensor data (here, we follow the journalistic metaphor which provides a guide to structure the presentation), and (3) *interactive exploration*, i.e. our user interface supports a mixed-initiative approach to interactive data exploration.

Our architecture includes a database that efficiently stores sensor data (thousands or millions of events) and the results of data analyses. The database represents raw sensor readings as *observed* events, each characterized by a set of attribute values. Our architecture also includes a component for data analysis to find abstractions over the input sensor data. The data analyzer uses general data mining algorithms to cluster, analyze trends, and describe the data in terms of the most common feature values, example instances, and exceptions. The automatic application of geospatial, temporal and domain-specific knowledge bases to the observed events produces *extended* and *aggregated* events, with new attribute values corresponding to abstractions of the observed attribute values.

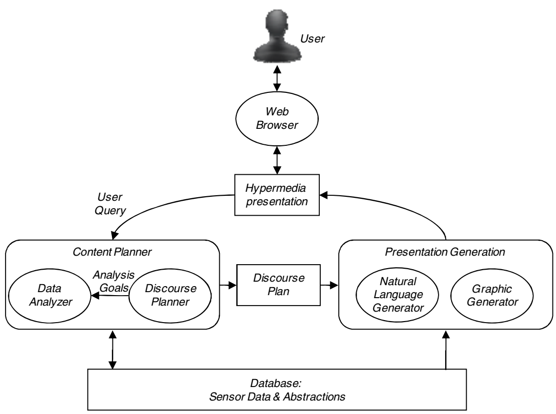


Fig. 2. Main components of our architecture for automatic generation of news from sensor data

In order to generate multimedia descriptions using the journalistic metaphor, our system must select from numerous potential ways of describing the input data. Therefore, it is not practically possible to use a few templates that help to generate these descriptions. Instead, we use a more flexible solution with a discourse planner in combination with natural language generation and graphics generation. Our discourse planner is conceived as a knowledge-based hierarchical planner. The knowledge base includes partial discourse patterns in the journalistic style, together with conditions for selecting each discourse pattern. The generated discourse plan includes information about the rhetorical structure of the discourse (with rhetorical relations [13]) and information about propositions in the discourse, represented as references to subsets of the input data (Figure 3). In our architecture the discourse planner operates in a loop with the data analyzer. The discourse planner invokes specific prefixed data analysis functions using the knowledge base of discourse patterns and the results of previous data analyses to progressively construct the discourse plan.

The component for *presentation generation* receives as input the discourse plan output by the discourse planner. The presentation generator generates interactive hypermedia presentations (in HTML and JavaScript) accessible by remote web browsers through the Internet. Our user interface supports a mixed-initiative approach to interactive data exploration. Some data analysis tasks are chosen and performed automatically by the system, and presented to the user as an inter-connected collection of multimedia stories through which the user can browse. The user can invoke other tasks herself by clicking on hyperlinks in a story or selecting subsets of the data via manipulation of a graphic.

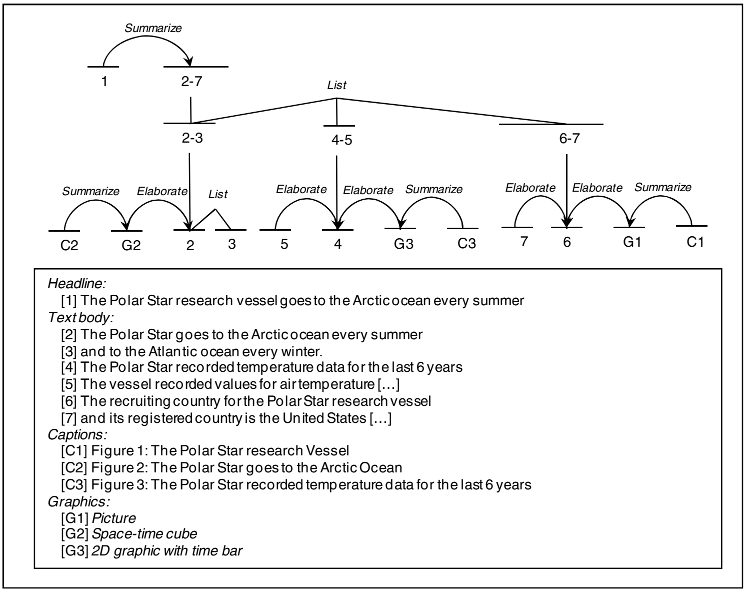


Fig. 3. Discourse structure of the presentation (simplified example)

The generation process involves picking elements off of the discourse plan and invoking text or graphics generation components accordingly. The *text generator* produces natural language text using two components, sentence planning and surface realization [20]. The figures that accompany texts are constructed on the fly by *the graphics generator.* The graphics generator can make use of several publicly available graphics tools, such as Google Charts, Dygraphs, SIMILE Widgets, and OpenStreetMaps.

4 Applications and Related Work

We developed a real-world application in the domain of hydrology following the user interface design described in this paper, using sensor data from a national information system (SAIH: Spanish acronym for Automatic Information System in Hydrology) in Spain. The application generates 20-30 pages of news every hour summarizing 44,736 sensor readings across Spain. For example, Figure 4 shows one of the generated pages with a headline (light rain at a few places in the Guadiana basin), text body (with hyperlinks), and two graphics: an animated illustration (the movement of a storm), and an interactive map. We evaluated the practical utility by comparing our application with existing web applications that present hydrological sensor data. The results of this evaluation showed that, with this system, users take up to 4.45 hours less to confirm the information in a summary of the hydrological data than they do given other existing systems for this data (with an average value of 3:45 hours for typical emergency scenarios).

Fig. 4. Example presentation in hydrology

The application in the hydrologic domain followed the journalistic style described in this paper. It also included a preliminary version of our architecture with domain specific knowledge bases and abstraction methods (more details about this application domain can be found at [16]). We have also applied our design partially to other domains: ship data from the VOSclim project [23], Twitter data with measures of sentiment [4], general geographical movements [17], and aircraft data [15]. As a result of these applications, we generalized some components of our design as they are described in this paper. We are currently working on a complete implementation of this design using different domains for evaluation with spatio-temporal data from sensors.

Our solution is related to multimedia presentation systems such as WIP [2], COMET [14], Autobrief [9], and Dart$_{bio}$ [3]. Compared to these systems, our system is original in the following aspects: (1) it is designed for a different problem (generation of descriptions about the meaning of sensor data), (2) it follows the journalistic style and constructs complex multi-page presentations, (3) it includes complex and interactive graphics (animations, maps, etc.), and (4) it has been evaluated in real-world domains (e.g., hydrology).

Our system also shares architectural elements with other data-to-text generation systems (for weather forecasting [21], or medicine [10]). However, we use a different data analysis approach, directed by the discourse planning process; and we produce multimedia presentations. Our system can also be related partially to user interface generation using high level models [8]. Compared to this, our solution also uses high level models (discourse plans with rhetorical relations) to generate output presentations.

Our proposal is also related to the field of computational journalism. In this field, our solution can be compared to research work related to automatic generation of news (e.g., [19]). In this type of solution, news is generated using text-to-text summarization techniques and web mining. In contrast, our system creates news stories from quantitative measures of sensor data, and automatically generates coordinated presentations of text and interactive graphics.

5 Conclusions

In this paper we have described an intelligent user interface, conceived as an automatically generated virtual newspaper, for interactive exploration of sensor datasets that combine text summaries and graphics. We have adopted the journalistic metaphor to make the results of data analyses easily understandable by non-expert users. This metaphor provides general criteria to automatically construct presentations (with headlines, types of discourse patterns, illustrations and specific page layouts). To support this type of interface, we have identified the main components of a general architecture combining and refining solutions from multimedia presentation systems and intelligent data analysis. We have applied this solution to real-world domains and shown its practical utility. This type of user interface is an effective solution to improve the utility of sensor networks.

The implementation of our architecture uses general-purpose reusable components. However, it also requires the development of certain domain-specific components, more specifically, domain-specific discourse patterns, knowledge bases and data analysis methods. Our future work includes the generalization of these domain-specific elements to facilitate reuse, as well as evaluation in different domains.

Acknowledgements. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement PIOF-GA-2009-253331 (Project INTERACTIVEX). This work was also partially supported by the Ministry of Science and Innovation of Spain within the VIOMATICA project (TIN2008-05837).

References

1. André, E.: The generation of multimedia presentations. In: Dale, R., Moisl, H., Somers, H. (eds.) A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text, pp. 305–327. Marcel Dekker Inc., New York (2000)
2. André, E., Rist, T.: The design of illustrated documents as a planning task. In: Maybury, M. (ed.) Intelligent Multimedia Interfaces, pp. 94–116. AAAI Press, Menlo Park (1993)

3. Bateman, J., Kleinz, J., Kamps, T., Reichenberger, K.: Towards Constructive Text, Diagram, and Layout Generation for Information Presentation. Computational Linguistics 27(3) (2001)
4. Barbosa, L., Feng, J.: Robust sentiment detection on Twitter from biased and noisy data. In: 23rd International Conference on Computational Linguistics, COLING (2010)
5. Card, S.K., Mackinlay, J.D., Shneiderman, B. (eds.): Readings in Information Visualization – Using Vision to Think. Morgan Kaufmann, San Francisco (1999)
6. Carenini, G., Moore, J.: A strategy for generating evaluative arguments. In: First Int'l Conf. on Natural Language Generation (2000)
7. Demberg, V., Moore, J.: Information presentation in spoken dialogue systems. In: 11th Conf. European Chapter of the Association for Computational Linguistics (2006)
8. Falb, J., Kavaldjian, S., Popp, R., Raneburger, D., Arnautovic, E., Kaindl, H.: Fully automatic user interface generation from discourse models. In: 14th International Conference on Intelligent User Interfaces (2009)
9. Green, N., Carenini, G., Kerpedjiev, S., Mattis, J., Moore, J., Roth, S.: AutoBrief: an Experimental System for the Automatic Generation of Briefings in Integrated Text and Information Graphics. International Journal of Human-Computer Studies 61(1), 32–70 (2004)
10. Hunter, J., Gatt, A., Portet, F., Reiter, E., Sripada, S.: Using natural language generation technology to improve information flows in intensive care units. In: 5th Conf. on Prestigious Applications of Intelligent Systems (2008)
11. Keim, D.A.: Information visualization and visual data mining. IEEE Trans. on Visualization & Computer Graphics 8(1), 1–8 (2002)
12. Kopp, S., Tepper, P., Cassell, J.: Towards integrated microplanning of language and iconic gesture for multimodal output. In: 6th Int. Conference on Multimodal Interfaces (2004)
13. Mann, W., Thompson, S.: Rhetorical Structure Theory: Toward a functional theory of text organization. Text 8(3), 243–281 (1988)
14. McKeown, K.R., Feiner, S.K.: Interactive multimedia explanation for equipment maintenance and repair. In: DARPA Speech and Language Workshop (1990)
15. Molina, M., Parodi, E., Stent, A.: Combining text and graphics for interactive exploration of behaviour datasets. In: 22nd Int. Conference on Tools with Artificial Intelligence (2010)
16. Molina, M., Flores, V.: A presentation model for multimedia summaries of behavior. In: 13th International Conference on Intelligent User Interfaces (2008)
17. Molina, M., Stent, A.: A knowledge-based method for generating summaries of spatial movement in geographic areas. International Journal on Artificial Intelligence Tools 19(4), 393–415 (2010)
18. Obis-Seamap, `http://seamap.env.duke.edu/`
19. Radev, D.R., Otterbacher, J., Winkel, A., Blair-Goldensohn, S.: NewsInEssence: Summarizing online news topics. Communications of the ACM 48(10), 95–98 (2005)
20. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge (2000)
21. Reiter, E., Sripada, S., Hunter, J., Yu, J., Davy, I.: Choosing words in computer-generated weather forecasts. Artificial Intelligence 67(1-2), 137–169 (2005)
22. UCR, Uniform Crime Reports, `http://www.fbi.gov/ucr/ucr.htm`
23. VOSclim, `http://www.ncdc.noaa.gov/oa/climate/vosclim/vosclim.html`

Domain Experts Tailoring Interaction to Users – An Evaluation Study

Helena Lindgren, Patrik J. Winnberg, and Peter Winnberg

Department of Computing Science, Umeå University,
SE-90187 Umeå, Sweden
{helena,patrikw,pw}@cs.umu.se

Abstract. This paper presents ACKTUS, a modeling tool for developing knowledge-based systems for the health domain, and an evaluation study of the system. The main purpose of the evaluation was to investigate whether the functionality and interaction design of ACKTUS was sufficiently intuitive for the domain experts to contribute with knowledge and to model the interaction design of the three end users' applications. Another purpose was to evaluate the applicability of the activity assessment protocol AAIMA for analysis. The study design was qualitative and formative, using observations and interviews with users to collect data. Three medical experts and two experts in occupational therapy participated, providing expertise in four different domains. The participants increased their understanding and skills during the evaluation period leading to improved knowledge-based applications. The AAIMA protocol proved to be useful and the results are fed into ongoing development work on developing the adaptive functionality of the ACKTUS systems.

Keywords: Formative evaluation, interaction design, end-user development and adaptation, knowledge modeling, e-health.

1 Introduction

Knowledge-based systems are to an increasing extent used by individuals for monitoring health and for supporting the daily work of health care professionals in the provision of care to individuals (e.g., [1]). In both cases it is of great importance to assure that the mediated knowledge has high quality (e.g. based in evidence-based medicine or other reliable sources) and that the interaction design of the system allows for tailored advice and personalized support in addition to transparency, in order for the end user to gain trust and confidence in the system's content and responses. For these purposes, the system ACKTUS (Activity-Centered Knowledge and interaction modeling Tailored to Users) is being developed, making use of Semantic Web-technology to model both the health-related knowledge and the interaction design of personalized knowledge-based systems [2], [3]. The knowledge engineering of such systems typically requires deep knowledge in the medical domain, knowledge of formal methods and an extensive amount of time to transform informal knowledge into formal structures (e.g., [4]). Because of this, the major aim of the ACKTUS project is to facilitate the modeling process while also making the

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 644–661, 2011.

domain professional take the interaction design of the end user application into consideration. For the domain professional to be able to also model in what way the end user will interact with the system, functionality that supports this has been integrated and extended during development. Research focuses on how the design of the modeling environment allows the domain professionals to accomplish creating what they envision being the functionality and interaction design of the future support systems. This paper presents ACKTUS and the results from a formative evaluation study, conducted as part of the development of ACKTUS. The overall research question is to what extent the domain expert is able to grasp the structure of ACKTUS and how well the system supports the development of this understanding. A related research question is how well the system supports the domain expert in designing the interaction for other end users. These questions are addressed by investigating the participants' performance in modeling tasks in authentic work, before and after the interface re-design, which aimed to improve the understanding of the system. This was partly done by adopting and adapting an activity assessment protocol presented in [5]. Therefore, an additional central research question is to what extent the adapted protocol is useful for capturing knowledge and skill development in users.

1.1 Tools for Knowledge Representation and Interaction Design

Guidelines for clinical practice are being developed for the purpose of improving care. These typically come in natural language and are not straightforwardly translated into a guideline interpretable by a computer. In order to facilitate this process a number of representation languages have been developed by different research groups specifically for the purpose of representing clinical guidelines and treatment protocols. These languages typically allow for describing "task-network models", where sets of interacting decisions and actions are to be carried out in sequence or in parallel over a period of time. Some examples are Asbru, PROforma, GLIF, EON, GUIDE, and PRODIGY (see comparisons done by Peleg et al. [6] and Isern et al [7]). These languages have in common that they provide methods for regulating a care process with pre- and post-conditions attached to plans. Tasks can be defined and associated to different points in a plan, where the task may involve data collection to be made by a health care professional. At these points typically forms appear where the required information according to the underlying clinical guideline can be entered.

Evaluations have been made of these languages within applied implementation projects of clinical guidelines, e.g, [8], [9], [10]. It has been shown that it takes time for the domain expert to learn the underlying structure [8]. In some cases, the knowledge acquisition phase has been facilitated by developing simplified graphical user interfaces (GUI) on top of the original GUI such as in the case of using a MediaWiki for modeling Asbru plans [10]. Another approach to address this issue is described by Seyfang and coworkers [4]. They developed a semi-structured format for simplifying the process on top of Asbru. In another project, a UML activity flow diagram was used when PROforma showed limitations in modeling decision flows in a graphical structure in an easy way [9].

Besides these languages, which are primarily developed for knowledge engineering purposes, we are not familiar with any generic knowledge modeling system that also allows the domain expert tailor the interaction with the end user application in a similar way as the system in focus for our evaluation study.

1.2 ACKTUS and the Application Domains in Focus for the Study

ACKTUS is being developed for the purpose of providing a modeling environment for domain experts who also may be potential end users of knowledge-based systems in the health domain. Another purpose of the system is managing knowledge integrated into knowledge-based systems. One goal is to design the system in such a way that end users are able to contribute to the knowledge in interaction with the system. Therefore, it is not expected that the end users or the domain experts participating in the knowledge modeling are familiar with formal models such as logic or formal ontologies. The task-network modeling languages described in the previous section are developed for modeling clinical practice guidelines and treatment protocols, mainly to support a care process within a care providing organization or between organizations. By contrast, in the current application projects, ACKTUS is used for synthesizing several knowledge sources into a support system in domains where the domain knowledge is incomplete and evolving. Another difference from the task-network languages is that the support systems developed using ACKTUS are not necessarily aimed at health care professionals. The end users in the current application projects are, besides health professionals, mining and construction workers and older persons in their homes. Therefore, ACKTUS also incorporates knowledge such as rules-of-thumb and best practice knowledge to be integrated apart from the knowledge expressed in clinical practice guidelines and other evidence-based medical knowledge sources. Preference orders among knowledge sources, level of expertise in knowledge domains and other contextual factors are taken into account in the reasoning about e.g., diagnosis in a patient case, or about adjustments in a work environment for preventing injuries.

The Semantic Web-application ACKTUS consists of a service-oriented architecture and integrates a formal terminology and information model in the form of an RDF/OWL ontology. The ontology captures 1) components used for tailoring interaction with the resulting knowledge applications, 2) components of argument-based reasoning and 3) components for modeling the user agents as actors in a reasoning process [2]. The ontology components are used by the domain experts for designing the content and interaction flows, and they put constraints on the modeling activity in order to support the domain expert in the modeling tasks. The outcome is a set of instantiated components and their relations, which function as structured tools for the end user in the interaction with the end user application. In addition to the end user application the ACKTUS architecture allows a developer, with sufficient permissions, to write applications that can access and manipulate the information using readily available Internet technologies like HTTP.

The key functionalities of ACKTUS are the following (Figure 1):

1. Integrate and model *knowledge sources*
2. Translate the content of knowledge sources into semi-formal *argumentation schemes*
3. Implement the schemes as formal *rules*
4. Define *reasoning contexts* to be activated by interaction protocols

5. Define and model *interaction protocols*
6. Define and model *assessment protocols* as sets of interaction protocols
7. Create rules for assessment flows between assessment protocols and interaction protocols
8. Apply and evaluate the composed system through end user system interfaces and with fictive patient or user cases

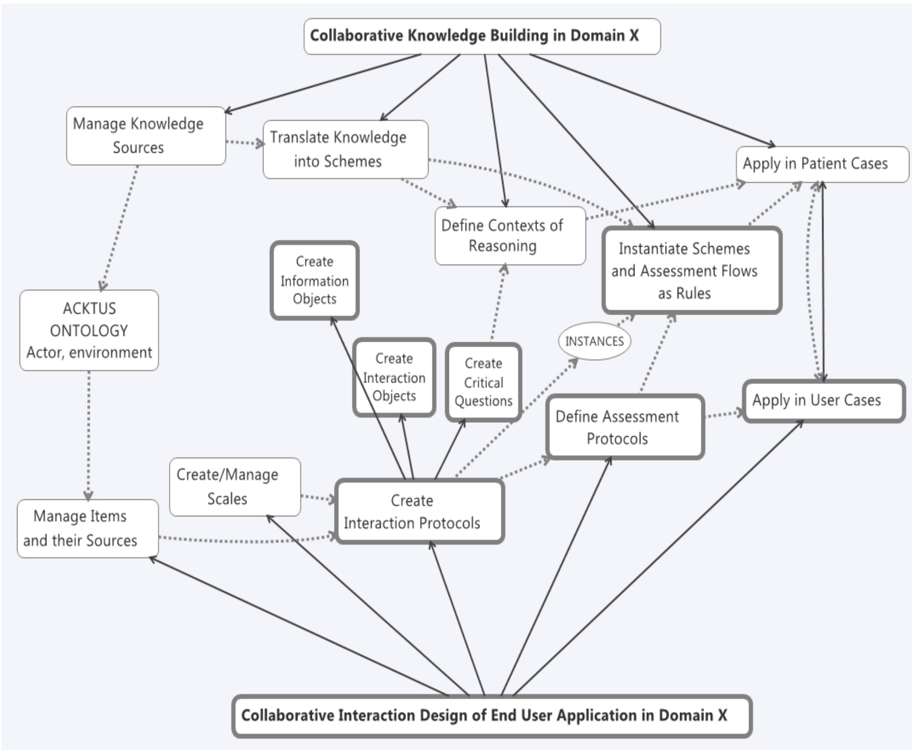


Fig. 1. Overview of ACKTUS functionality in terms of major tasks and their subtasks (indicated by arrows). Dotted arrows represent a flow of activity. Tasks focused in our evaluation is marked.

The different functionalities are used to different extent in the application domains, due to their different characteristics. In the dementia domain, where care provision is highly dependent on evidence-based clinical practice guidelines, the tasks 1-3 are crucial and the validation of the work needs to be thoroughly done [11]. Since diagnostic reasoning is performed differently by novice and expert physicians [12] (e.g., in diagnosing dementia diseases) the notion of *reasoning context* is used (Task 4). The reasoning contexts are defined by the domain expert, and frame the topics to investigate at different levels of activity, and may be activated following the individual end user's need for support. The end user is a health professional responsible for diagnosing patients with a suspected dementia disease.

In the domain of occupational health and environmental medicine, less evidence-based knowledge is available, making the tasks 1-3 less focused. Furthermore, a large portion of the information is embedded in the work environment and the motivations for using parts of the application are rooted in the end user's current life situation. Here a larger emphasis is put on the interaction design and flow for the purpose of generating interest and motivation in the end user for taking work-related health issues seriously (Tasks 5-8). The same applies to the occupational therapy domain, where the therapist designs (tailors) the interaction to be accomplished by the client subjected to rehabilitation assessment and intervention [13]. The end users of these applications are private persons, with different life situations motivating the use of a tailored support system. These applications may be supplementary and may share knowledge between agents, such as in the case when the occupational therapist's client is a person developing, or is already diagnosed with an early stage of dementia, or is a person living with a person suffering from dementia.

In this evaluation study we focus on the tasks 5-8 since these are about equally important to the participants, who work in different application projects in our study during the evaluation period. The study then supplements an earlier study focusing the modeling of clinical guidelines in the dementia domain using the tasks 1-3 and 8 [3].

2 Methods

A qualitative and formative user-centered evaluation of the end-user graphical user interface of the ACKTUS system was performed using a general action research methodology. The evaluation was performed in two phases, representing two iterations of the development of the system as a whole, including the user interface design. In the first phase the methods activity analysis and cognitive walkthrough were used to form a baseline for user observations and interviews. A re-design was performed before the second phase based on the results of the first phase. During the second phase further evaluations were made using observation of and interviews with users. The methods used and the procedure is described in more detail in this section.

The evaluation study followed an action research methodology, which is a collective and participatory research method to achieve change or learning. In accordance with action research methodology, the project was carried out in cycles of planning, acting, observing the results in a natural environment and reflecting upon them, all the while involving those responsible for and affected by the practice to be changed [14].

The system and its context were analyzed using activity theory as base for analysis [14], generating a structured view of the workflow and the actions possible to perform in the system. Tasks were chosen from this set of actions for the users to perform in the user evaluation sessions. The chosen actions did not cover all aspects of the user interface, but instead those that were most relevant to the users in their ongoing modeling work using the system. The results from the user observations were then analyzed using an activity assessment protocol adapted from Lindgren [5], order to identify where in the interaction problems arose and to what degree the user managed to perform independent work with the system. In the following subsections the activity analysis and the user evaluation study are further described.

2.1 Activity Analysis

The options and available actions of the system were analyzed using activity theoretical models. The work that can be performed by a user was divided into two overarching activities on an abstract level, which are carried out by actions of different levels of concreteness. The two main activities to be executed in interaction with the system were identified as: 1) *Collaborative knowledge building in domain X*, 2) *Collaborative interaction design of end user application in domain X* (Figure 1). Firstly, the activity of collaborative knowledge building describes how medical health professionals input knowledge into the ACKTUS system interpreted into different levels of complexity. Secondly, how the knowledge is presented and interactively mediated to an end-user is modeled in the collaborative interaction design activity.

These activities are achieved through a number of possible actions using the ACKTUS system as a mediating tool. For instance, knowledge building is achieved by defining knowledge sources and structuring the knowledge into schemes and rules, while the interaction design includes actions such as managing scales, questions and assessment protocols.

Activity is performed in communities of actors; in this case, the collaborative nature of building knowledge and structuring the interaction flow for utilizing that knowledge. A single user will rarely build the knowledge from scratch, as they will typically work in relation to others' ongoing and finished work already entered into the system. The interaction will also develop over time as the knowledge in the system and conditions develop. For instance, the interaction will look different for the first users who enter the first pieces of knowledge than for later users. From this set of activities and actions, a task analysis was developed, listing all options and actions that the interface makes available, from using buttons and controls to e.g. reviewing information that is presented. This task analysis was expanded upon for use in the cognitive walkthrough segment of the evaluation.

AAIMA Protocol for Assessment. A modified version of an instrument for assessing ability in performing computer-supported activity (protocol for Assessment of Autonomy in Internet-Mediated Activity - AAIMA) was used [5] (Table 1). The activity analysis was adapted into a protocol based on this model for use in the observations of users in determining what level of autonomy and complexity a user achieves in performing an action with the system as a tool. Each activity and action is described with three levels of complexity according to the model [5].

For actions, the levels are organized as follows. On the lowest, most basic level (denoted Level 0), the user must understand the purpose of the task and be able to find the section of the interface where the task may be performed. Without an understanding of the purpose of the action, it cannot be executed in a goal-driven manner, which means it could not be considered an action (since actions are goal-based). That is, if a user were to randomly click in the interface and happen to perform the goals of the action by chance, he is still not described as having performed or accomplished the action. On a higher level, Level 1, the user is able to edit information that is already entered into the system with an understanding of the object's components (i.e. that a rule needs premises and a conclusion to function).

Table 1. AAIMA-protocol [5] adapted to the study. We omit here the columns showing needs and motives and limit the table to showing the main activity and four sub-actions in focus.

Act. level	Activity description
Activity	**Collaborative Interaction Design of End User Application in Domain X**
Level 0	Understanding the purpose of the interaction design but does not participate
Level 1	Contributing by defining basic units of information
Level 2	Contributing by designing assessment flows
Level 3	Structuring interaction flow and framework
Action	**Model Interaction Protocols**
Level 0	Finding interaction protocols, understanding their relation to items and scales, understanding the difference between information and interaction objects and critical questions
Level 1	Editing existing interaction protocols, reviewing definitions
Level 2	Creating new interaction protocols with appropriate scales or questions
Level 3	Full understanding of the role of interaction protocols in structuring interaction flows, reviewing relevance of scales, etc.
Action	**Model Assessment Protocols**
Level 0	Navigating to assessment protocols and understanding their structure and purpose
Level 1	Editing existing assessment protocols, defining concepts
Level 2	Creating new assessment protocols
Level 3	Reviewing an assessment protocol's structure by testing it
Action	**Instantiate Assessment flows as Rules**
Level 0	Finding rules, understanding scheme structure and how their content could be fully expressed in rules
Level 1	Editing existing rules and understanding the purpose of its individual components (premises, conclusion, critical questions)
Level 2	Creating new rules, specifying their components and relation to a scheme
Level 3	Understanding the function of rules and their role in reasoning, evaluating whether a set of rules encompass all the content of a scheme
Action	**Apply in User Cases**
Level 0	Navigating to user cases, understanding their relation to assessment protocols and rules
Level 1	Understanding the function of characteristics such as role, motive etc.
Level 2	Following assessment flows in a purposeful manner according to a specific user case
Level 3	Able to effectively utilize user cases to evaluate the validity of the interaction flows of the system, assessing relevance of assessment protocols and questions to a certain type of user

On Level 2, the user is also able to add new knowledge into the system, with an understanding of the knowledge's interaction and relation to other pieces of information in the system (i.e. that an *assessment protocol* contains a set of *interaction protocols* or other assessment protocols). For the third and highest level, a complete understanding of the action and its context is required, and being able to review and analyze the relevance and validity of the knowledge that exists in the system.

If a level of any task causes problems (*breakdowns*), but the user manages to complete them under the guidance of a more experienced peer, the task is said to lie within the user's zone of proximal development (ZPD) [16]. This means the task has the possibility of being performed independently by the user within the near future through training and development. The notion of *breakdown* is used to describe the situation when the user is interrupted in completing the task, possibly due to limitations in the user interface design or the individual's understanding of the task.

2.2 User Evaluations

User evaluation sessions were carried out with five domain expert users in total. Two were experts in the field of environmental health, a domain in which the ACKTUS system is currently being applied (ACKTUS-EKF). These were chosen as they are a part of the project and have a say in its development, and represent the user group consisting of the medical professionals that will enter and structure the system's knowledge. They are also specialized in the field of the other user group, the end-user construction and mining workers. In addition, they are specialized in two supplementary sub-fields of environmental health, with a corresponding division of responsibility in the knowledge modeling process. A third medical professional, involved in the development of the dementia diagnosis subsystem (ACKTUS-Dementia), was chosen for the same reasons and participated in the second phase of the evaluation. This user had participated earlier by modeling clinical guidelines and rules corresponding to the tasks 1-3 described in previous section [3]. Two additional users, experts in the occupational therapy (rehabilitation) domain, participated. In their cases the evaluation was part of an introductory demonstration for the purpose of initiating a process of knowledge and interaction modeling using the subsystem ACKTUS-Rehab.

The users were asked to either continue tasks that they were to perform in the system in the course of their work, or perform certain predetermined tasks if they currently had nothing to work on. The areas covered by their work and the tasks were decided from what was determined to be most relevant to their work in the system at the time and included 1) managing *interaction protocols*, 2) managing *assessment protocols* and 3) creating/managing *assessment flows* using *rules*. The users were also asked questions about their experience with the interface, what was problematic or what they would like to change, etc. The users who tested the Rehab subsystem were given example tasks to finish that demonstrated how the system could be applied to their domain. Following this, they continued to model their own chosen examples with the same purpose, having gained understanding of the purpose of the system. The same procedure was done when the dementia expert was introduced to the tasks.

While the three medical professionals had experience with parts of the system, they were not familiar with all the tasks. The ACKTUS-EKF users were at a beginner level in regards to managing rules, and the ACKTUS-Dementia user had not attempted to define assessment protocols or use rules to define assessment flows before.

During the first phase of evaluation, the parts of the interface in focus for evaluation were in an initial, basic stage of development. Based on the users' performance and the other methods of evaluation, design suggestions were generated and implemented. Significant updates were made to the interfaces of interaction and assessment protocols in particular, and some updates were made to the interface for creating rules. These updated versions of the interface were used by the participants in the second phase, generating further design alterations to be implemented in the future.

The users' interactions with the system and comments during the session were recorded on video. A total of 12 hours of active modeling work was recorded part from recorded interviews. The actions mediated by the interface (clicks, typing, scrolling etc.) were then transcribed and analyzed using the aforementioned activity protocol. The transcription was divided into segments according to the actions of the protocol that the user was trying to achieve at a given time. Each action was then scored for each user by what level of complexity and with what degree of autonomy that an action was completed. Comments and suggestions were also noted and assembled as design suggestions.

3 Results

The results from the evaluation study involving domain expert users are further described in the following subsections with a focus on learning and developing understanding of the system detected and interpreted using the AAIMA instrument. The outcome of using ACKTUS in terms of developed end-user applications is described and aspects concerning redesign are motivated.

3.1 Using the AAIMA Protocol for Detecting Learning and Skill Development in the Use of ACKTUS

A progress in ability to use ACKTUS for their modeling purposes was seen in the users participating in the study. The results from the AAIMA protocol analysis are summarized in Table 2, which provides an overview of the results. In the table each action and activity is marked with the highest level (a number between 0-3) that the user managed to achieve over the course of 1-3 evaluation sessions during each phase. Additionally, for each activity level, how automatized each action was performed (i.e. to what extent the action caused breakdowns) was also noted, but this is excluded from the table for brevity's sake. The ZPD notation in the table designates that the user was able to complete the action at the given level of complexity, but only with some aid or guidance from the evaluator, indicating that the activity may be performed independently within the nearest future (i.e., the next session). The differences in performance between the two evaluation phases can be seen in Table 2. In the following description we use the term *participants* to distinguish the professionals participating in the evaluation study from the future end users of the resulting applications.

Phase 1. In the first phase of observations, none of the participants reached Level 3 of the overall activity, that is, structuring an interaction flow and framework, since they did not manage to construct rules for instantiating assessment flows. This was also technically limited at the time of evaluation. They did reach activity Level 2, in that they were able to organize components of knowledge and interaction components in the system to create assessment protocols. As such, what was limiting the interaction

Table 2. Overview of results from analyses using the activity protocol in Phase 1 and 2. U3 and U5 did not participate in the first phase and U5 was new to the system in Phase 2.

Activity/action	U1 PH	U2 PH	U3 PH	U4 OT	U5 OT
PHASE 1	*<60 min*	*<60 min*		*<60 min*	
Collaborative Interaction Design of End User Application in Domain X	2	2 (ZPD)	N/A	2 (ZPD)	N/A
Manage Interaction Protocols	2	2	N/A	2 (ZPD)	N/A
Manage Assessment Protocols	3	3 (ZPD)	N/A	3 (ZPD)	N/A
Instantiating Assessment flows as Rules	1	0	N/A	N/A	N/A
Apply in User Cases	N/A	N/A	N/A	N/A	N/A
PHASE 2	*102+150min*	*60+63 min*	*45+44 min*	*41 min*	*40 min*
Collaborative Interaction Design of End User Application in Domain X	3 (ZPD)	3 (ZPD)	3	3 (ZPD)	3 (ZPD)
Manage Interaction Protocols	3	2	3	3	2
Manage Assessment Protocols	3	3	3	3 (ZPD)	3
Instantiating Assessment flows as Rules	2 (ZPD)	2 (ZPD)	3	3 (ZPD)	2 (ZPD)
Apply in User Cases	3	3	3	3	2

at this stage seems to have been the understanding of and application of rules. The only action that reached Level 3 in the first phase, full understanding and reviewing contents, was *managing assessment protocols* (Figure 2). Level 3 caused problems for one user however, mostly concerning understanding what objects would appear when testing the protocol and how to choose which would appear (i.e. adding questions to the protocols). To summarize the first phase, it seems that the problems for the users were in the understanding of certain actions more than their execution. That is to say, the users were able to edit and create new objects (critical questions etc.) but had trouble with distinguishing different kinds of objects, or understanding the objects' role and relationships in interaction flows (e.g. when an interaction object would come into play). Obtaining a full impression or overview of how things relate and function appeared to be problematic at this stage. Therefore, the system's presentation of these aspects was worked on before Phase 2, for instance, in the labeling of controls and objects, titles, explanations of concepts and natural structure, progression and mappings, etc. (some results are shown in Figures 3 and 4).

Redesign of ACKTUS. The different evaluation methods yielded different design issues to be addressed. The results from the first phase of the user observations and activity analysis indicated that obtaining a full impression or overview of how objects

Fig. 2. Initial version of the assessment protocol, used in Phase 1

relate and function in the system is problematic. The system's presentation of these aspects needed to be addressed, for instance in the labeling of controls and objects, using clear titles, explaining concepts, using a natural structure and suggesting the next step in a workflow, reducing information load to facilitate user decisions, etc. The results also showed that the rule editor requires specific attention (Figure 5). In addition, the cognitive walkthrough, performed concurrently with the first phase of user observations, identified a number of recurring issues, chiefly concerning visibility, recognizability and understandability of controls and objects in the interface.

In Figure 2 an assessment protocol is shown as an example of an object with the previous version of the interface. Other objects followed the same color theme and design. In the case of assessment protocol, the interface did give some hints as to the structure of the content, by displaying the contents of a sub-protocol with indentation. However, in general there were many types of controls available, unsorted by importance or suggested workflow. This was true for other objects in the system as well, and may have contributed to the lower scores in the AAIMA assessments in the first phase of evaluations.

These issues were addressed in the re-design following two overarching themes, namely providing objects in the interface with unique, recognizable and recurring appearances, and hiding certain controls and information when they are not needed. Objects (such as rules, critical questions, assessment protocols etc.) were given unique appearances that recur throughout the interface wherever said objects are used or referenced (Figures 3, 4). The workflow in the system was improved by removing

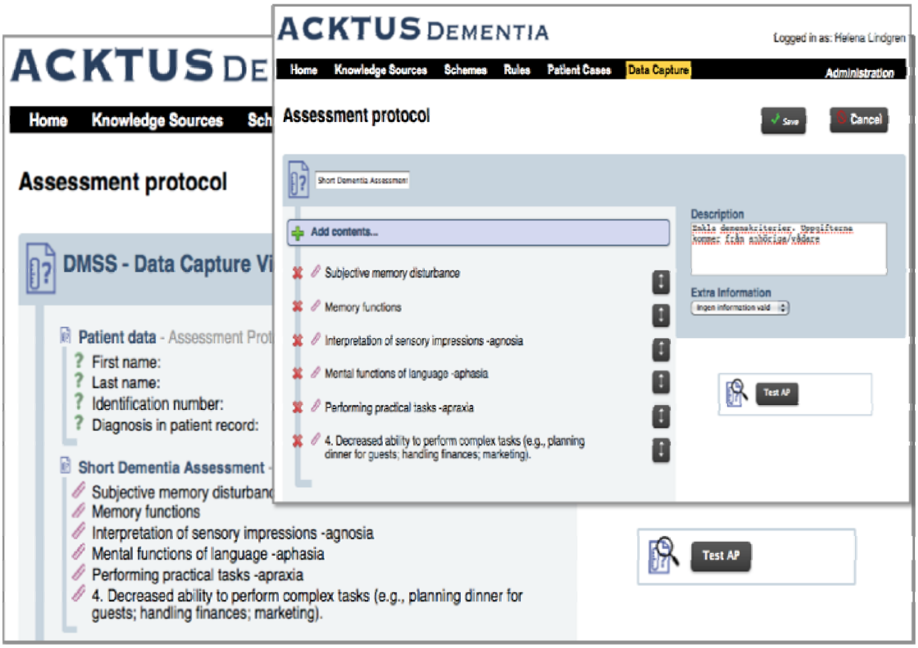


Fig. 3. Assessment protocols in normal and edit view created by User 3 in Phase 2

unnecessary steps and to provide more feedback at certain points in the interaction. The connection between controls and their affected components and between related objects were strengthened, and interface improvements were done to support a natural workflow, weighing content and information visually according to its relative importance to achieving the task at hand.

Some of the changes implemented for the second phase of evaluations are exemplified in Figure 3 and Figure 4. Two participants increased their performance in the managing of interaction protocols to Level 3, indicating a higher understanding of the role of the objects in question in the system as a whole. Also, novice users reached high levels of performance and understanding even from only a short introductory session. Changes such as the ones displayed in Figure 3 and 4 may have contributed to this increased understanding. For one, each object now has a separate color theme, which should help the user to identify and separate objects more quickly.

Another change that may have contributed to the users' increased understanding is the edit mode (Figure 3). Hiding editing controls when first viewing an object allows the user to focus on getting an overview and understanding its content before deciding what to do with the object.

Consistent use of colors, along with icons for each object, should help the user get a sense of "what goes where" in the system. Additionally, the spine to the left in the assessment protocol is meant to further emphasize the structure of it, again to assist the user in recognizing the relations between different kinds of objects (Figure 4, 5).

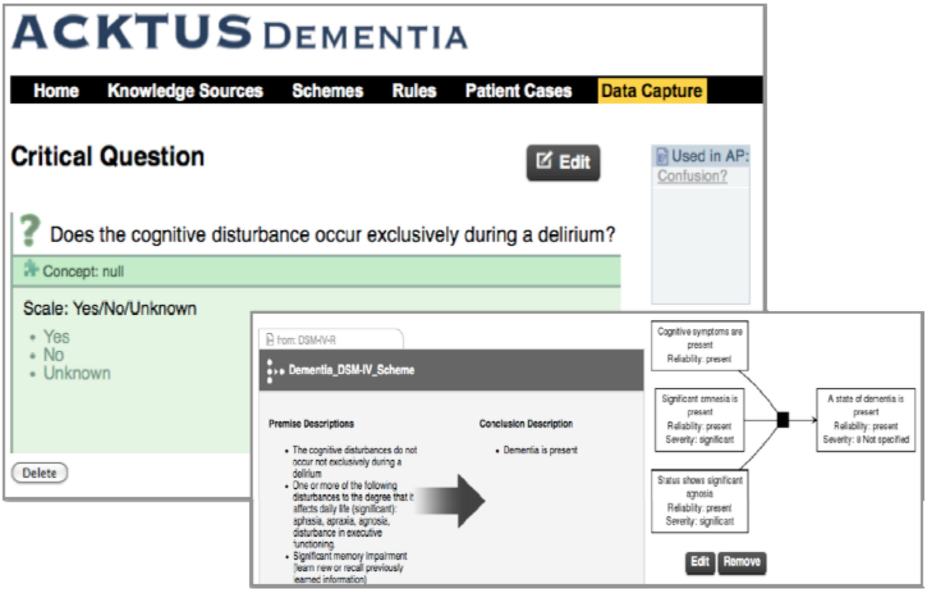


Fig. 4. Critical question (interaction protocol) view, and the scheme and rule view in Phase 2

Phase 2. In the second phase of user observations, the participants did reach Level 3 of the activity, some needing occasional aid from the evaluators (i.e. the activity lay within their ZPD), since they were also able to create rules within the system (Table 2). However, their proper understanding of the rules and their application was limited, and the concepts, terms and workflow used in the rule creation process still caused breakdowns, though these were handled by the individual in most cases. Still, a progress was seen in, e.g., User 1. In Phase 2 this user was able actually to complete rules that govern question presentation and order, with some guidance from the evaluator. This was not accomplished in the first phase of evaluation, when he only managed to edit parts of existing rules but not create his own rules.

The possibility to apply and evaluate the results in user cases was not included in the first phase of the evaluation study. In the second phase the participants were able to evaluate the content they had entered into assessment protocols by viewing it in a prototype of the end user interface, and from there enter responses in the manner of specific types of users, in order to see whether these fictive users would be shown the dedicated tailored content (e.g., Figure 5). This task generally caused no problems for the participants; indeed, a few of them even chose to view the content as the end user would see it, even when editing basic content.

3.2 Outcome of the Evaluation Sessions in Terms of Developed Applications

During the evaluation sessions, the participants continued with tasks related to their own work and were able to complete tasks yielding authentic results that will be usable in practice in the future.

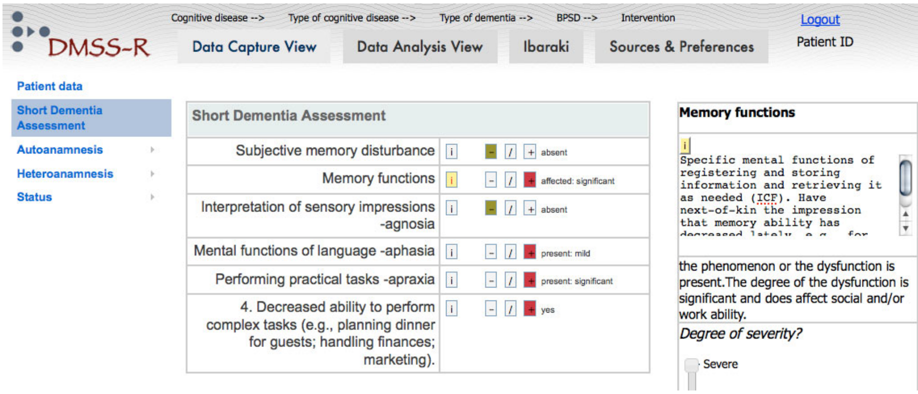


Fig. 5. DMSS end user prototype visualizing the short dementia assessment protocol

For instance, in Phase 2 User 1 worked with two assessment protocols that will be used for assessing environmental health risk factors, concerning vibration specifically, for end users of the EKF subsystem. User 1 was able to review an extensive amount of questions, correcting mistakes such as spelling errors, reviewing the order the questions were presented in when answering a specific way, and editing these factors when necessary. For presentation purposes, it was also necessary for the user to create sub-assessment protocols with which to group the questions. All in all, an assessment protocol that had been developed was fine-tuned and made presentable during the participant's evaluation session. Another assessment protocol was reviewed and adjusted, giving information about what remains to be done. What caused the user the most pause was managing the rules used for deciding question order and presentation conditions.

In the case of User 3, this physician had been involved in developing a knowledge base for dementia integrated in a stand-alone prototype, subjected to evaluations with medical professionals in Asia. Following their request of a quick assessment protocol that would fit their workload and resources (preferably to be finished within three minutes, which is, however, in practice impossible), he modeled during a single session a minimal assessment protocol for diagnosing dementia by reusing components integrated in the full assessment protocol. He was able to redesign the assessment flow with some assistance from a knowledge engineer to follow the associated clinical guideline using rules, and modeled critical questions to appear in the user interface for guiding the end user. He also combined the outcome with an additional short protocol for assessing Alzheimer's disease. This was accomplished within 30 minutes. This 3-minutes protocol was seen as a guided entrance into the full protocol in the complex patient cases, and as such it was seen as a major contribution to the development and tailoring of the dementia support system to different user needs (see Figures 3-5).

In the case of User 4 and User 5, the setting was an initiation of a collaborative project where User 5 modeled content she will have use of in an end user situation involving older adults. User 4 functions as the leader of their common project and had a larger perspective on the use of ACKTUS in their setting, leading to her to envision a diagnostic tool that integrates research-based knowledge sources as well. Following her idea, she modeled general knowledge for this purpose, supplementary to the knowledge modeled by User 4. The modeled content was anticipated to be used by older adults in a dialogue system, where parts could be re-used and possibly adapted by other professionals in the domain.

An example of the interaction designs produced by the participants during the evaluation period is shown in Figure 5. To summarize, the results of the modeling done by the participants are accustomed, personalized interaction flows adapted to given end users' needs and requests. For instance, in the case of DMSS (Figure 5), the minimal assessment protocol for diagnosing dementia gives the end user the option of using this protocol to quickly assess a patient's mental state and only continue to the larger, more detailed protocol if they decide that this is necessary. In the case of EKF, the questions and information shown to the end user is decided by their previous responses to questions, according to the rules defined by the participants. At the end of the dialogue, the end-user receives personalized advice based on their answers.

4 Discussion

Defining assessment protocols generally caused little problems for the users in Phase 2. While the users did reach Level 3 in both phases, User 2 and User 4 faced breakdown situations during phase 1 and the activity lay within their ZPD. This was not the case in the second phase. One reason for this increase in skills could be the major revisions in the interface that underwent between the two evaluation phases (see Figures 3-5). This is also the case for the task of creating interaction protocols, which had received improved functionality and presentation since the first phase, most likely reflected in the increased performance for the users 1, 2 and 4.

One notable change is the increased performance in the task of creating rules. User 1 and User 2 increased their score from barely being able to edit rules or not wanting to participate, to being able to create new rules with some guidance in phase 2. The beginner users (User 4 and User 5) also scored high in the second phase. However, since only parts of the intended interface changes for managing rules had been implemented at the time of the second phase of evaluation due to time constraints, this change likely has other causes as well, such as an increase in motivation or learning. The EKF users may for instance have gained insight into potential applications of rules and therefore were more motivated to learn about them before and during the second phase. These two users' initial perception of the modeling tasks was that they would describe the content and a programmer then would implement the content. Their attitude changed when presented with the system and when starting to model the content, though with some delay. Another motivational factor for the two EKF users was probably the approval of, and encouragement for, the inclusion of personal advice for the end users, coming from a project reference group, which had not been approved at the time for Phase 1. They put large interest in their last sessions on modeling personal advice to appear for end users, which were to be demonstrated for the reference group. User 4, who scored 3 when creating rules, immediately saw potential applications for the rules and the system in their line of work and as such was highly motivated to test and learn.

User 4 expected a different procedure for relating the assessment protocols to the rules, making her somewhat unwillingly adjust to the procedure implemented in ACKTUS. It will be investigated whether this alternative procedure should be supported by the system in the future.

Since a convenience sample was used for selecting the participants (i.e., all professionals involved in modeling tasks during the evaluation period) and due to the fact that the sample contains only five participants, the results are only representative for the group. However, the results indicate that it is possible to learn and use the re-designed system after only a short introduction with the correct motivation grounded in a need for tailoring a knowledge system to specific users or user groups.

Adapting and applying the AAIMA protocol yielded a structured baseline for evaluation and for a continuing evaluation during the ongoing development of ACKTUS and the different end user applications. It also gave insight into a user's comprehension and capability in using the system by incorporating elements of activity theory, such as ZPD. The results from using the protocol to score user performance provided a basis for a structured comparison both between different users and between different sessions with the same user. From this, the development and learning of a participant using the system can be extrapolated, providing insight into areas of concern in the interface and how well the system supports learning. The protocol did not eliminate the subjectivity in determining whether, for instance, a user understands the relationship between different kinds of objects, but as long as the activity levels are interpreted and applied consistently according to the same criteria, the subjectivity is reduced. A way to further investigate its inter-rater reliability would be to have several evaluators apply it to the same set of data and compare their results. The protocol was originally used to assess the skills in computer-mediated activity in elderly users [5], and has in this study been used with participants from the medical professional domain. In this study it was shown that the protocol is useful for a significantly different domain, provided the protocol is adapted based on an activity analysis of the application domain. This adaptation need may prevent the application of the protocol in cases where there are limited resources for assessing the use context. Furthermore, the protocol may show less applicability also in the cases where an actual use situation with dedicated end users cannot be arranged for evaluation. The major contribution with the AAIMA protocol is that it supplements other methods in that it supports a structured theory-based interpretation of the data that can be collected e.g., in a study of focus shifts in activity execution [17]. In addition, it allows for detecting changes in knowledge and behavior over a longer period of time.

5 Conclusions

A pilot evaluation study was conducted of ACKTUS, an early prototype for modeling knowledge and interaction design of knowledge-based support systems to be used in the health domain. The purpose was to investigate how well the system supports knowledge modeling and interaction design for domain experts in authentic use situations, as well as their development of a conceptual understanding of the system. Consequently, we focused on users with minor or no experience from using the application. Five health-care professionals with expertise in four different domains were observed using the system in 1-3 sessions each, and were interviewed. All participants were working on particular applications relevant for their domain as part of three different research projects. The concept of “zone of proximal development” (ZPD) was used to assess the development of skills, following the AAIMA assessment protocol [5] adapted to the result from an activity analysis done of the activity in focus for evaluation. The protocol proved to be a valuable instrument for detecting and describing the development of the users’ skills and understanding of the system. This evaluation also demonstrates the applicability of the protocol to a new user group and a new type of system than it was originally designed for.

Between the phases the ACKTUS user interface was re-designed based on the results from the initial phase. The results show that the users who participated in both phases increased their understanding and skills, and that the two new users who entered in Phase 2 were able to reach the same level of understanding as the three who had used the system in Phase 1. In Phase 2 they were able to model components and interaction design and had gained a sufficient overall understanding of the system. The re-design of the user interface was interpreted as the major reason, and the fact that a certain amount of learning took place between sessions was interpreted as a minor factor. A more important factor was the motivational level in the users. These results illustrate the development of understanding, and that the ACKTUS application does not require extensive training and learning to use, in contrast to other approaches for integrating medical knowledge presented in the literature.

The results are fed into further re-design of the user interface, with the aim at providing novice users of the ACKTUS modeling environment a transparent and intuitive interaction design. When this is accomplished, also end users may contribute with knowledge as part of a community of users sharing a common living knowledge repository. The AAIMA protocol will be used as baseline in future evaluation studies, enabling the detection of development of skills and understanding. The protocol is also being integrated into ACKTUS for developing adaptive functionality using automated methods for detection of activity and skills.

Acknowledgments. The authors would like to thank the professionals participating in studies and development work. The project is partly funded by VINNOVA (Sweden’s innovation agency) and Emil and Wera Cornell foundation.

References

1. Kaplan, B.: Evaluating informatics application – clinical decision support systems literature review. Int. J. Med. Inform. 64, 15–37 (2001)
2. Lindgren, H., Winnberg, P.: A Model for Interaction Design of Personalised Knowledge Systems in the Health Domain. In: Grasso, F., Paris, C. (eds.) Proc. 5th International Workshop on Personalisation for e-Health 2010, Casablanca, Morocco (2010)
3. Lindgren, H., Winnberg, P.: Evaluation of a Semantic Web Application for Collaborative Knowledge Building in the Dementia Domain. In: Proc. E-Health 2010, Casablanca, Morocco (2010)
4. Seyfang, A., Miksch, S., Marcos, M., Wittenberg, J., Polo-Conde, C., Rosenbrand, K.: Bridging the Gap between Informal and Formal Guideline Representations. In: Brewka, G., Coradeschi, S., Perini, A., Traverso, P. (eds.) ECAI 2006. Frontiers in Artificial Intelligence and Applications, vol. 141, pp. 447–451. IOS Press, Amsterdam (2006)

5. Lindgren, H.: Personalisation of Internet-Mediated Activity Support Systems in the Rehabilitation of Older Adults – A Pilot Study. In: Proc. Personalisation for e-Health 2009, pp. 20–27 (2009)
6. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., Stefanelli, M.: Comparing computer-interpretable guideline models: a case-study approach. J. Am. Med. Inform. Assoc. 10, 52–68 (2003), doi:10.1197/jamia.M1135
7. Isern, D., Moreno, A.: Computer-based execution of clinical guidelines: A review. Int. J. Med. Inform. 77(12), 787–808 (2008)
8. Shalom, E., Shahar, Y., Maimon, M.: Can physicians structure clinical guidelines? Experiments with a mark-up-process methodology. In: Riaño, D. (ed.) K4HelP 2008. LNCS (LNAI), vol. 5626, pp. 67–80. Springer, Heidelberg (2009)
9. Sutton, D.R., Taylor, P., Earle, K.: Evaluation of PROforma as a language for implementing medical guidelines in a practical context. BMC Med. Inform. Decis. Mak. 6, 20 (2006)
10. Eccher, C., Seyfang, A., Ferro, A., Stankevich, S., Miksch, S.: Bridging an Asbru Protocol to an Existing Electronic Patient Record. In: Riaño, D., ten Teije, A., Miksch, S., Peleg, M. (eds.) KR4HC 2009. LNCS (LNAI), vol. 5943, pp. 14–25. Springer, Heidelberg (2010)
11. Lindgren, H.: Towards personalised decision support in the dementia domain based on clinical practice guidelines. UMUAI, doi:10.1007/s11257-010-9090-4
12. Patel, V., Kaufman, D., Arocha, J.: Emerging paradigms of cognition in medical decision-making. Journal of Biomedical Informatics 35, 52–75 (2002)
13. Lindgren, H., Nilsson, I.: Designing Systems for Health Promotion and Autonomy in Older Adults. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 700–703. Springer, Heidelberg (2009)
14. Davidson, E., Heslinga, D.: Bridging the IT Adoption Gap for Small Physician Practices: An Action Research Study on Electronic Health Records. Information Systems Management 24(1), 15–28 (2007)
15. Kaptelinin, V., Nardi, B.A.: Acting With Technology: Activity Theory and Interaction Design. MIT Press, Cambridge (2006)
16. Vygotsky, L.: Mind in Society. Harvard University Press, London (1978)
17. Bodker, S.: Applying activity theory to video analysis: How to make sense of video data in HCI. In: Nardi, B. (ed.) Context and Consciousness. Activity Theory and Human Computer Interaction, pp. 147–174. MIT Press, Cambridge (1996)

Identifying Relationships between Physiological Measures and Evaluation Metrics for 3D Interaction Techniques

Rafael Rieder[1,2], Christian Haag Kristensen[1], and Márcio Sarroglia Pinho[1]

[1] Pontifical Catholic University of Rio Grande do Sul, Av. Ipiranga, 6681, Partenon
CEP 90619-900 Porto Alegre, RS, Brazil
[2] University of Passo Fundo, BR 285, Km 171, São José
CEP 99052-900 Passo Fundo, RS, Brazil
rafaelrieder@gmail.com,
{christian.kristensen,marcio.pinho}@pucrs.br

Abstract. This project aims to present a methodology to study the relationships between physiological measures and evaluation metrics for 3D interaction techniques using methods for multivariate data analysis. Physiological responses, such as heart rate and skin conductance, offer objective data about the user stress during interaction. This could be useful, for instance, to evaluate qualitative aspects of interaction techniques without relying on solely subjective data. Moreover, these data could contribute to improve task performance analysis by measuring different responses to 3D interaction techniques. With this in mind, we propose a methodology that defines a testing protocol, a normalization procedure and statistical techniques, considering the use of physiological measures during the evaluation process. A case study comparison between two 3D interaction techniques (ray-casting and HOMER) shows promising results, pointing to heart rate variability, as measured by the NN50 parameter, as a potential index of task performance. Further studies are needed in order to establish guidelines for evaluation processes based on well-defined associations between human behaviors and human actions realized in 3D user interfaces.

Keywords: usability metrics, physiological measures, interaction techniques.

1 Introduction

In order to evaluate the characteristics of three-dimensional user interfaces (3DUI), like presence and immersion, methods and tools commonly used to evaluate two-dimensional user interfaces can be applied, such as prototypes, questionnaires and formative and summative tests. These instruments are able to get relevant usability metrics also in 3DUIs, like variables to measure system performance, user task performance and user preferences. The first two measures result in objective data for assessing, respectively, the computer or graphics system performance, and the quality of performance of specific tasks in the 3D application. The third measure results in

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 662–679, 2011.

subjective data for assessing the user satisfaction while using an interface. The evaluation of these measures is important because they allow that different 3DUI elements, such as spatial perception, multimodal interaction and sensory stimulus are considered.

However, the adaptation of these tools to evaluate 3DUIs can lead to incomplete assessment of the particular characteristics of these applications, such as the use of non-conventional devices and 3D interaction techniques (ITs) [6]. These characteristics tend to influence the user performance and the user satisfaction, which requires a process to evaluate its various resources based on the user experience level.

A recent alternative used to evaluate interfaces it is the use of the physiological measures. According to Malik *et al* [22], the physiological monitoring provides information about the user's physiological balance, and its measures are associated with stress. Researches in the Virtual Reality (VR) area have been using this type of measurement to assess the user's physical and mental effort on the 2D games [17][18][19][20] and to evaluate presence and user comfort in immersive virtual environments (VEs) [7][8][16][23][24]. However, there are no studies about the relationship between physiological measures and metrics focused on the evaluation of the quality of ITs.

The use of physiological measures can still address other two classical problems in the evaluation of 3DUIs. The first concerns the collect of objective measures, which in some cases require modifications in the source code. In these situations, it is not always possible or desirable to alter an application, due to the complexity of the system [2] or the limited availability of development time [28]. The second problem concerns the reliability of subjective metrics, which may have influenced their results by external factors to the interaction process, such as user's physical and mental efforts, or user's cognitive mediation, such as omission or summarization of information.

Physiological measures, therefore, offer objective responses that are not controlled by the person, they are associated with factors such as fatigue and irritation, and provide data related to the organism's behavior. These are measures that can indicate, for example, the adaptation periods to a new device or new IT, because the user's stress level can be viewed along the timeline. Besides, they can aid in the comprehension of the performance results and answers of questionnaires. Doing so, physiological responses may complement the current methods of assessment [5][14][25], allowing the understanding of the interaction process as a whole, and contributing to increasing the quality of the VR applications.

In order to evaluate whether physiological measures may be or not a substitute for objective and subjective usability metrics, statistical methods of multivariate data analysis can be used. According to Hair *et al* [13], these methods allow analyzing simultaneously the influence of multiple measures on each subject or variable under investigation, regardless of the complexity or the context in which these variations occur. So, it is possible to verify if results obtained with physiological measures are able to indicate the same problems identified by traditional usability metrics.

Therefore, this work describes a methodology for assessing the quality of ITs in immersive VEs, comparing physiological measures and evaluation metrics for 3DUIs using multivariate data analysis. Our methodology contemplates the use of a testing protocol, data normalization and exclusion processes, and statistical methods for exploratory data analysis and regression analysis, in order to discover relationships between variables that contribute to the evaluation process, complementing or assisting in the interpretation of results. In the same way, our methodology also determines the physiological measures able to indicate the same results expected by traditional usability measures, and these may eventually replace usual measures in projects in which the simplification of the testing stage is desirable. So, it is possible to reduce the dependency on subjective data, and to avoid changes to collect performance data in complex software.

This paper is organized as following: Section 2 presents the related work, whereas Section 3 describes the developed methodology. In Section 4, a case study is presented to evaluate the use of this approach. Section 5 shows a discussion about the results obtained using this methodology. Finally, Section 6 concludes the paper highlighting the potential of our approach.

2 Related Work

The use of physiological measures is a recent alternative in the evaluation of graphical interfaces. Latest researches apply this resource to measure presence and cybersickness in immersive VEs, and user's physical and cognitive effort in videogames.

Meehan *et al* [23][24] used physiological monitoring to measure presence in a stressful VE. These researches used heart rate (HR), skin conductance (SC) and skin temperature measures in four different experiments to compare participants' physiological reactions to a non-threatening virtual room and their reactions to a stressful virtual height situation. According to the authors, HR satisfied all the requirements for a reliable, valid, sensitive and objective measure of presence in a stressful VE. In addition, HR showed correlation with the well-established presence questionnaire. SC had some of the properties desired to measure presence, and skin temperature did not.

Slater *et al* [26] reported the difficulty to measure presence subjectively, and highlighted the need to evaluate alternatives that provided objective data to overcome, or, at least, supplement the use of questionnaires. With this in mind, the authors conducted an experiment to explore the relationship between physiological responses, breaks in presence and utterances by virtual characters towards the participants using a virtual bar scenario. The results showed that changes in HR and SC point to occurrence of breaks in presence during the interaction process. Changes in HR also indicated the moments when an avatar speaks to the subjects, whereas heart rate variability (HRV) parameters pointed to differ between participants with different social anxiety scores, classified by questionnaire.

Brogni *et al* [7][8] followed a similar approach to previous work. They studied the use of physiological responses to determine the impact of visual realism and the user stress level during the interaction in an urban VE. In this scenario, the sense of presence was subjected to the texture quality and the appearance of avatars. The results showed that HRV parameters indicated the reducing of level stress as time goes by. The authors also reported that these parameters were associated with the level of visual realism based on the texture strategy used in their experiments.

Kim *et al* [16] presented a study about the use of physiological measures to reduce cybersickness in immersive VEs. Their approach proposed a system based on 11 physiological signals, which detected the cybersickness and automatically reduced the user's field of view and slowed the travel velocity in the VE. The results indicated a significantly lower frequency of cybersickness when users use this system, compared to a situation in which it is not used. The authors also highlighted the gastric tachyarrhythmia as a physiological measure of a better outcome.

On the other hand, Lin *et al* [17][18] studied the use of physiological measures as a metric to evaluate the usability of video games. The experiments considered the use of HR, SC and blood pressure (BVP) measures to assess the user's performance and satisfaction while tasks were realized in three different difficulty game levels. The authors reported the SC measure as relevant in comparisons between game levels and performance user groups, indicating that a greater number of errors and difficulty consequently increases the SC. This work also studied the relationship between physiological measures and frustration events, which showed a range of more than 5% in SC in 70% of the frustration cases analyzed. Other measures did not showed significant results.

Complementary studies of Lin *et al* [19][20] added measures like pupillary response, eye tracking and HRV parameters to assess the user cognitive efforts in videogames. The results also showed the relationship between HRV parameters and game levels. Pupillary response and eye tracking presented promise results, but it was not consistent as SC and HRV measures.

From the literature review, it can be noticed that there are no studies exploring the relationship between physiological measures and usability metrics used to evaluate the quality of three-dimensional ITs in immersive VEs. The following section presents a methodology designed for this purpose.

3 Methodology

Firstly, the following sections present the platform for testing, the physiological measures, the task performance measures and the questionnaires used by our approach. Secondly, we present the steps to apply our methodology, which include the use of a test protocol, normalization and exclusion methods for physiological measures, and statistical analysis methods.

3.1 Platform for Testing

In order to illustrate the use of our methodology, we built a virtual room with four numbered books, distributed on the floor inside the user's field of vision, as presented in Figure 1. Two well-known ITs also were implemented to select and manipulate objects: ray-casting [3] and HOMER [4]. These techniques were chosen because they are commonly used as parameters to evaluate new ITs. With ray-casting, the user selects and manipulates objects using a virtual light ray, with the ray's direction specified by the user's hand. With HOMER, the user selects an object using the ray-casting technique, and manipulates it using a virtual hand, which instantly moves to the current object's position and attaches to it.

The user's task is to get the books, turning them as necessary, and place them in transparent areas marked on the floor. The books must be organized with their front cover visible to the user, and their spine toward to the left side of the virtual room. At the end of task, the books will be placed in ascending order, from left to right. The VE also provides two visual feedbacks to indicate when an object is ready for selection, and an aural feedback to inform collisions occurred into the virtual room.

This application was built using C++, OpenGL, GLUT, irrKlang[1], and the SmallVR[2] toolkit, which simplifies the development of VR applications by abstracting many implementation aspects such as device control and scene graph management, while maintaining the GLUT structure for the program.

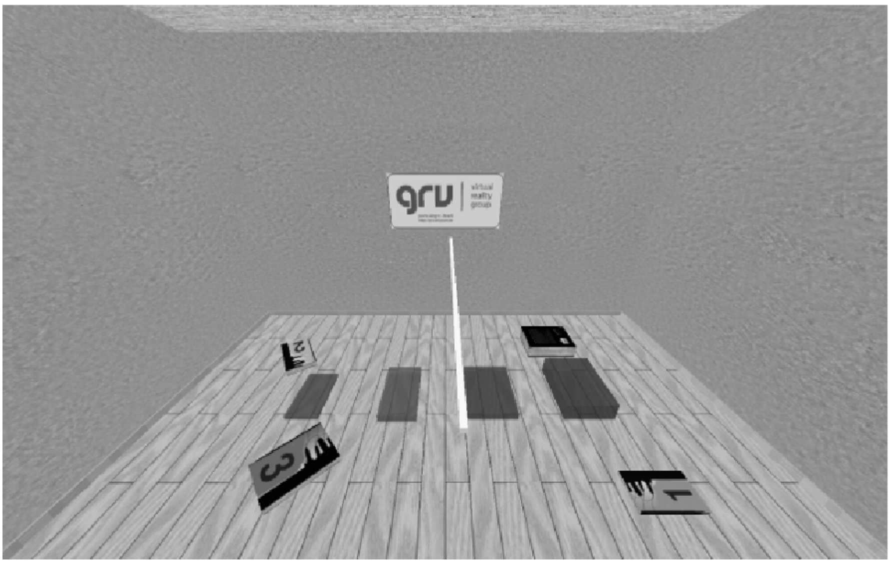


Fig. 1. The virtual room application built to our experiment

In order to explore the VE and use the ITs resources, we used i-Glasses Head Mounted Display (HMD) and Polhemus FastTrak motion tracker with two tracking points enabled for interaction. The first tracking point was used to track the user's head movements, whereas second was used in the user's dominant hand to select and manipulate objects. Grab and release user's actions were confirmed using a push-button attached on the second tracking-point.

The physiological monitoring used an electrocardiogram (ECG) sensor and a SC sensor, on a non-invasive way. These equipments were connected to the Procomp Infiniti encoder, which captured and sent the physiological responses to the Biograph Infiniti software for data processing. Three electrodes connected to the ECG sensor were fixed on the user's wrists with rubber wrist straps, whereas two electrodes connected to the SC sensor were strapped to two fingers of user's non-dominant hand using finger bands. All these solutions are manufactured by Thought Technology[3].

[1] http://www.ambiera.com/irrklang
[2] http://grv.inf.pucrs.br/projects
[3] http://www.thoughttechnology.com

Figure 2 presents the device's default configuration, highlighting the push-button attached to one of the tracking-points, and positions of the ECG and SC sensors.

Fig. 2. Subject wearing physiological and VR devices during the experiment

3.2 Physiological Measures

For this work we used the HR and SC physiological measures, collected by ECG and SC sensors, respectively. Our approach also includes the use of seven different HRV measures, generated by time domain and frequency domain methods from short-term 5-minutes recordings, according to Malik *et al* [22].

This way, the following physiological measures were selected:

– Mean SC, expressed in micro-Siemes (μS);
– Mean HR, expressed in beats per minute (bpm);
– Standard deviation of the NN[4] interval (SDNN), expressed in milliseconds (ms);
– Number of interval differences of successive NN intervals greater than 50 ms (NN50);
– Proportion derived by dividing NN50 by the total number of NN intervals (pNN50);
– Mean total power in very low frequency range (VLF), expressed in milliseconds squared (ms^2);
– Mean total power in low frequency range (LF), expressed in ms^2;

---

4 On an ECG recording, NN are the normal-to-normal intervals between adjacent QRS complexes.

– Mean total power in high frequency range (HF), expressed in ms^2;
– Ratio between the LF and HF measures (LF/ HF).

3.3 Task Performance Measures

In order to get the task performance during the experiment with ITs, the following measures were defined. These data were collected in real time by the own VE application:

– *Total time*: entire period of the interaction process, measured in seconds (s);
– *Accuracy*: distance between the center of an object and the center of its target position[5], measured in meters (m);
– *Collisions*: amount of collisions occurred with an object;
– *Grabs*: number of attempts to grab an object.

3.4 Questionnaires

Two questionnaires were created for this work. The pre-test questionnaire obtained demographic information and user's past experiences with VEs, whereas the post-test questionnaire evaluated the user preferences in relation to the interface, tasks, and ITs. A progressive scale of 1 to 7 was set to evaluate each question in both instruments, from a lower to a higher concept.

Pre-test questionnaire asked about the age, gender and education of each participant, and was composed of five questions. These questions were about the VR level of knowledge, the number of times there was presence in VE experiments, the number of times there were physiological measures, the sense of discomfort when interacting with new interfaces, and the sense of discomfort when interacting with graphical interfaces like games and VEs.

Post-test questionnaire was composed of seven questions. The aim of these questions were to evaluate the influence of wired devices in the user's performance, the level of irritation generated by the fact of requiring to the user pick up again the objects after each collision, the level of pressure generated by the time limit to complete the task, the level of satisfaction in using the ray-casting and the HOMER techniques, the level of confidence in performing the task correctly, and the level of satisfaction with the visual and aural feedbacks presented by the VE. Moreover, this questionnaire also provided areas to describe the occurrence of discomfort situations, and the moments of more irritation or more difficulty during the interaction process.

3.5 Test Protocol

In order to test our methodology, we designed a protocol based on related work presented in Section 2. Our approach supports steps to the simultaneous collection of task performance, subjective evaluation and physiological data during the experience.

The test protocol was established with nine stages, which can be executed in approximately 45 minutes, in the following order:

[5] The target position means the desired position of the object in the VE.

- Apply pre-test questionnaire;
- Prepare devices;
- Collect baseline data;
- Start training – first IT;
- Start experiment – first IT;
- Start training – second IT;
- Start experiment – second IT;
- Release devices;
- Apply post-test questionnaire.

"Apply pre-test questionnaire" stage contemplates the trainer and experiment presentations, the distribution of informed consent form to read, and the filling of pre-test questionnaire. After this it is provided to read the VE instructions, which explains how to execute tasks using the two ITs. The estimate duration to finish this stage is eight minutes. According to Kim *et al* [16], this period is also important to stabilize the physiological responses before next stages.

"Prepare devices" stage considers the arrangement of physiological sensors and VR equipments in the user's body. Firstly, the user is invited to turn off electronic devices, to remove watch and bracelets and to accommodate in a chair, sitting in a comfortably way. After this the trainer cleans the user's wrists with alcohol gel, and applies a conductive gel into the ECG sensors to reduce noise caused by electrical resistance of the skin. The physiological sensors are fixed as mentioned in Section 3.1 and, finally, the user wears the HMD. For this stage, we estimate approximately three minutes to complete it.

"Collect baseline data" is started after the user feels comfortable with the devices. The trainer requests the user to put his/her arms on his/her legs in a rest position. In this step, VR devices remain turned off, and it is asked for the user to keep his/her eyes open. Based on Slater *et al* [26] and Brogni *et al* [8] collecting physiologic data is done in two minutes and it will be the comparison base for the interaction process.

In the next stage, "Start training – first IT", the subject begins to interact with the VE to learn to use the devices and the first IT. Selection and manipulation tasks can be performed during three minutes.

Again, the subject interacts with the same VE and IT in the "Start experiment – first IT" stage, and must perform all tasks in seven minutes. After this period, the subject has three minutes to rest with all devices off.

The same procedures are applied to the next two stages for training and experiment of the "second IT".

It is important to highlight for an unbiased analysis, the use of two techniques must be balanced according the number of participants in an experiment. For this reason, our protocol divides the experiment between "first IT" and "second IT".

The above times, defined to complete the training and experiment stages, are based on Lin *et al* [17][18][19][20] and Brogni *et al* [7][8] works, whereas the interval to rest between stages is based on Kim *et al* [16].

During the "Release devices" stage, VR and physiological equipments are removed from the user. The trainer applies procedures to clean the user's wrist and devices, using dry wipes and dusters. The estimate time is four minutes to do it.

At last, “Apply post-test questionnaire” stage contemplates the subjective evaluation of test, which the user is invited to answer the post-test questionnaire. A brief period also is addressed for comments and thanks. For this stage, we estimate approximately five minutes, based on Slater *et al* [26] and Lin *et al* [20] works.

3.6 Normalization and Exclusion Methods for Physiological Measures

In order to statistically compare physiological responses and different usability variables, it is necessary to define data normalization and exclusion methods to the HR and SC measures.

For the SC data normalization, we adopted a scale of 0 to 1, which it attributed the minimum value of 0 to a lowest SC value, and the maximum value of 1 to a highest SC value. This procedure was applied to each user’s SC signal, generated a new and normalized Mean SC measure. So, SC values became uniform and preserved the individual characteristics of each subject.

By contrast, for the HR measures we needed to apply a procedure to exclude some data, because the adopted way to collect this physiological response is susceptible to generation of noises in the HR signal. The procedure eliminated participants who presented HR values outside the normal range for a human. The exclusion criterion was executed in the following order:

- HR rest: based on the baseline data, subjects were excluded from the dataset when their mean HR was below 60 bpm or above 100 bpm. According to Guyton and Hall [12], typical healthy resting HR in adults is 60–100 bpm;
- HR max: subjects were excluded from the dataset when their mean HR during the experiment was above to the maximum HR, which it was estimated from the Tanaka formula [27], presented by the Equation 1;
- HR target: subjects were excluded from the dataset when their mean HR during the experiment was above to the target HR, which it was estimated from the Karvonen method [15], presented by the Equation 2. In order to use this method, we determined an intensity level of 50% to the interaction task, since the physical effort during the interaction process can be considered within a moderate activity zone [1], as a result of the subjects being seated and performed spatial movements using their arms and head during the test.
- HR min: subjects were excluded from the dataset when their mean HR during the experiment was below 60 bpm.

$$HR_{max} = 208 - (0{,}7 \text{ x Age}) \tag{1}$$

$$HR_{target} = [(HR_{max} - HR_{rest}) \text{ x (Intensity level \%)}] + HR_{rest} \tag{2}$$

3.7 Statistical Analysis Methods

In order to verify the relationships between different measures, we chose to use multivariate data analysis methods. According to Hair *et al* [13], these methods are able to investigate, simultaneously, multiple measures about each subject or object under study.

This work adopted the following analytical steps:

- Apply methods for exploratory data analysis to summarize, test the normality, detect outliers of the data, and use techniques to verify correlations between variables. This approach allows to identify the consistency and distribution of the data, and avoid the redundant variables;
- Apply multiple regression techniques to generate prediction models, considering methods to select relevant variables and its coefficients of determination. This approach allows discovering what measures are associated to the task performance and subjective responses.

For the exploratory data analysis, we defined the following tests:

- Summarize data: Descriptive statistics;
- Normalization: Kolmogorov-Smirnov test;
- Outliers' detection: stem and leaf and box-plots;
- Correlations: Pearson's coefficient to linear relationships, and Spearman's coefficient to nonlinear relationships.

The stepwise regression was chosen to create regression models and selects the predictor variables. In this process, each regression model may have one or more predictor variables and their coefficient of determination (r^2). These coefficients inform the power of these measures have to explain the variability of results. In this project, these results indicate whether a physiological measure can substitute or not a traditional usability measure. An analysis of variance (ANOVA) is also applied to test the significance of the regression model.

In order to generate regression models, our analysis selected only physiological measures with results statistically significant in the correlation tests ($p < 0{,}05$).

4 Case Study

In order to evaluate the effectiveness of physiological measures as indexes of quality to 3DUIs, this section aims to present a case study using our methodology, according to the definitions shown in Section 3.

Our evaluation included 54 healthy participants, 28 men and 26 women aged between 17 and 57 years old. Their tests were scheduled during a period of two weeks.

The subjects were also distributed into two equal groups (14 men and 13 women), in order to balance the use of ITs. The group "A" used as first IT the ray-casting technique, whereas the group "B" used as first IT the HOMER technique.

4.1 Relationships between Physiological and Task Performance Measures

According to the Section 3.7, it is necessary to apply a set of multivariate data analysis methods to assess the relationship between physiological and task performance measures, as presented in Sections 3.2 and 3.3. Thus, it is possible to identify which physiological responses are able to indicate task performance, or whether they can at least assist the interpretation of the results.

First of all, we used the testing protocol to collect the physiological, task performance and user preferences measures. After this we applied the normalization and exclusion procedures, in order to adjust our physiological datasets for the statistical comparisons. During this last stage, we detected some abnormal HR measures in 22 subjects, which needed to be discarded. Because of this situation, the original dataset had to be subdivided into two new groups: a dataset for SC measures, which included all the experiment participants (54 subjects), and another dataset for HR and HRV measures, which included only 32 subjects.

In the next stage, we applied the statistical methods for exploratory data analysis and multiple regression, looking for physiological measures able to indicate task performance.

Since our methodology was applied, two physiological measures (NN50 e HF) had a statistically significant relationship with two task performance measures ("Total time" and "Accuracy"). However, only one of these relationships indicated, on the regression model, strongly statistically significant results by both techniques ("Total time" x NN50, $p < 0{,}01$), as presented in the Table 1.

According to the results of the Table 1, the "accuracy" task performance measure only has statistically significance with the NN50 and HF physiological measures, for experiments using ray-casting technique.

Table 1. Regression models for physiological and task performance measures with strong correlation

Interaction Techniques	Task Performance Measures	Physiological Measures	Regression r^2 (%)	ANOVA F-test	p(value)
HOMER	Total Time	NN50	28,83%	12,15	0,00**
	Accuracy	NN50	7,26%	2,35	0,13
	Accuracy	HF	2,35%	0,72	0,59
Ray-Casting	Total Time	NN50	61,98%	48,91	0,00**
	Accuracy	NN50	43,28%	22,89	0,00**
	Accuracy	HF	31,33%	13,69	0,00**

On the other hand, NN50 physiological measure may be considered as the variable with the most associated with the "Total time" measure, because results were strongly significants for both experiments, independently of two techniques ($p < 0{,}01$). Based on Table 1, the NN50 physiological measure is able to indicate the user task performance, for the "total time" measure, with a statistical power (r^2) of 61,98% to the experiments using ray-casting technique, and 28,83% to the experiments using HOMER technique.

We also generated a regression model using the NN50 and "Total time" means, in order to join the statistical power of the selected physiological response in a single model, independently of two techniques. The result showed a coefficient of determination of 45,16% (ANOVA, $p < 0{,}01$, $F = 24{,}70$).

However, our results presented intermediary statistical power values. The coefficients of determination showed values far from an index close to perfect correlation ($r^2 = 100\%$), showing that the variance of NN50 measure cannot explain, alone and exactly, the variance of "Total time" measure. In other words, we can say that the NN50 physiological measure still cannot be used to replace the "Total time" measure during a task performance evaluation.

4.2 Relationships between Physiological and User Preferences Measures

Using the same approach presented in the Section 3.7, we also applied a multivariate data analysis to verify the relationships between physiological and user preferences measures, already presented in the Sections 3.2 and 3.4. In this study, we also used the data normalization and exclusion procedures, defined in the Section 3.6, which subdivided our dataset in two new sets (54 subjects for the SC measure, and 32 subjects for the HR and HRV measures).

In order to compare the questionnaire answers and physiological measures, we generated new physiological measure means from the values of experiences using the two ITs. Results can be visualized in the Table 2.

Table 2. Regression models for physiological and user preferences measures with strong correlation

Questionnaires	User Preferences Measures	Physiological Measures	Regression	ANOVA	
			r^2 (%)	F-test	p(value)
Pre-test	Question 1	SC	11,44%	6,71	0,01*
	Question 1	NN50	15,26%	5,40	0,03*
	Question 2	NN50	13,82%	4,81	0,03*
	Question 2	LF	10,65%	3,57	0,07
	Question 2	HF	7,50%	2,43	0,13
	Question 4	SC	10,49%	6,09	0,02*
Post-test	Question 4	HR	12,83%	4,42	0,04*
	Question 4	NN50	17,02%	6,15	0,02*
	Question 4	HF	23,39%	9,16	0,01*
	Question 7	LF/HF	13,72%	4,77%	0,03*

Firstly, tests were applied to verify the relationships between physiological measures and pre-test questionnaire answers. In this study, NN50 and SC physiological measures presented statistically significant relationships with questions addressed the level of knowledge about VR (Question 1), the experience with non-conventional devices in VEs (Question 2) and the tendency to feel discomfort or irritation when interact with new interfaces (Question 4).

Based on these analysis, we may note that the regression models presented in the Table 2 showed significant results ($p < 0,05$) for SC and NN50 measures as predictors of assessment, but with low statistical power for the Question 1 (SC, $r^2 = 11,44\%$; NN50, $r^2 = 15,26\%$), Question 2 (NN50, $r^2 = 13,82\%$) and Question 4 (SC, $r^2 = 10,49\%$). In this way, we can say that these physiological measures – when solely used – still cannot be employed to indicate the user level of knowledge about VR, the user level of experience in VEs, and the user level of irritation during learning in new graphical interfaces.

Secondly, we applied the same tests to verify the relationships between physiological measures and post-test questionnaire answers. In this study, we did not compare the physiological measure means and the answers related to the Questions 4 and 5, because these questions aimed to evaluate the ITs, separately. In this case, the Question 4 responses were compared with physiological measures collected during the user experiences with the ray-casting technique, and Question 5 responses were compared with physiological data of the experiences using HOMER technique.

This study presented only one physiological measure (LF/HF) with statistically significant relationship ($p < 0,05$) for the evaluation about the quality of visual and aural feedbacks displayed during the interaction process (Question 7). However, only 13,72% of the variance of LF/HF can explain the variance of the Question 7 responses. So, we can say that the LF/HF is not able to indicate this item evaluation.

Comparisons between physiological measures and Questions 4 and 5, which aimed to evaluate the level of satisfaction in using the ITs, presented statistically significant relationship ($p < 0,05$) only between HR, NN50 and HF measures and the Question 4 answers (ray-casting technique evaluation), as shown in the Table 2. The generated regression models also showed determination coefficients of low explanatory power (HR, $r^2 = 12,83\%$; NN50, $r^2 = 17,02\%$; HF, $r^2 = 23,39\%$), impossible to indicate the level of satisfaction with the use of ray-casting technique through physiological measures.

The post-test questionnaire also evaluated aspects about physical discomfort, sense of irritation and difficulties to perform tasks during the interaction process. The subjects' responses showed that these sensations did not affect the user's performance during the test, pointing no significant results.

5 Discussion

Based on the previous section, physiological measures still cannot be considered as indexes of user task performance and user preferences measures for VEs. From the results, the physiological responses only indicate user behavior tendencies during the interaction process, which can help as an additional resource to understand usability metrics during the evaluation process.

Comparisons between physiological and task performance measures only highlight the NN50 measure, which presented statistically significant results for both evaluated techniques and statistical power near an acceptable level to the "Total time" measure. According to the Section 3.2, the NN50 attests the amount of interval differences of successive NN intervals greater than 50 ms, which indicates the level of stabilization of the heart rhythm. Figure 3 shows a correlation plot between "Total time" and "NN50" measures, considering both experiences using the two ITs. This figure also indicates the trend of the NN50 increases as the "Total time" increases too.

This result can be interpreted on two different points of views. Firstly, experiments completed in less time show more concentrated subjects, which also use more physical effort to perform the tasks, compared with those who spend more time. On the other hand, we can say that the subjects spend more time learning to use the 3DUI, which ends up leaving them more relaxed and with their HRs in levels around a baseline measurement. Anyway, it is a promising measure to be explored and evaluated again in future research.

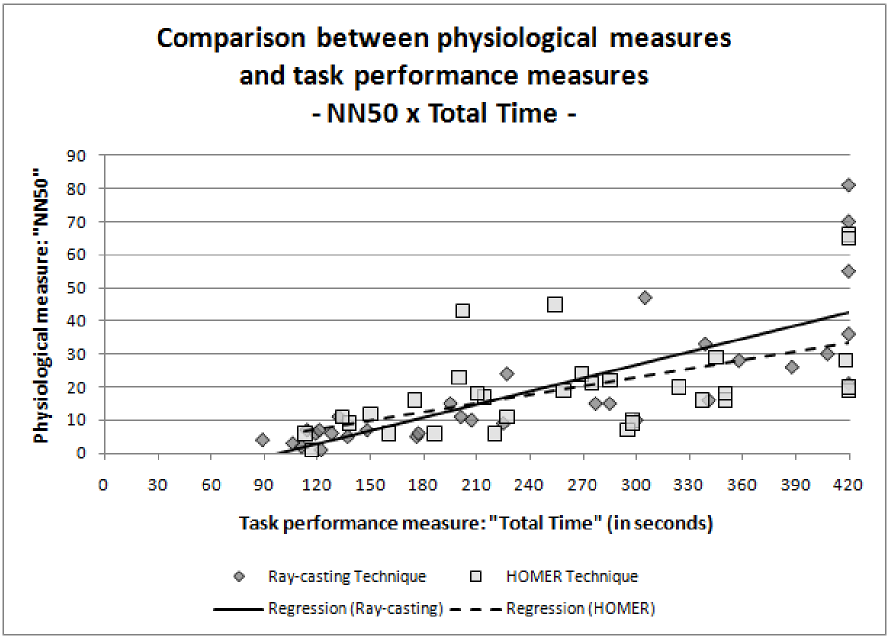


Fig. 3. Correlations and trends between "Total time" and "NN50" using the two ITs

Comparisons between physiological user preferences measures presented some relevant measures, such as SC, NN50 and LF/HF – but none of them showed significant statistical power. Probably, these low relationships can be associated with the first application of the pre- and post-test questionnaires, and their progressive scales.

With this in mind, we applied some tests to verify the questionnaire's reliability. Firstly, Cronbach's alpha [10] and split-half tests [11] were used, which presented the results shown in Table 3. According to Hair *et al* [13], the reliability coefficients indicate a good assessment tool when their index results in values above 0,7, which were not observed in this analysis. A probably reason for this result is the sample size used (54 subjects), which was below the literature recommendations, considering the scale dimension used by the instruments. As our scale ranged from 1 to 7, it would be necessary to apply our questionnaires to 70 subjects, at least, which guarantees obtaining more reliable results.

We also applied a factor analysis to the questionnaires to verify their relevance. According to Malhotra [21], it is recommendable to have the largest possible number of factors to this analysis, which are generated based on the number of questions, satisfying at least 60% of the total variance. According to the Table 4, for the pre-test questionnaire, 63,54% of the total variance can be explained using two factors, considering eigenvalues greater than 1,00. For the post-test questionnaire, Table 5 presents that 67,60% of the total variance can be explained using three factors, but only two of them have eigenvalues greater than 1,00. These results are also corroborated by low values of the Kaiser-Meyer-Olkin (KMO) tests, which examine the appropriateness of factor analysis.

These results recommend the review of the progressive scale used in our questionnaires, before new evaluation. Certainly, these changes will result in a more refined analysis to discover the relationships between physiological measures and subjective data.

Table 3. Reliability tests for the pre- and post-test questionnaires

Questionnaires	Cronbach's Alpha	Guttman Split-Half Coefficient	Items
Pre-test	0,50	0,10	5
Post-test	0,35	0,48	7

Table 4. Factor analysis of the Pre-test questionnaire (KMO measure = 0,51)

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	1,92	38,30	38,30
2	1,26	25,24	63,54
3	0,92	18,48	82,02
4	0,60	12,05	94,07
5	0,30	5,93	100,00

Table 5. Factor analysis of the Post-test questionnaire (KMO measure = 0,60)

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2,24	32,04	32,04
2	1,50	21,36	53,40
3	0,99	14,20	67,60
4	0,73	10,46	78,06
5	0,70	10,05	88,11
6	0,49	6,93	95,04
7	0,35	4,96	100,00

Moreover, our evaluation was partially hampered because some data were discarded during the normalization and exclusion procedures. Comparisons involving HR and HRV measures had a loss of almost 60% of data.

According to Combatalade [9], despite the precaution taken in relation to skin preparation, conductive gel application, electrode placement and user instructions, it is very difficult to save HR data absolutely clean and no noise. It forces the use of a normalization process to the HR signal, especially to detect two types of artifacts: missed beats and extra beats.

Missed beats can occur when the signal is so distorted that the software is unable to identify the beat pattern and only picks up on the next good beat, which results in long inter beat interval value. On other hand, extra beats can occur when the program confuses a distortion in the signal for a beat and detect two or more beats, when there should be only one, resulting in short inter beat interval value at the end of analysis.

In order to reduce artifacts, a software solution to process HR signals and analyze HRV measures can be adopted. In this case, it is important to be assisted by a medical professional in order to ensure that the data cleaning does not interfere in future results.

It is important to mention that there are also real natural physiological events, similar to these artifacts, like premature trial or ventricular contractions. As a result of this, it is recommended to follow a specific protocol if there is presence of subjects with heart diseases and/or under medication.

Another suggestion to minimize the occurrence of noise in HR signal, it is the use of self-adjustable wrist straps, which prevents the electrodes become tightened or loosened on the user's wrists. It is possible to use a non-invasive device fixed on the user's chest, closer to the heart, which reports the HR to the ECG sensor without using wires.

At last, it is also recommended a collaborative effort between Computer Science and Medicine experts, in order to define guidelines allowing a better understanding about the user's behavior during the interaction process using physiological measures to do it.

6 Conclusions

This work presents a methodology for assessing the quality of ITs in 3DUIs, based on the use of physiological measures during the interaction process. A case study comparing two ITs was executed, in order to apply this new method through a well-defined testing protocol, two procedures to normalize and exclude some samples from physiological dataset, and the application of multivariate data analysis methods, which allow to highlight the relationships between physiological measures and usability metrics.

The use of our methodology pointed to promising results and interesting tendencies. We except our methodology to evolve and to be consolidate through new comparisons with other ITs, not only to evaluate selection, manipulation or navigation tasks in VEs, but also to consider another features of 3DUIs, like computer stereo vision, different degrees of freedom and multiple sensory stimuli.

Physiological measures still cannot be considered as substitutes of user task performance or user preferences. However, these measures can be used as a complementary resource for the interpretation of usability metrics commonly observed during the evaluation processes, highlighting the relationship between NN50 psychological measure and "Total time" task performance measure. We recommend a more detailed study of this measure for future work, in order to contribute and qualify the ITs evaluation process.

Our research also presents a mode to validate the proposed questionnaires, as a means of establishing better relationships between physiological measures and subjective user preferences. This analysis pointed to the importance of scale adjustment for subjective evaluation before testing, based on the available sample size of subjects. This fact may have contributed to measures with significant differences, such as SC and NN50, have shown low statistical power in this work.

Regarding the comparisons between physiological measures and usability metrics, we highlighted the need to use multivariate data analysis techniques during the evaluation process. These statistical methods allowed the understanding and the detailed case study of relationships between different measures, encouraging the formulation of more refined conclusions and the indication of trends for future studies in this promising area.

Finally, we emphasize that our project is a first step to define a specific methodology to the 3DUI evaluation process, considering the use of physiological measures and appropriated statistic techniques to compare multiple datasets. The continuity of our research involves new case studies and a multidisciplinary effort between Computer Science and Medicine experts to create guidelines for clear and wide association of the user's behavior and their actions performed in 3DUIs using different interaction resources.

References

1. American College of Sports Medicine: ACSM's Advanced Exercise Physiology. Lippincott Williams & Wilkins, Philadelphia (2005)
2. Bisbal, J., Lawless, D., Wu, B., Grimson, J.: Legacy Information Systems: Issues and directions. IEEE Software 16(5), 103–111 (2002)
3. Bolt, R.A.: "Put-That-There": Voice and Gesture at the Graphics Interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, pp. 262–270. ACM, New York (1980)
4. Bowman, D.A., Hodges, L.F.: An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In: Proceedings of the 1997 Symposium on Interactive 3D Graphics, pp. 35–38. ACM, New York (1997)
5. Bowman, D.A., Gabbard, J.L., Hix, D.: A Survey of Usability Evaluation in Virtual Environments: classification and comparison of methods. Presence: Teleoperators and Virtual Environments 11(4), 404–424 (2002)
6. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces: theory and practice. Addison-Wesley, Boston (2004)
7. Brogni, A., Vinayagamoorthy, V., Steed, A., Slater, M.: Variations in Physiological Responses of Participants during Different Stages of an Immersive Virtual Environment Experiment. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, pp. 376–382. ACM, New York (2006)
8. Brogni, A., Vinayagamoorthy, V., Steed, A., Slater, M.: Responses of Participants during an Immersive Virtual Environment Experience. The International Journal of Virtual Reality 6(2), 1–10 (2007)
9. Combatalade, D.: Basics of Heart Rate Variability Applied to Psychophysiology. Technical report MAR953-00, Thought Technology Ltd. (2010)
10. Cronbach, L.J.: Coefficient Alpha and the Internal Structure of Tests. Psychometrika 16(3), 297–334 (1951)
11. Guttman, L.: A Basis for Analyzing Test-Retest Reliability. Psychometrika 10(4), 255–282 (1945)
12. Guyton, A.C., Hall, J.E.: Textbook of Medical Physiology. Saunders, Philadelphia (2005)
13. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis. Prentice Hall, New Jersey (2005)

14. Hix, D., Hartson, H.: Developing User Interfaces: ensuring usability through product & process. John Wiley & Sons, New Jersey (1993)
15. Karvonen, M.J., Kentala, E., Mustala, O.: The Effects of Training on Heart Rate: a longitudinal study. Annales Medicinae Experimentalis et Biologiae Fenniae 35(3), 307–315 (1957)
16. Kim, Y.Y., Kim, E.N., Park, M.J., Park, K.S., Ko, H.D., Kim, H.T.: The Application of Biosignal Feedback for Reducing Cybersickness from Exposure to a Virtual Environment. Presence: Teleoperators and Virtual Environments 17(1), 1–16 (2008)
17. Lin, T., Omata, M., Hu, W., Imamiya, A.: Do physiological data relate to traditional usability indexes? In: Proceedings of the 17th Australia Conference on Computer-Human Interaction, pp. 1–10. Computer-Human Interaction Special Interest Group of Australia, Narrabundah (2005)
18. Lin, T., Imamiya, A., Omata, M., Hu, W.: An Empirical Study of Relationships Between Traditional Usability Indexes and Physiological Data. Australasian Journal of Information Systems 13(2), 105–117 (2006)
19. Lin, T., Imamiya, A.: Evaluating usability based on multimodal information: an empirical study. In: Proceedings of the 8th International Conference on Multimodal Interfaces, pp. 364–371. ACM, New York (2006)
20. Lin, T., Imamiya, A., Mao, X.: Using Multiple Data Sources to get Closer Insights into User Cost and Task Performance. Interacting with Computers 20(3), 364–374 (2008)
21. Malhotra, N.K.: Marketing Research: an applied orientation. Prentice Hall, New Jersey (2006)
22. Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal 17(3), 354–381 (1996)
23. Meehan, M.F., Insko, B., Whitton, M., Brooks Jr, F.P.: Physiological Measures of Presence in Stressful Virtual Environments. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, pp. 645–652. ACM, New York (2002)
24. Meehan, M.F., Razzaque, S., Insko, B., Whitton, M., Brooks Jr, F.P.: Review for Four Studies on the Use of Physiological Reaction as a Measure of Presence in Stressful Virtual Environments. Applied Psychophysiology and Biofeedback 30(3), 239–258 (2005)
25. Rosson, M., Carroll, J.: Usability Engineering: scenario-based development of human-computer interaction. Morgan Kaufmann, San Francisco (2001)
26. Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., Friedman, D.: Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment. Presence: Teleoperators and Virtual Environments 15(5), 553–569 (2006)
27. Tanaka, H., Monahan, K.D., Seals, D.R.: Age-predicted Maximal Heart Rate Revisited. Journal of the American College of Cardiology 37(1), 153–156 (2001)
28. Tullis, T., Albert, W.: Measuring the User Experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann, San Francisco (2008)

Comparing User Experience and Performance in SecondLife and Blackboard

Alistair G. Sutcliffe and Amal Alrayes

Manchester Business School, University of Manchester
Booth Street West, Manchester M15 6PB, UK
ags@man.ac.uk

Abstract. Collaborative problem solving was compared in SecondLife (SL) and Blackboard (BB) and both technologies were compared with a face-to-face (FTF) control condition. There were no performance differences overall, although FTF was quicker and preferred, followed by BB and SL. BB was perceived to be more usable, whereas SL provided better user experience. Worse performance was indicated by dislike of avatar interaction in SL, and poor user experience in BB, whereas better performance was associated with engagement with avatars, and better usability in BB. The affordances for collaboration in each technology are discussed, with reflections on the mixed methods approach using qualitative and quantitative data analysis.

Keywords: Collaborative problem solving and learning, Affordances, Mixed methods evaluation.

1 Introduction

The shortcomings of collaborative virtual environments (CVEs), such as a poor sense of presence, and limited non-verbal communication, have been pointed out by several authors [11]. However, CVEs are clearly successful for multiplayer games (such as World of Warcraft [5]), which has led educators to explore the potential of Second Life as a CVE that might motivate collaborative learning [9,22]. Some studies of collaboration in CVEs have suggested that presence and user experience can be superior to conventional 2D interfaces since avatars provide improved awareness of others and shared tasks [23,31]. However, comparisons of performance between CVEs and real-world equivalents have failed to show any clear advantage for virtual worlds. For example, in training medical students in interviewing skills for diagnosis, the VE performed as well as real patients but was less satisfying [13]. SL can facilitate sharing experience and personal information, although Neustaedter et al.'s [17] qualitative study suggests that SL complements real-life experiences rather than being an effective substitute for them.

Several issues and criteria have been proposed in Computer Supported Collaborative Work (CSCW) evaluation frameworks [21], such as group characteristics, situation factors (context), individual characteristics, task properties, group process, and task

P. Campos et al. (Eds.): INTERACT 2011, Part III, LNCS 6948, pp. 680–696, 2011.

and group outcomes. Neale et al. [16] added shared awareness to this list; however, as they pointed out, shared awareness is complex and involves knowledge of others' roles, identities, goals, and activities. Comparing affordances for collaboration in 3D virtual environments and conventional 2D interfaces was an initial motivation for this study. Evaluations in collaborative technology have tended to follow either an ethnographic approach to investigate the context of use in depth, or more focused experimental analyses directed towards specific questions about collaboration (e.g. [19]). A second motivation for this study was to investigate the mixed methods approach for understanding how affordances for collaboration contribute to user experience as well as performance and learning.

In this paper we investigate interaction and collaboration in SecondLife (SL) as an example of a 3D desktop CVE and Blackboard (BB) as a more traditional collaborative interface. The study did not examine collaborative learning per se, where longer-term investigation would be necessary; instead, our focus was on understanding how user experience and motivation may create sufficiency conditions for collaborative learning in one Computer Supported Collaborative Learning (CSCL) technology (BB) and one potentially more engaging technology (SL) which is being adapted for CSCL purposes. Two research questions are addressed: first, will the motivation of interacting in SL improve performance and user engagement in comparison to BB? Performance is this case was effectiveness of collaborative problem solving. The second question concerned investigating the quality of the user experience and possible influences of experience on performance with different technologies, using a mixed (qualitative and quantitative) methods approach. In following sections of this paper, related work in CVEs and CSCL technologies is reviewed, then the experimental methods are described. This is followed by quantitative data analysis and qualitative investigations into user experience.

2 Related Work

Studies of CSCL environments have reported improved levels of learning [8] and participation [7,12] compared to a face-to-face (FTF) interaction. In contrast, other studies have found that students working in CSCL environments can perceive discussions as more confusing [28] and less productive [25]. Furthermore, CSCL environments can produce lower levels of student participation [13], higher levels of conflict [10], poor group cohesion [25] and lower levels of satisfaction [1]. These contradictions suggest further in-depth studies of the causes of either failure or success of CSCL technology are necessary.

Familiarity among participants is well known to have a beneficial effect on technology mediated group work [19], since knowledge of others can enhance social awareness and organisation of work. However, in CSCL previous research is divided about the effect of group members' familiarity and performance. Newcomb and Bagwell [18] found collaboration between friends produced more intense social activity, more frequent conflict resolution, more effective task performance, greater equality, and loyalty. Familiar groups may also help creation of effective shared problem-solving spaces [2,3].

In contrast, Maldonado et al. [15] found a strong negative correlation between prior friendship and groupwork performance. Furthermore, working with friends can yield lower-quality outcomes because friends indulge in more non-task related behaviour, with stronger pressures to agree, and unwillingness to be critical of each others' ideas [6,34]. Groups of friends may also disagree more frequently [24] and become more concerned with resolving disagreements [18]. Thus it appears the jury is out on the effect of social relationships in collaborative learning.

User experience (UX) has many definitions arising from research on aesthetics [29,30] which extended traditional concepts to show that users' response to interaction has an affective dimension. However, aesthetics alone fails to capture the interactive elements of user experience especially in virtual worlds such as SL [27], so we adapt concepts from virtual reality [31,33] to augment measures of UX.

3 Methods

The hypothesis that interactive 3D worlds would motivate more effective collaboration and learning was investigated by comparing cooperative problem solving in SL and BB, a leading 2D collaborative learning environment. The role of affordances for technology-mediated collaboration were investigated to evaluate how 2D user interface features in BB compared to 3D interface in SL.

Sixty three participants (30 male, 33 female) took part in the experiment, organised in 21 teams with 3 members in each team. The participants were 24 undergraduate students, 8 postgraduate (Masters) students and 31 postgraduate (PhD) students at Manchester University. The participants were allowed to select their own team members. The majority (15 out of 21) of the teams were mixed gender. All participants were familiar with BB so no further training was necessary. Only one had used SL so training in user interface operation, avatar control and communication was given.

The teams had to solve three analogous versions of a survival prioritisation problem. Three different analogues were used (Lost in the Desert, Lost at Sea and Survival on the Moon) to minimise direct learning effects across repeated trials. Participants were presented with lists of items which might be useful or useless for survival in each scenario environment. The goal in each scenario was to collaboratively decide the best ranking of the items in order to survive. The items and context differed so no direct transfer of problem solutions was possible, although meta-level and process learning, i.e. approach to the problem, and organising collaborative problem solving was possible. The prime hypothesis therefore focused on performance:

Since motivation and user engagement with CSCL technologies are important precursors for effective learning, the secondary motivation for the study was to investigate the differences in interaction and collaboration between the two technologies (SL and BB) by comparing them with the control FTF condition.

A within-groups repeated measures design compared BB, a standard non-graphical user interface and SL, a 3D graphical environment with avatar-mediated interaction, with a FTF control condition. The order of tasks and technology conditions was counterbalanced. The independent variables were modality conditions and collaborative functions provided by the technology:

FTF: all modalities plus co-presence. Support for collaboration: notepad for handwritten records.
BB: asynchronous text message, e-mail and chat. Support: message threading, exchange of documents as attachments, electronic notes created using Word.
SL: visual communication via avatars, movement, position and limited gesture; near-synchronous text chat. Support: manipulation of text objects and lists in the virtual world.

In the control FTF condition the three team members were co-present in one room so they could converse naturally to arrive at a solution. They had access to the problem narrative and notepads to make lists or notes. The solution was recorded as a word processed list of objects in rank order. In BB each team member worked on a separate PC connected to BB (Figure 1a). The solution was presented as a word processed list of items ranked in order of their utility for survival. In SL each team member worked on a separate PC and was assigned an avatar of their own gender. They could move, pick up and manipulate numbered objects which could be placed in slots by item on a text list to denote the rank order. Hence the solution could be collaboratively constructed and was presented as a ranked list created in the 3D environment. Communication was by text chat bubbles associated with each avatar (see Figure 1b).

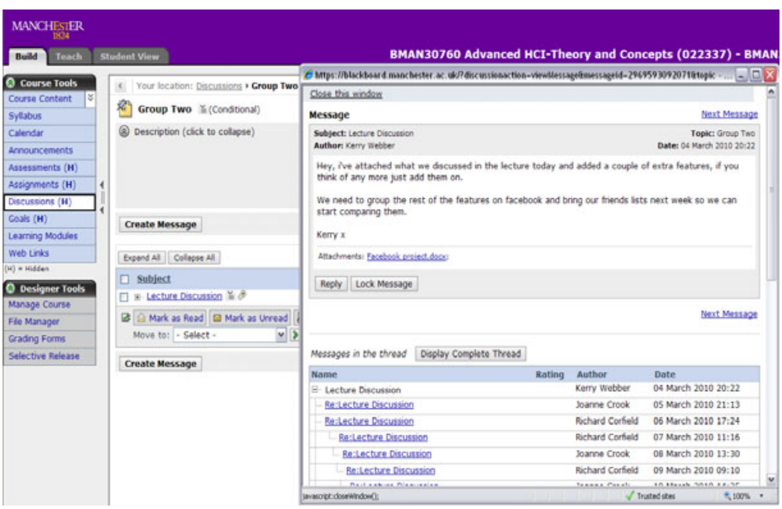


Fig. 1a. Blackboard user interface showing a group discussion thread

Fig. 1b. Screen dump showing SecondLife with group discussion

Voice communication was not used, since we wished to focus on the avatar versus conventional UI interaction, rather than compare communication modalities. Furthermore text chat is commonly used in SL when audio output can be annoying in public contexts such as co located class with several groups. Text communication had the added advantage of recorded transcripts of conversations for qualitative analysis. No verbal or FTF visual contact was allowed in either the BB or SL conditions.

After completing the task, each participant filled in a questionnaire with four sections rating user experience on 1-7 Likert scales (affect, general measures of UX, interaction quality, and presence). A further questionnaire rated satisfaction and perception of their own performance (decisions made, challenge, motivation) and overall preference. The experiment ended with a team interview to investigate the participants' experience, feelings and reflections on collaboration using the two technologies and in the control conditions. Questions probed the participants' perceptions of the quality of interaction, interactive experience with the technologies, critical incidents and usability problems with reflections on user experience in each condition. The participants' behaviours were observed during the experiment by taking notes and video recordings in SL sessions. BB chat room logs were saved for further analysis of users' discourse. Observations focused on participants' interaction with the technologies, artefacts in the real world or representations in BB and SL, and patterns of communication within the groups.

To summarise, the dependent variables were: group performance (task completion times, errors) individual participants' ratings of the quality of interaction and collaboration with each technology and FTF. The post-test rating data was ranked by means within each condition to simply presentation of the results. Performance differences between the technologies and scenario tasks were tested by ANOVAs, and post hoc tests. Qualitative data analysis focused on patterns of collaboration (observation and video logs) and the participants' recollection of the effectiveness of group working, critical incidents and breakdowns. Interview transcripts were coded following the open coding and axial/categorical aggregation conventions [26], then the transcripts were inspected for excerpts which illustrated the more frequent topics and issues reported by the respondents.

4 Results- Quantitative Data

Social closeness was calculated by summing the relationship strength reported by each group member on an 8-point scale and calculating the percentage of the maximum possible score (48). None of the groups contained three close friends (M 23.7%, range 12-65%) and only five groups reported an aggregate closeness >50%. No correlations were found between performance and social closeness in the groups. Indices for group interaction were calculated from the overall number of contributions factored by the contribution ratio:

$$\text{Interaction} = \text{TotContrib} * (\text{highest/lowest member contributions})$$

There were no correlations between these measures and performance accuracy and times, so the quantity of interaction within the groups did not appear to influence performance overall or in any of the conditions.

Not surprisingly, since BB was the university's standard collaborative learning environment, most participants had greater experience with BB (M 4.91) than with SL (M 2.13 on a 7-point scale, range 1=once to 7=daily use). The participants had limited knowledge of the scenarios (M 2.91) so for most it was a novel problem.

Performance

Expert solutions for each problem analogue were taken from the literature and used to calculate the group performance by scoring goodness-of-fit for the priority order of the ideal solutions against each group's solution. There was a significant difference between the task scenarios $F(2,8)=74.043$, $p<0.000$, with solutions for the Moon survival scenario being more accurate than the other two, but there was no main effect on performances between the three environments, and no interaction; see Table 1.

Table 1. Percentage of correct scores by scenario and technology

	Desert	Sea	Moon
FTF	47	43	70
SL	49	47	74
BB	53	49	76

All groups completed the task successfully although the average scores in the Desert and Sea scenarios showed room for considerable improvement; see Table 1. There were no apparent learning effects and the groups did not improve their performance after successive exposures to the task.

Groups completed the task more rapidly FTF, as might be expected (M 11.23 minutes), compared to 26.80 mins for SL and 24.00 mins for BB; while the Desert scenario was completed more quickly (M 18.47) than the Moon (M 21.53) or Sea scenarios (M 22.03). There was a significant main effect for technology (F=13.5

(2,28) p<0.001) but not for task. FTF was faster than the technologies but there were no differences between SL and BB (post hoc tests p<0.05). The Moon task was completed more accurately although it took slightly longer than the Desert scenario. In the technology conditions, FTF was quicker, but BB was more accurate, although these difference were small and not significant.

Group interaction

Interaction within the groups, measured by message exchange and chat posting, was more frequent in BB (M 113.87 messages), than SL (M 107.40). There were slightly fewer interactions for the Moon task (M 103.55) than for Desert (M 106.20) or Sea (M 122.15). Although there were no significant main effects, the interaction (scenario x technology) was marginally significant F(2,9) 3.48, p<0.05, with more interactions in SL for the Sea and Moon tasks and more in BB for the Desert task. FTF interactions were not recorded since dialogue exchanges were too rapid for accurate recording, so no direct comparison was possible; however, informal observation suggested that FTF interaction was most frequent.

Post-test questionnaire ratings

There were significant differences in nearly all the post-test rating scores, with the control condition tending to be most favoured, followed by BB with SL in third position (p>0.001 ANOVAs, with post hoc tests on individual ratings). The results of post hoc tests are reported in rank order of mean scores where, by default, differences between all three combinations were at least p<0.05, apart from the table cells shaded to show where the differences between the technologies were not significant even though they did differ from FTF. In the affect measures, illustrated in Table 2, FTF was first for pleasure and joy, while SL was first for both the positive emotion of surprise and the negative emotions of anxiety, frustration, fear and disgust. In all categories except pleasure, BB had lower mean scores than SL, but these differences were not significant.

Table 2. Rank order of means of the affect rating scales
shading = no significant difference between technologies (BB and SL)

Measure	Rank order 1-2-3 by means		
Pleasure	FTF	BB	SL
Surprise	SL	BB	FTF
Anxiety	SL	BB	FTF
Joy	FTF	SL	BB
Frustration	SL	BB	FTF
Disgust	SL	BB	FTF
Fear	SL	BB	FTF

In the user experience measures illustrated in Table 3, not surprisingly FTF was most positive; however, SL had higher means than BB for all measures except the motivation for repeated use (use again), although these differences were not significant.

Table 3. User experience rank order by means

Hold your attention	FTF	SL	BB
Feel excited	FTF	SL	BB
Good mood after using	FTF	SL	BB
Use again	FTF	BB	SL
Vivid memory	FTF	SL	BB
Memory good or bad	FTF	SL	BB

The FTF control condition was most favoured, then BB was a clear winner over SL in all measures, with a significant difference for clear design ($p>0.05$ post hoc test); see Table 4.

Table 4. Interaction quality measures, rank order by means

Convenient use	FTF	BB	SL
Easy to use	FTF	BB	SL
Easy to navigate	FTF	BB	SL
Clear design	FTF	BB	SL

Interaction quality measures indicated that SL was perceived as complex and challenging than BB ($p>0.05$) while BB gave the best awareness of external events, possibly reflecting the information-intensive user interface. SL was ranked second to FTF in engagement and (less) awareness of the user interface, realism indicating some benefit from presence and immersion in the 3D graphical world and avatars; but BB was rated better for natural feel than SL ($p<0.05$) and being absorbed, so the usability problems in SL (Table 5) may have disturbed the sense of presence.

Table 5. Ranks order by means for interaction quality and presence from [33]
Shading = no significant difference between technologies (BB and SL)

Good pace (speed) of interaction	FTF	BB	SL
Pace of interaction interesting	FTF	BB	SL
How complex was the interaction	SL	BB	FTF
How challenging	SL	BB	FTF
How engaged	FTF	SL	BB
Awareness of user interface	FTF	SL	BB
Natural feel	FTF	BB	SL
Awareness external events	BB	SL	FTF
Totally absorbed	FTF	BB	SL
Realistic	FTF	SL	BB

As expected, overall satisfaction favoured FTF; (see Table 6); $F(2,267)=45.88$, $p<0.000$); while participants were more satisfied with BB than SL, although this difference was only marginally significant ($p<0.05$ in post hoc tests).

Table 6. Environment satisfaction

Environment	Mean	SD
FTF	6.14	1.148
BB	4.90	1.41
SL	4.35	1.504

Satisfaction with decisions made followed the same pattern, illustrated in Table 7 ($F(2,267)=10.725$, $p<0.001$) with BB in second place; the difference between BB and SL was not significant, although FTF was better than both BB and SL ($p<0.01$, post hoc tests).

Table 7. Satisfaction with decisions made

Environment	Mean	SD
FTF	6.17	1.199
BB	5.50	1.105
SL	5.13	1.561

SL was considered to be more of a challenge than the other conditions (Table 8), probably because it was less familiar to the participants ($F(2,267)=20.469$, $p=<0.0001$), while FTF was best for overall motivation ($F(2,267)=20.469$, $p<0.001$) with BB in second place and SL third, although the means were not significantly different.

Table 8. Challenge and overall motivation scores

		Mean	SD
Challenge	FTF	3.31	2.053
	SL	4.71	1.637
	BB	4.16	1.476
Overall motivation	FTF	5.93	1.428
	SL	4.60	1.654
	BB	4.81	1.413

To summarise, performance was no better FTF than with SL and BB technologies, which is surprising since the CSCW technologies can constrain discussion and fluid interaction. The control FTF was rated best on experience, operational ease of use, positive affect and overall satisfaction. SL was in second position for user experience items, whereas BB was second on usability measures, although most differences were not significant. SL evoked mixed emotions with both surprise and fear, anxiety and frustration. For overall satisfaction there no significant difference between SL and BB.

5 Results: Qualitative Data Analysis

The patterns of positive/negative comments in the post-test interviews agreed with the quantitative measures, demonstrating that most individuals preferred FTF interaction over SL or BB, citing natural communication as the main reason (30%), followed by ease of interaction (22%) and rapid interaction (18%). An equal number of positive and negative comments (44) were reported for BB, with familiarity (20%) and ease of use (19%) heading the list, followed by simple interface, easy to focus and communication (11%). The more frequent negative comments were boring interaction (30%), poor response time (25%), didn't like text-only communication (18%), and hindered discussion (14%).

SL attracted more negative than positive comments overall (55%). Among the more frequent positive aspects were enjoyable and fun (41%), close to reality (34%) and interaction quality (15%). In contrast, the downsides were unfamiliarity (14%), distracting interface (14%), and navigation problems (15%), with most other issues relating to complexity, general frustration and usability issues.

The general themes which emerged for BB reflected users' perceptions of a simple and familiar interface which was effective for the task in hand, although it was boring and not very exciting. The collaborative affordances were limited to communication and document exchange. In contrast, SL was perceived to be more dynamic, stimulating and interesting, but with downsides of being distracting and annoying. Collaboration afforded by avatars was motivating in terms of curiosity but not seen as relevant to the task. Several comments were made that the avatars were not faithful representations and the absence of facial expressions reduced the sense of presence. Furthermore, the limited gestures and difficulties in moving them reduced the effectiveness of avatars, with several respondents commenting that interaction evolved into a chat room format making the avatars irrelevant. For example:

> SL is very restrictive, as not only do you have to interact with a keyboard, you also have to use a mouse to interact and move around. The avatars are a distraction, because even though they represent you, they do not show emotion or body language. Using the mouse to interact with items is hard.

> SL provided too many distractions. For example, moving around and moving boxes made it harder to make a good choice. BB was a far simpler interface which made it easier to focus on the task.

Comments in favour of SL usually focused on the user experience and novelty:

> I really enjoyed the SL experience. Maybe as it is my first time so it's quite novel. I also felt hugely immersed in this environment in comparison to BB. BB wasn't much fun. SL was more absorbing due to the avatars and looking at how they behaved in SL.

And one of the few who did not rate FTF first commented:

> I prefer using SL and BB over FTF because you get in contact with different types of people without the need to get involved with them in real life. And SL is more fun, interesting and less boring that FTF.

Pointing to and manipulating the numbered objects in SL was infrequent, and most groups discussed the prioritisation order; one individual was nominated to move the object to record the result.

Discussion was effective in all groups; however, individuals in four groups reported excluding or ignoring the opinions of others, and six groups commented that, given the opportunity to repeat the task, they would aim to be more inclusive and improve discussion.

The quantitative data suggested that overall there was little to choose between BB and SL; in contrast, the qualitative data analysis suggested different reasons for liking or disliking each technology. The interviews were analysed to see if the performance in the five best and five worst groups showed any patterns that might link users' reactions to the technologies and their performance.

Comparison of the Top 5 and Bottom 5 groups

Performance among the groups was normally distributed, so groups in the tails of the distribution were investigated in more depth.

Comments made by the five best and five worst performing groups were compared, to tease apart possible reasons for success and failure between the conditions. In FTF, the comments of the top and bottom five groups were nearly all positive and followed the general pattern of favouring natural communication ability to see facial expressions and non-verbal communication, use speech efficiently, and the social advantages of being able to get to know other group members. The best performing groups made more comments overall and more positive (60%) than negative comments; conversely, the five worst performing groups made more negative (57%) than positive comments. The distribution of comments by valency is shown in Table 9.

Table 9. Valency of comments expressed as percentage of total comments separately for the top 5 and bottom 5 groups

	Best 5 groups		Worst 5 groups	
	+ve	-ve	+ve	-ve
SL	30	15	22	35
BB	30	35	10	33

For the five best groups, positive comments were evenly distributed, but they made more negative comments about BB. However, most of these comments were criticisms about the need for improving its functionality. The five worst groups made a similar number of negative comments about both technologies, with most comments concerning poor usability. These groups were more positive about SL, with comments on the engagement of the 3D world and avatars. So, even though both groups made a similar number of negative comments about BB, for the better performers these were complaints about missing requirements whereas the worst performers were reporting usability problems. This tendency was also apparent for SL.

For BB, the five top groups reported that communication was quick, easy and efficient, with downsides of not being exciting and text-only communication being limited. In contrast, the five bottom groups reported more negative comments such as "boring interface, limitation of text-only interaction, slow, frustrating and difficult for effective discussion".

In SL, the reasons were quite different. The better performing groups all made favourable comments about the sense of presence, interesting and exciting interaction, reporting novel and exciting experience with few negative comments. In contrast, the poorly performing groups all made frequent criticism that SL was difficult to use and navigate, the graphics and interaction felt artificial, also that it was unfamiliar and frustrating to use, although there were some positive comments on presence and exciting interaction.

This analysis, in combination with the quantitative data for all groups, suggests that BB worked well for most groups, apart from a few poor performers who appear to have experienced more usability problems and become bored by the text-only interface. SL on the other hand worked well only when groups were motivated by the excitement and novelty of the 3D world, but for most, poor usability hindered its effectiveness.

Order effects

There was no evidence that the order in which the groups experienced the technologies or control condition affected performance; neither did the scenario order show any performance differences. However, the participants' comments did reveal interesting order effects. All six groups who encountered the FTF condition first commented that this allowed them to get to know each other quickly and help negotiate ways of collaborating. These collaborative processes enabled these groups to work more effectively in the technology conditions. Conversely, participants in seven of the groups who experienced one of the technology conditions commented that it would have helped to get to know each other first FTF. The following three quotes illustrate the advantages of meeting FTF before experiencing CMC technologies:

> We started from FTF; that made us more familiar with each other. It is better for us in the following discussions. That's why we seemed to be the first one to finish the tasks. The worst way is starting from SL as it is difficult to handle and it doesn't make sense to discuss in SL. I know there are differences between SL and BB but not that much different in these two tasks.

> Doing FTF first was very beneficial as we got to know and established a working process. Then we were able to take our online tasks in similar process to real life. I feel this allowed us to work quicker and more productively. So definitely it is beneficial in my opinion for a group to work together offline before online together.

> I had FTF, BB and SL because the FTF exercise was first; it was easier to organise a strategy, and it influenced the other exercise strategies and made it easier to carry them out. Because FTF was first, I understood the way the other participants thought and interacted through their body language and emotions, by interacting with them and understanding how to work with them in a group.

These comments reflect the limitations of communication in the technologies which do not appear to have been mitigated either by avatars in SL or shared awareness functions in BB.

However there were a few comments which did favour the affordances of the technologies:

> Shyness and not knowing group members hindered. Easier to start discussion with complete strangers on BB and SL.

Learning and Reflection

Most of the groups (76%) reported that they did develop a common strategy to solve the problem and 4 out of 5 of the worse performing groups were among the groups without a common strategy. However nearly all the groups (86%) reported that their strategy did not change over iterations in the task, so there was little evidence of process learning. When asked in the final interview to reflect on how their performance could be improved, 19/21 of the groups suggested improvements. The majority concerned process (31%) concerned process such as better ways of structuring discussion, voting or agreeing prioritisation; 18% cited the need for more information on the problem domains, while 17% cited the technology and hindering performance, in particular the need for voice communication and better shared awareness of the problem space (item list and priorities agreed). Five groups reported social interaction could be improved (11% comments) while the remaining comments concern various issues such as the need for more time. Three of the bottom 5 performing groups encountered these social problems.

While learning may have emerged with more trials, the fact that most groups did form a strategy and the strategy did not change implies the technologies did not help learning or performance; furthermore, the reported social problems indicate that social awareness also needs to be improved.

In general reflections, technology was mentioned in 26% of comments spread across all groups. The need for better shared awareness and common ground was cited frequently; however, there are differing views on the affordances of technology compared to FTF, as illustrated in the following excerpts:

> Working with people who are complete strangers is much more convenient in SL and BB because the discussion will be more specific, focus, and aim to solve the problem. And not to be shy. BB is a pretty plain interface but more straight forward and easy to read. In SL users need to adjust location of the avatar properly in order to get a proper view, navigation error will caused the avatar to move out of sight. Some people are more comfortable to share information via online rather than FTF. In the chatting room we can chat and discuss anything.

> I was more willing to argue my point in the FTF discussion whereas I was more inclined to avoid confusion in SL and BB hence exclude only the necessary opinions.

Overall neither technology provided effective affordances for collaboration and learning, and the groups did not appear to improve their modus operandi in spite of being aware that improvements could be made.

6 Discussion

Our findings demonstrate a positive outcome for the first research question: user experience with SL was richer than BB and more motivating; however, this did not produce better performance. So the answer to the second question we posed is that user experience in terms of affect and interaction does not appear to improve performance, although there was no difference between Blackboard and Second Life. The reasons for success or failure of the two technologies produced a complex picture. SL did not produce the expected benefit from the motivation of 3D interaction, probably because of usability problems encountered with the avatars. BB in contrast was perceived as being more usable, even though for some groups it was considered to be boring and not a stimulating user experience. FTF was expected to be most effective and indeed it was quickest and rated best on experience and positive emotions. However, FTF did not produce more accurate results, while BB did have a marginal, although non-significant, advantage. FTF may have been too easy for the participants so they did not pause to reflect and hence validate their solutions. In contrast, BB may have encouraged more reflection since communication was slower and messages were persistent.

Social relationships and overall group activities did not correlate with performance, so our results appear to agree with previous findings that social relationships have no positive effect on learning performance [15,24]. However, the degree of social closeness in most of our groups was limited, so our findings only relate to weak ties rather than stronger relationships.

The post-test attitude scores indicated a general pattern that BB was preferred for usability and overall, whereas SL was rated to be more exciting in terms of user experience. However, qualitative analysis of the five top and bottom performing groups showed that users may be divided into different cohorts: those who prefer the excitement of 3D experience and those who prefer more conventional task-oriented interfaces. The users' comments indicated that these reactions may have influenced performance in the tails of the distribution.

The avatars and affordances for collaboration provided by SL for the experimental task were limited. The avatars did not appear to motivate many users and did not provide a rich form of communication, contrary to the intuitions of media richness theory [4]. The numbered blocks did afford prioritisations but few users found this to be a natural form of collaboration. SL may be more advantageous in tasks where physical manipulation is necessary (e.g. construction, assembly and design). In the survival scenarios, the task was essentially abstract, and only involved discussion and prioritisation. As some users commented, the avatars hindered communication, which evolved into a chat session. Use of voice in SL may have enhanced chat and possibly increased social presence; however, whether use of the voice modality would have enhanced the affordances of the avatars is an open question for future study. Theoretical models indicate that concurrent motor control (of avatars) and verbal communication can increase workload [32] so avatars in non-physical tasks may be more of a hindrance than a help.

BB was familiar for many of the participants, which may explain its advantage in usability and some of the adverse reactions to the less familiar SL. However, BB was compared unfavourably with FTF, because of its poor communication and weak affordances for collaboration. Topics threads in message exchanges were not used; instead, most groups used the chat facility and a Word document for shared awareness of their prioritised list. In conclusion, the affordances for collaboration provided by both technologies were limited, as demonstrated by the experimental order which enabled groups using FTF first to establish an effective modus operandi that they subsequently transferred to the technology conditions.

The significance of social awareness within groups was apparent from the importance attached to FTF interaction and the users' reflection on their experience. So while prior social relationships may not be necessary for collaboration [15,34], some social familiarity does appear to promote group interaction and performance. When technology is being used to mediate learning, a prior meeting FTF is advisable to help participants develop common groups so they can adapt their modus operandi to the technology. Although shared awareness was only cited by some of our users, it did appear implicitly in many reflections on the limitations of both SL and BB technology, reinforcing Neale et al.'s [16] view that such functions need to be improved.

The mixed methods approach we adopted showed the advantage of combining quantitative and qualitative data. The performance data and attitude scores set the scene for the overall capabilities of the technologies, but the reasons for success or failure needed qualitative data to discover the reasons for users' experience, which may have influenced performance. However, the experimental paradigm we used is still limited by the scope of tasks and experience that can be investigated, and longitudinal studies of collaboration in context are also essential for assessing the effectiveness of collaborative technologies [16]. Mixed methods approaches in experimental designs can however produce useful insight for comparing the affordances of collaborative technologies.

The implications of this study are, first, to caution against the rush to adapt apparently engaging technologies such as SL for collaborative learning, since we did not find the expected motivation bonus. There also appeared to be a usability penalty. However, these downsides might have been mitigated by more extensive training. The second implication concerns collaboration more generally, where the order effect, we found, indicates that giving groups FTF contact before interacting via CSCW/CSCL technology can help users develop a modus operandi for collaboration and adapt more effectively to the technologies. A third implication reflects limitations of the task. SL may be more acceptable and effective in tasks requiring physical interaction and hence active avatar roles, so it may be important to consider the collaborative task and learning objectives when introducing SL. In our future studies we will explore different tasks and the effect of stronger social relationships where avatars might mediate a strong sense of social presence.

References

1. Baltes, B., et al.: Computer-mediated communication and group decision making: A meta analysis. Organizational Behavior and Human Decision Processes 87(1), 156–179 (2002)
2. Barron, B.: When smart groups fail. Journal of the Learning Sciences 12(3), 307–359 (2003)

3. Brown, J., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educational Researcher 18(1), 32–42 (1989)
4. Daft, R.L., Lengel, R.H., Trevino, L.K.: Message equivocality, media selection, and manager performance: Implications for information systems. MIS Quarterly 11(3), 355–366 (1987)
5. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: The life and death of online gaming communities: A look at guilds in World of Warcraft. In: Proceedings: CHI 2007, pp. 839–848. ACM Press, New York (2007)
6. Dutson, A., et al.: Review of literature on teaching engineering design through project oriented capstone courses. Journal of Engineering Education 86(1), 17–25 (1997)
7. Fjermestad, J.: An analysis of communication mode in group support systems research. Decision Support Systems 37, 239–263 (2004)
8. Hertz-Lazarowitz, R., Bar-Natan, I.: Writing development of Arab and Jewish students using cooperative learning (CL) and computer-mediated communication (CMC). Computers & Education 39(1), 19–36 (2002)
9. Hew, K.H., Cheung, W.S.: Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. British Journal of Educational Technology 41(1), 33–55 (2010)
10. Hobman, E., et al.: The expression of conflict in computer-mediated and face-to-face groups. Small Group Research 33, 439–465 (2002)
11. Irani, L., Hayes, G.R., Dourish, P.: Situated practices of looking: Visual practice in an online world. In: Proceedings: CSCW 2008, pp. 187–197. ACM Press, New York (2008)
12. Janssen, J., et al.: Visualization of participation: Does it contribute to successful computer-supported collaborative learning. Computers & Education 49, 1037–1065 (2007)
13. Johnsen, K., Raij, A., Stevens, A., Scott Lind, M.D., Lok, K.: The validity of a virtual uuman experience for interpersonal skills education. In: Proceedings CHI 2010, pp. 1049–1058. ACM Press, New York (2010)
14. Johnson, S., Aragon, S., Shaik, N., Palma-Rivas, N.: Comparative analysis of learner satisfaction, and learning outcomes in online, and face-to-face learning environments. Journal of Interactive Learning Research 11(1), 29–49 (2000)
15. Maldonado, H., Klemmer, S., Pea, R.: When is collaborating with friends a good idea? Insights from design education. In: Proceedings: CSCL 2009 (2009)
16. Neale, D.C., Carroll, J.M., Rosson, M.B.: Evaluating Computer-Supported Cooperative Work: Models and frameworks. In: Proceedings CSCW 2004, pp. 112–121 (2004)
17. Neustaedter, C., Fedorovskaya, E.: Capturing and sharing memories in a virtual world. In: Proceedings: CHI 2009, pp. 1161–1170. ACM Press, New York (2009)
18. Newcomb, A., Bagwell, C.: Childrens friendship relations: A meta-analytic review. Psychological Bulletin 117, 306–347 (1995)
19. Olson, G.M., Olson, J.S.: Technology support for collaborative workgroups. In: Olson, G.M., Malone, T.W., Smith, J.B. (eds.) Coordination Theory and Collaboration Technology, pp. 559–584. Lawrence Erlbaum, Mahwah (2001)
20. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
21. Pinelle, D., Gutwin, C.: A review of groupware evaluations. In: Proceedings: IEEE WETICE, pp. 86-91 (2000)
22. Salmon, G.: The future for (second) life and learning. British Journal of Educational Technology 40(3), 526–538 (2009)

23. Schroeder, R., Heldal, I., Tromp, J.G.: The usability of collaborative virtual environments and methods for the analysis of interaction. Presence, Teleoperators and Virtual Environments 15(6), 655–667 (2006)
24. Shah, P., Jehn, K.: Do friends perform better than acquaintances? The interaction of friendship, conflict, and task. Group Decision and Negotiation 2, 149–165 (1993)
25. Straus, S.: Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. Human–Computer Interaction 12, 227–266 (1997)
26. Strauss, A., Corbin, J.: Basics of qualitative research. Sage Publications, London (1998)
27. Sutcliffe, A.G.: Designing for user engagement: aesthetic and attractive user interfaces. In: Carroll, J.M. (ed.) Synthesis Lectures on Human Centered Informatics. Morgan Claypool, San Rafael CA (2009)
28. Thompson, L., Coovert, M.: Teamwork online: The effects of computer conferencing on perceived confusion, satisfaction and postdiscussion accuracy. Group Dynamics 7, 135–151 (2003)
29. Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. Interacting with Computers 13(2), 127–145 (2000)
30. Tractinsky, N., Zmiri, D.: Exploring attributes of skins as potential antecedents of emotion in HCI. In: Fishwick, P. (ed.) Aesthetic Computing. MIT Press, Harvard (2005)
31. Tromp, J.G., Steed, A., Wilson, R.: Systematic usability evaluation and design issues for collaborative virtual environments. Presence: Teleoperators and Virtual Environments 12(3), 241–267 (2003)
32. Wickens, C.: Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomics Science 3(2), 150–177 (2002)
33. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence 7, 225–240 (1999)
34. Zajac, R., Hartup, W.: Friends as co-workers: Research review and classroom implications. The Elementary School Journal 98, 3–13 (1997)

Author Index